Statistica Sinica Preprint No: SS-2024-0105	
Title	Efficient Estimation of the Accelerated Failure Time Model with Auxiliary Aggregate Information
Manuscript ID	SS-2024-0105
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202024.0105
Complete List of Authors	Huijuan Ma, Manli Cheng, Yukun Liu, Donglin Zeng and Yong Zhou
Corresponding Authors	Yukun Liu
E-mails	ykliu@sfs.ecnu.edu.cn
Notice: Accepted author version.	

Statistica Sinica

Efficient Estimation of the Accelerated Failure Time Model by Incorporating Auxiliary Aggregate Information

Huijuan Ma[1], Manli Cheng[1], Yukun Liu[1], Donglin Zeng[2], and Yong Zhou[1]

[1]*East China Normal University and* [2]*University of Michigan*

Abstract: With the rapidly increasing availability of aggregate data in the public domain, there has been a growing interest in synthesizing information from individual-level data and aggregate data. This article studies the maximum full likelihood estimation method to integrate the auxiliary information in the estimation of the accelerated failure time model. To overcome the computational challenges in maximizing full likelihood, we propose a novel one-step estimator, where the maximum conditional likelihood estimator without combining any auxiliary information is chosen as an initial estimator. We establish the consistency and asymptotic normality of the proposed one-step estimator and show that it is more efficient than the initial estimator. The asymptotic variance of the proposed one-step estimator has a closed form and is easily estimated by the plug-in rule. Simulation studies show that the proposed one-step estimator yields an efficiency gain over existing approaches. The proposed methodology is illustrated with an analysis of a chemotherapy study for Stage III colon cancer.

Key words and phrases: Accelerated failure time model; Empirical likelihood; Information synthesis

1. Introduction

The accelerated failure time (AFT) model is an important and attractive alternative to the Cox proportional hazards model (Cox, 1972) in regression analysis of censored data. It directly relates the log failure time to covariates and has a straightforward physical interpretation. Sometimes it can provide a more accurate or more concise summarization of data than the Cox model (Zeng and Lin, 2007). There has been a vast literature on the semiparametric analysis of the accelerated failure time model. Well-known methods include rank-based approaches (Prentice, 1978; Tsiatis, 1990; Wei et al., 1990; Ying, 1993; Yang, 1997; Jin et al., 2003; Zhou, 2005), least squared approaches (Buckley and James, 1979; Ritov, 1990; Lai and Ying, 1991; Jin et al., 2006), and semiparametric efficient approaches (Zeng and Lin, 2007; Ding and Nan, 2011; Lin and Chen, 2013).

In medical research and precision medicine, comprehensive individual-level data are recognized as the best source of evidence to produce desirable estimates of the effect of a medicine or therapy. However, concerns such as privacy restrictions and high costs usually make it challenging to collect as much individual-level data as possible in practice. With the rapidly increasing availability of aggregate data in the public domain, there is an urgent need to develop statistical methods with improved efficiency by incorporating information from both individual-level and publicly available aggregate data (Huang et al., 2016; Qin et al., 2022). The problem of synthesizing censored survival data with auxiliary information has been studied under various semiparametric models, including the Cox proportional hazards model (Huang et al., 2016; Huang and Qin, 2020; Sheng et al., 2021; Gao and Chan, 2023; Su et al., 2023; Shang, 2022), the additive hazard model (He et al., 2019; Ding et al., 2023; Shang and Wu, 2023), the semiparametric transformation model (Cheng et al., 2023), the additive–multiplicative hazard model(Shang and Wang, 2017), and the nonmixture cure model (Han et al., 2022). For the accelerated failure time model, Sheng et al. (2020) proposed an improved version of the generalized method of moments (Hansen, 1982, GMM) by combining auxiliary information with conventional weighted log-rank estimating equations. In general, rank-based approaches may not be semiparametric efficient, and the involved nonsmooth functions pose severe numerical challenges.

In this paper, we propose a maximum full likelihood estimation method to integrate the auxiliary information in the estimation of the accelerated failure time model. The full likelihood is composed of the conditional likelihood for the survival time given the covariate and the marginal likelihood of the covariate. The external auxiliary information, which is allowed to come from a population different from the target population, is incorporated into the marginal likelihood by using the classical empirical likelihood. Without auxiliary information, the proposed maximum full likelihood estimator reduces to the maximum conditional likelihood estimator. To bypass the calculation problem of the maximum full likelihood estimator, we construct a novel one-step estimator after investigating the first-order approximations of the aforementioned estimators with and without incorporating auxiliary information.

The one-step estimator is easy to calculate and has the same limiting distribution as the maximum full likelihood estimator.

This paper makes three contributions to the accelerated failure time model. First, without incorporating auxiliary information, we introduce a completely nonparametric conditional likelihood estimation method, together with an expectation-maximization (EM) algorithm. A smoothing technique is adopted to overcome the computational challenges posed by the discontinuity of step functions. The maximum conditional likelihood estimators for the regression parameter and the distribution of the error term are shown to be asymptotically joint normal. And the former estimator achieves the semiparametric efficiency lower bound, which implies that our method is no less efficient than the aforementioned rank-based approaches and least squared approaches. Compared with the semiparametric efficient approaches of Zeng and Lin (2007), Ding and Nan (2011) and Lin and Chen (2013), our method has less computation burden, is less sensitive to tuning parameters, and automatically produces an estimator for the error-term distribution. Second, when auxiliary information is available, we propose a full likelihood estimation approach to combine individual-level data with external auxiliary aggregate data, which are transformed into a system of estimating equations. The maximum full likelihood estimator is shown to be asymptotically normal and more efficient than the maximum conditional likelihood estimator, although its numerical calculation is challenging. Third, we propose a novel one-step estimator that is easy to implement and its asymptotic variance has a closed form and can be estimated by the commonly-used plug-in rule.

There are notable differences between our method and existing approaches to combining survival data with auxiliary information. As a pioneer work in this area, Huang et al. (2016) transformed the likelihood and auxiliary survival information that involve an infinite-dimensional parameter into a system of estimating equations that involves a finite-dimensional parameter, and then used empirical likelihood to incorporate auxiliary information. Their techniques were later extended to other models; see, for example, He et al. (2019), Huang and Qin (2020), Sheng et al. (2020, 2021), Ding et al. (2023), and Su et al. (2023), among others. Unlike them, we handle the finite-dimensional parameter and a nuisance nonparametric function simultaneously. Han et al. (2022) proposed a sieve method to estimate the infinite-dimensional parameter, while we develop an empirical likelihood approach to estimate the nonparametric error distribution function. Although Gao and Chan (2023) considered a general semiparametric model framework, their methodology and one-step estimator are not directly applicable to the AFT model. This is because the target parameter and the nuisance function under the AFT model are bundled in the likelihood function and the efficient influence function involves the derivative of the nuisance function, which has to be estimated separately. In contrast, we study the semiparametric estimation theory for the bundled parameters and our proposed EM algorithm can easily estimate the derivative of the nuisance function. Gao and Chan (2023) required their criterion function to be smooth enough, however the profile empirical likelihood function under the AFT model involves indicator functions, which are not smooth. Also our reparameterization technique and initial parameter estimation are different from Gao and Chan (2023). The bundle property not only poses serious numerical and theoretical challenges, but further differentiates our estimation framework from the existing ones.

2. Efficient estimation with auxiliary information

2.1 Model setup and full likelihood function

Let T, C and Z denote the failure time, censoring time and a p-dimensional vector of covariates, respectively. We assume the following accelerated failure time model

$$\log(T) = Z^\top \beta + \epsilon, \tag{2.1}$$

where β is a p-dimensional vector of unknown parameters, and ϵ is a random error independent of Z. The random error ϵ has distribution function $F(\cdot)$, density function $f(\cdot)$ and hazard function $\lambda(t) = f(t)/\{1 - F(t)\}$, which are all left completely unspecified. In the presence of right-censoring, we observe $X = \min(T, C)$ and $\delta = I(T \leq C)$, where $I(\cdot)$ is the indicator function. For a random sample of size n, the observed individual-level data are summarized as $(X_i, \delta_i, Z_i), i = 1, \ldots, n$.

We assume that T and C are independent given the covariate Z, and that the distribution of C does not depend functionally on β, that is, the censoring is noninformative. Given β, let $e_i(\beta) = \log(X_i) - Z_i^\top \beta$. The full likelihood function based on the observed individual-level data is proportional to

$$\prod_{i=1}^{n} \big\{f(e_i(\beta))\big\}^{\delta_i} \big\{1 - F(e_i(\beta))\big\}^{1-\delta_i} \{dF_z(Z_i)\}, \tag{2.2}$$

where $F_z(\cdot)$ is the distribution function of Z.

In addition to the individual-level data, suppose that certain aggregate information from external resources is available. For generality, we allow the external information to come from a population that is slightly different from the target population. Specifically, we assume that the external data follow an accelerated failure time model with the same regression parameter β but the error term has a slightly different hazard function

$$\bar{\lambda}(t) = \rho\lambda(t), \tag{2.3}$$

where $\rho > 0$ is an unknown scale parameter. When $\rho = 1$, it reduces to the homogeneous case where the individual-level data and aggregate information are from the same population.

Suppose that the aggregate external information is summarized by estimating equations

$$E\{\Psi(\beta, \rho, F; Z)\} = 0, \tag{2.4}$$

where $\Psi(\beta, \rho, F; Z) = (\Psi_1(\beta, \rho, F; Z), \ldots, \Psi_J(\beta, \rho, F; Z))^\top$ consists of J estimating functions. As one example, we assume the auxiliary subgroup survival probabilities are available (Huang et al., 2016; Sheng et al., 2020). Let Ω_j be the jth subgroup, and ζ_j be the corresponding t_j^*-year survival probability reported from external sources, $j = 1, \ldots, J$. Under model (2.3), for the target population, $\mathrm{pr}(T > t_j^* \mid Z \in \Omega_j) = \zeta_j^{1/\rho}$. To make use of such information, we can take $\Psi_j(\beta, \rho, F; Z) = I(Z \in \Omega_j)\{1 - F(\log t_j^* - Z^\top \beta) - \zeta_j^{1/\rho}\}$, where Ω_j, t_j^* and ζ_j are all known.

2.2 Empirical likelihood approach for estimation

The log-likelihood function, or the logarithm of (2.2), can be expressed as

$$\ell_n(\beta, F) + \sum_{i=1}^{n} \log\{dF_z(Z_i)\}, \tag{2.5}$$

where

$$\ell_n(\beta, F) = \sum_{i=1}^{n} \left[\delta_i \log\{f(e_i(\beta))\} + (1-\delta_i)\log\{1 - F(e_i(\beta))\}\right] \tag{2.6}$$

is the log conditional likelihood and $\sum_{i=1}^n \log\{dF_z(Z_i)\}$ is the log marginal likelihood. In the principle of the empirical likelihood method (Owen, 1990; Qin and Lawless, 1994), we model F_z by a step function $F_z(z) = \sum_{i=1}^n q_i I(Z_i \le z)$, where $Z_i \le z$ holds elementwise and q_i's should satisfy

$$q_i \ge 0, \quad \sum_{i=1}^{n} q_i = 1, \quad \sum_{i=1}^{n} q_i \Psi(\beta, \rho, F; Z_i) = 0. \tag{2.7}$$

The last constraint follows from equation (2.4).

By the Lagrange multiplier method, the marginal likelihood $\sum_{i=1}^n \log\{dF_z\,(Z_i)\} =$

$\sum_{i=1}^{n}\log(q_i)$ subject to (2.7) is maximized at $q_i = n^{-1}\{1+\nu^\top\Psi(\beta,\rho,F;Z_i)\}^{-1}$, where ν is the solution to

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\Psi(\beta,\rho,F;Z_i)}{1+\nu^\top\Psi(\beta,\rho,F;Z_i)} = 0. \tag{2.8}$$

Accordingly, after profiling F_z out in (2.5), we have the profile empirical log-likelihood (up to a constant),

$$L_n(\beta,\rho,\nu,F) = \ell_n(\beta,F) - \sum_{i=1}^{n}\log\{1+\nu^\top\Psi(\beta,\rho,F;Z_i)\}, \tag{2.9}$$

where ν satisfies equation (2.8). It is natural to estimate (β,ρ,F) by the maximum full likelihood estimator $(\hat\beta_{\mathrm{au}},\hat\rho_{\mathrm{au}},\hat F_{\mathrm{au}})$, which is the maximizer of (2.9) with respect to (β,ρ,F). However, as pointed out by Zeng and Lin (2007), after profiling out the nonparametric function F in (2.9), the objective function cannot achieve its maximum for finite β. In other words, the calculation of $\hat\beta_{\mathrm{au}}$ is intractable, which makes it impractically useful. To bypass this dilemma, in the next subsection we propose a one-step estimator that is easy to calculate and has the same limiting distribution as $\hat\beta_{\mathrm{au}}$. We start with an initial estimator, which is the maximum likelihood estimator without auxiliary information, and update it based on the first-order approximation of $(\hat\beta_{\mathrm{au}},\hat\rho_{\mathrm{au}},\hat F_{\mathrm{au}})$. Any existing semiparametric efficient estimator without using auxiliary information (e.g. Zeng and Lin, 2007; Ding and Nan, 2011; Lin and Chen, 2013) could serve as an initial estimator in the proposed one-step procedure. However, they are more or less inconvenient in computation: they may either be unstable

(Zeng and Lin, 2007), require selecting the number of splines and knots (Ding and Nan, 2011) or already use a one-step procedure (Lin and Chen, 2013). We suggest a computationally simple estimator that bears the same asymptotic efficiency.

In the absence of auxiliary information, the covariates Z_i's have no likelihood contribution on F and β. The log-likelihood function of (β, F) is $\ell_n(\beta, F)$ in (2.6). We propose a two-step profile procedure to maximize $\ell_n(\beta, F)$. As demonstrated by Vardi (1989), for fixed β, to maximize $\ell_n(\beta, F)$ with respect to F, it is sufficient to consider discrete distribution functions with support points $e_j(\beta)$'s, namely, $F(x \mid \beta) = \sum_{j=1}^n p_j(\beta) I(e_j(\beta) \le x)$, where

$$p_j(\beta) \ge 0, \; j = 1, \dots, n, \qquad \sum_{j=1}^n p_j(\beta) = 1. \tag{2.10}$$

The complete-data log-likelihood is $\sum_{j=1}^n \log\{dF(\log(T_j) - Z_j^\top \beta \mid \beta)\} = \sum_{j=1}^n [\delta_j \log\{ p_j(\beta)\} + (1-\delta_j) \sum_{i=1}^n I\{\log(T_j) - Z_j^\top \beta = e_i(\beta)\} \log\{p_i(\beta)\}]$. We show in the Appendix that, in the E-step of the $(u+1)$th cycle, the conditional expectation of the complete-data log-likelihood given all the observed data is

$$\sum_{j=1}^n w_j^{(u)}(\beta) \log(p_j(\beta)), \tag{2.11}$$

where

$$w_j^{(u)}(\beta) = \delta_j + \sum_{i=1}^n \left\{ (1-\delta_i) \frac{p_j^{(u)}(\beta) I(e_j(\beta) > e_i(\beta))}{\sum_{v=1}^n p_v^{(u)}(\beta) I(e_v(\beta) > e_i(\beta))} \right\} \tag{2.12}$$

and $p_i^{(u)}(\beta)$'s are the values of $p_i(\beta)$'s in the uth cycle. The M-step is to maximize (2.11) under the constraints in (2.10). The maximizer is

$$p_j^{(u+1)}(\beta) = \frac{1}{n} w_j^{(u)}(\beta), \quad j = 1, \ldots, n. \tag{2.13}$$

Denote the final estimate of $p_j(\beta)$ by $\breve{p}_j(\beta)$. Given β, the nonparametric maximum likelihood estimator (MLE) of F is

$$\breve{F}(x \mid \beta) = \sum_{j=1}^{n} \breve{p}_j(\beta) I(e_j(\beta) \leq x). \tag{2.14}$$

Having $\breve{F}(x \mid \beta)$, we may obtain the MLE of β by maximizing $\ell_n(\beta, \breve{F}(x \mid \beta))$ with respect to β. This is challenging as the objective function is nonsmooth. As pointed out by Zeng and Lin (2007), because the objective function $\ell_n(\beta, \breve{F}(\cdot \mid \beta))$ depends only on the ranks of $e_i(\beta)$ and these ranks are stable as β becomes extreme, it can not achieve its maximum for finite β. To bypass this dilemma, we propose to replace $\breve{F}(x \mid \beta)$ in $\ell_n(\beta, \breve{F}(x \mid \beta))$ by a smoothed version of it (Horowitz, 1992), i.e.

$$\tilde{F}(x \mid \beta) = \sum_{j=1}^{n} \tilde{p}_j(\beta) K((x - e_j(\beta))/\sigma), \tag{2.15}$$

where $\tilde{p}_j(\beta)$ is the $\breve{p}_j(\beta)$ with the indicator function $I(s < t)$ in (2.12) replaced by $K((t - s)/\sigma)$. Here $K(\cdot)$ is a smooth distribution function, e.g. the standard normal distribution function, and $\sigma > 0$ is a smoothing parameter. Following Zeng and Lin

Algorithm 1 Calculation of the estimator $(\tilde{\beta}, \tilde{F})$

Input: $p_j^{(0)}(\beta) = 1/n$ for any given β, $j = 1, \ldots, n$.

Step 1 Given β, do the following.

Step 1(a) Set $p_j^{(0)}(\beta) = 1/n, j = 1, \ldots, n$ and calculate $w_j^{(0)}(\beta)$ through (2.12).

Step 1(b) For $u \geq 0$, calculate $w_j^{(u+1)}(\beta)$ and $p_j^{(u+1)}(\beta)$ through (2.12) and (2.13), respectively. Repeat this process until $p_j^{(u+1)}(\beta)$ converges, and write the final quantity as $\breve{p}_j(\beta)$.

Step 1(c) Calculate $\tilde{F}(x \mid \beta)$ in (2.15) and return $\tilde{F}(x \mid \beta)$.

Step 2 Calculate $\tilde{\beta} = \text{argmax}_\beta \ell_n(\beta, \tilde{F}(\cdot \mid \beta))$ and $\tilde{F}(\cdot) = \tilde{F}(\cdot \mid \tilde{\beta})$.

Output: Return $(\tilde{\beta}, \tilde{F})$.

(2007), we recommend the use of $\sigma = csn^{-1/3}$, where $c > 0$ is a constant and s is the sample standard deviation of $\log(X) - Z^\top \beta$ (with β replaced by an initial parameter value) among all subjects. Let $\tilde{\beta} = \arg\max_\beta \ell_n(\beta, \tilde{F}(\cdot \mid \beta))$ and $\tilde{F}(\cdot) = \tilde{F}(\cdot \mid \tilde{\beta})$. Our procedure of calculating $(\tilde{\beta}, \tilde{F})$ can be summarized as Algorithm 1.

As only Step 1(b) involves iteration and Steps 1 and 2 need not be repeated, Algorithm 1 is computationally efficient. Once $\tilde{F}(\cdot \mid \beta)$ and $\tilde{\beta}$ are obtained, it is convenient to obtain estimates for other parameters. For example, given β, smoothed estimators of the density function $f(t)$ and its derivative $\dot{f}(t)$ are $\tilde{f}(t \mid \beta) = \sum_{j=1}^n \tilde{p}_j(\beta)\dot{K}((t - e_j(\beta))/\sigma)/\sigma$ and $\dot{\tilde{f}}(t \mid \beta) = \sum_{j=1}^n \tilde{p}_j(\beta)\ddot{K}((t - e_j(\beta))/\sigma)/\sigma^2$, where $\dot{K}(\cdot)$ and $\ddot{K}(\cdot)$ are the first two derivatives of $K(\cdot)$. Accordingly, given β, smoothed maximum conditional likelihood estimators of the hazard function $\lambda(t)$ and its derivative $\dot{\lambda}(t)$ are $\tilde{\lambda}(t \mid \beta) = \tilde{f}(t \mid \beta)/\{1 - \tilde{F}(t \mid \beta)\}$ and $\dot{\tilde{\lambda}}(t \mid \beta) = \dot{\tilde{f}}(t \mid \beta)/\{1 - \tilde{F}(t \mid \beta)\} + \{\tilde{f}(t \mid \beta)\}^2/\{1 - \tilde{F}(t \mid \beta)\}^2$, respectively. These estimators have closed forms and are readily used for various purposes including variance estimation later on. For example, an efficient estimator and its variance depend on $\lambda(t)$ and $\dot{\lambda}(t)$ (Lin and Chen, 2013), which can be estimated by $\tilde{\lambda}(t \mid \tilde{\beta})$ and $\tilde{\dot{\lambda}}(t \mid \tilde{\beta})$, respectively.

Our simulation results demonstrate that the proposed estimator in the absence of auxiliary aggregate information is insensitive to the tuning parameter σ, and has negligible biases and smaller standard deviations than Zeng and Lin (2007)'s estimator and Lin and Chen (2013)'s estimator. In terms of computational efficiency, Lin and Chen (2013)'s estimator is the best, our estimator is slightly inferior and both of them are more efficient than Zeng and Lin (2007)'s estimator. See Section S6.2 of the Supplementary Material. When the sample size is large, the computational efficiency of Lin and Chen (2013) is evident, and we recommend taking it as an initial estimator in the proposed method the presence of auxiliary information.

2.3 One-step estimator

The proposed one-step estimator relies on the first-order approximation of $(\hat{\beta}_{\mathrm{au}}, \hat{\rho}_{\mathrm{au}})$. We use $g(t) = \log\{\lambda(t)\} = \log[\{dF(t)/dt\}/\{1 - F(t)\}]$ to represent a distribution $F(t)$. For convenience, we shall exchangeably use g and F as an argument and denote $O_i = (X_i, \delta_i, Z_i)$. For theoretical development, following Ding and Nan (2011), we express the log-likelihood in terms of g. The log conditional likelihood is written as $\ell_n(\beta, g) = \sum_{i=1}^n l(\beta, g; O_i)$, where $l(\beta, g; O_i) = \int g(t) dN_i(t, \beta) - \int Y_i(t, \beta) \exp\{g(t)\} dt$ with $N_i(t, \beta) = \delta_i I(\log(X_i) - Z_i^\top \beta \le t)$ and $Y_i(t, \beta) = I(\log(X_i) - Z_i^\top \beta \ge t)$.

The log-likelihood function (2.9) becomes $L_n(\beta,\rho,\nu,g) = \ell_n(\beta,g) - \sum_{i=1}^n \log(1 + \nu^\top \Psi(\beta,\rho,g;Z_i))$. The maximum likelihood estimator is $(\hat\beta_{\rm au},\hat\rho_{\rm au},\hat g_{\rm au}) = \text{argmax}_{\beta,\rho,g} \min_\nu L_n(\beta,\rho,\nu,g)$.

Write the derivative of $l(\beta,g;O)$ with respect to β as $\dot l_\beta(\beta,g;O)$. The functional derivative of $l(\beta,g;O)$ with respect to g along the direction h is denoted as $\dot l_g(\beta,g;O)[h]$. We show in the Supplementary Material that

$$\dot l_\beta(\beta,g;O) = -Z\Big\{\int \dot g(t)dN(t,\beta) - \int Y(t,\beta)\exp\{g(t)\}\dot g(t)dt\Big\},$$
$$\dot l_g(\beta,g;O)[h] = \int h(t)dN(t,\beta) - \int Y(t,\beta)\exp\{g(t)\}h(t)dt,$$

where $\dot g(\cdot)$ is the first derivative of $g(\cdot)$. Similarly, $\Psi_j(\beta,\rho,F;Z)$ summarizing auxiliary information can be rewritten as $\Psi_j(\beta,\rho,g;Z) = I(Z\in\Omega_j)\Big[\exp\Big\{-\int I(s \le \log t_j^* - Z^\top\beta)e^{g(s)}ds\Big\} - \zeta_j^{1/\rho}\Big]$, $j = 1,\ldots,J$. The first order ordinary and functional derivatives of $\Psi(\beta,\rho,g;Z) = (\Psi_1(\beta,\rho,g;Z),\ldots,\Psi_J(\beta,\rho,g;Z))^\top$ are written as

$$\dot\Psi_\beta(\beta,\rho,g;Z) = (\dot\Psi_{1,\beta}(\beta,\rho,g;Z),\ldots,\dot\Psi_{J,\beta}(\beta,\rho,g;Z)),$$
$$\dot\Psi_\rho(\beta,\rho,g;Z) = (\dot\Psi_{1,\rho}(\beta,\rho,g;Z),\ldots,\dot\Psi_{J,\rho}(\beta,\rho,g;Z))^\top,$$
$$\dot\Psi_g(\beta,\rho,g;Z)[h] = (\dot\Psi_{1,g}(\beta,\rho,g;Z)[h],\ldots,\dot\Psi_{J,g}(\beta,\rho,g;Z)[h])^\top,$$

where $\dot\Psi_{j,\beta}(\beta,\rho,g;Z) = I(Z\in\Omega_j)ZS(\log t_j^* - Z^\top\beta)\exp\{g(\log t_j^* - Z^\top\beta)\}$, $\dot\Psi_{j,\rho}(\beta,\rho,g;Z) = I(Z\in\Omega_j)\rho^{-2}\zeta_j^{1/\rho}\ln(\zeta_j)$, and $\dot\Psi_{j,g}(\beta,\rho,g;Z)[h] = -I(Z\in\Omega_j)S(\log t_j^* - Z^\top\beta)\int I(s \le$

$\log t_j^* - Z^\top\beta)e^{g(s)}h(s)ds$ with $S(\cdot) = 1 - F(\cdot)$ and $j = 1, \ldots, J$. More detailed derivations are given in the Supplementary Material, where we also adapt the current data setting to notations in empirical process and develop theoretical results. Define $h_1^*(t, \beta_0) = -\dot{g}_0(t)E\{ZY(t, \beta_0)\}/E\{Y(t, \beta_0)\}$, where $\dot{g}(\cdot)$ is the first order derivative of $g(\cdot)$, and $h_2^*(t, \beta) = (h_{21}^*(t, \beta), \ldots, h_{2J}^*(t, \beta))^\top$ with

$$h_{2j}^*(t, \beta_0) = \frac{E[I(Z \in \Omega_j)I(t \le \log t_j^* - Z^\top\beta_0)\{1 - F(\log t_j^* - Z^\top\beta_0)\}]}{E\{Y(t, \beta_0)\}}, \quad j = 1, \ldots, J.$$

The two vector-valued functions are the so-called least favorable directions in semiparametric likelihood theory. Define $A = -E\big\{\dot{\Psi}_\rho(\beta_0, \rho_0, g_0; Z)\big\}$, $B = E\big\{\dot{\Psi}_\beta(\beta_0, \rho_0, g_0; Z) - \{\dot{\Psi}_g(\beta_0, \rho_0, g_0; Z)[h_1^*]\}^\top\big\}$, $\chi(\beta_0, \rho_0, g_0; O) = \Psi(\beta_0, \rho_0, g_0; Z) - \dot{l}_g(\beta_0, g_0; O)[h_2^*]$, $\iota(\beta_0, g_0; O) = \dot{l}_\beta(\beta_0, g_0; O) - \dot{l}_g(\beta_0, g_0; O)[h_1^*]$,

$$\Sigma = E\big\{\iota(\beta_0, g_0; O)^{\otimes 2}\big\}, \tag{2.16}$$

and $Q = E\big\{\chi(\beta_0, \rho_0, g_0; O)^{\otimes 2}\big\}$, where $U^{\otimes 2} = UU^\top$ for any vector U.

Let $(\tilde{\beta}, \tilde{F}) = \text{argmax}_{\beta, F}\ell_n(\beta, F)$ and $\tilde{\rho}$ be the solution of $\sum_{i=1}^n \sum_{j=1}^J \Psi_j(\tilde{\beta}, \rho, \tilde{F}; Z_i) = 0$. Our one-step estimator of β is constructed based on $(\tilde{\beta}, \tilde{\rho})$ and first-order linear approximations of $(\hat{\beta}_{\text{au}}, \hat{\rho}_{\text{au}})$ and $(\tilde{\beta}, \tilde{\rho})$. It can be verified (See Section S4 in the

Supplementary Material) that

$$\begin{pmatrix} \hat{\beta}_{\mathrm{au}} - \beta_0 \\ \hat{\rho}_{\mathrm{au}} - \rho_0 \end{pmatrix} = \begin{pmatrix} I_{p\times p} & 0_{p\times J} & 0_{p\times 1} \\ 0_{1\times p} & 0_{1\times J} & 1 \end{pmatrix} \times \begin{pmatrix} \Sigma & B & 0 \\ -B^\top & Q & A \\ 0 & -A^\top & 0 \end{pmatrix}^{-1} \begin{pmatrix} n^{-1}\sum_{i=1}^n \iota(\beta_0, g_0; O_i) \\ n^{-1}\sum_{i=1}^n \chi(\beta_0, \rho_0, g_0; O_i) \\ 0 \end{pmatrix} + o_p(n^{-1/2}), \tag{2.17}$$

and

$$\begin{pmatrix} \tilde{\beta} - \beta_0 \\ \tilde{\rho} - \rho_0 \end{pmatrix} = \begin{pmatrix} I_{p\times p} & \Sigma^{-1}B & 0_{p\times 1} \\ 0_{1\times p} & (1_{1\times J}A)^{-1}1_{1\times J}(B^\top\Sigma^{-1}B + Q) & 1 \end{pmatrix} \times \begin{pmatrix} \Sigma & B & 0 \\ -B^\top & Q & A \\ 0 & -A^\top & 0 \end{pmatrix}^{-1} \begin{pmatrix} n^{-1}\sum_{i=1}^n \iota(\beta_0, g_0; O_i) \\ n^{-1}\sum_{i=1}^n \chi(\beta_0, \rho_0, g_0; O_i) \\ 0 \end{pmatrix} + o_p(n^{-1/2}), \tag{2.18}$$

where $I_{p\times p}$ is the identity matrix of size p, $0_{m\times n}$ is a $m \times n$ dimensional matrix with zeros, and $1_{m\times n}$ is a $m \times n$ dimensional matrix with ones.

The approximations in (2.17) and (2.18) motivate us to construct a one-step estimator of (β, ρ) as

$$\begin{pmatrix} \hat{\beta}_{\text{os}} \\ \hat{\rho}_{\text{os}} \end{pmatrix} = \begin{pmatrix} \tilde{\beta} \\ \tilde{\rho} \end{pmatrix} + \begin{pmatrix} 0_{p\times p} & -\hat{\Sigma}^{-1}\hat{B} & 0_{p\times 1} \\ 0_{1\times p} & -(1_{1\times J}\hat{A})^{-1}1_{1\times J}(\hat{B}^\top\hat{\Sigma}^{-1}\hat{B}+\hat{Q}) & 0 \end{pmatrix}$$
$$\times \begin{pmatrix} \hat{\Sigma} & \hat{B} & 0 \\ -\hat{B}^\top & \hat{Q} & \hat{A} \\ 0 & -\hat{A}^\top & 0 \end{pmatrix}^{-1} \begin{pmatrix} n^{-1}\sum_{i=1}^n \iota(\tilde{\beta}, \tilde{g}; O_i) \\ n^{-1}\sum_{i=1}^n \chi(\tilde{\beta}, \tilde{\rho}, \tilde{g}; O_i) \\ 0 \end{pmatrix}, \quad (2.19)$$

where

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^n \left\{ \dot{l}_\beta(\tilde{\beta}, \tilde{g}; O_i) - \dot{l}_g(\tilde{\beta}, \tilde{g}; O_i)[\tilde{h}_1^*] \right\}^{\otimes 2}, \quad \hat{Q} = \frac{1}{n}\sum_{i=1}^n \left\{ \Psi(\tilde{\beta}, \tilde{\rho}, \tilde{g}; Z_i) - \dot{l}_g(\tilde{\beta}, \tilde{g}; O_i)[\tilde{h}_2^*] \right\}^{\otimes 2},$$
$$\hat{B} = \frac{1}{n}\sum_{i=1}^n \left\{ \dot{\Psi}_\beta(\tilde{\beta}, \tilde{\rho}, \tilde{g}; O_i) - \{\dot{\Psi}_g(\tilde{\beta}, \tilde{\rho}, \tilde{g}; O_i)[\tilde{h}_1^*]\}^\top \right\}, \quad \tilde{A} = -\frac{1}{n}\sum_{i=1}^n \Psi_\rho(\tilde{\beta}, \tilde{\rho}, \tilde{g}; Z_i),$$

with $\tilde{g}(\cdot) = \log\{\tilde{\lambda}(\cdot)\}$,

$$\tilde{h}_1^*(t, \tilde{\beta}) = -\frac{\dot{\tilde{\lambda}}(t)}{\tilde{\lambda}(t)} \frac{\sum_{i=1}^n Z_i I(\log(X_i) - Z_i^\top\tilde{\beta} \geq t)}{\sum_{i=1}^n I(\log(X_i) - Z_i^\top\tilde{\beta} \geq t)}, \quad \tilde{h}_2^*(t, \tilde{\beta}) = (\tilde{h}_{21}^*(t, \tilde{\beta}), \ldots, \tilde{h}_{2J}^*(t, \tilde{\beta}))^\top,$$
$$\tilde{h}_{2j}^*(t, \tilde{\beta}) = \frac{\sum_{i=1}^n I(Z_i \in \Omega_j) I(t \leq \log t_j^* - Z_i^\top\tilde{\beta})\{1 - \tilde{F}(\log t_j^* - Z_i^\top\tilde{\beta})\}}{\sum_{i=1}^n I(\log(X_i) - Z_i^\top\tilde{\beta} \geq t)}, \quad j = 1, \ldots, J.$$

An important case is the homogeneous case where $\rho_0 = 1$. In this situation, we omit ρ in the previous notations. For example, we write $\Psi(\beta, \rho, g; Z)$ and $\chi(\beta, \rho, g; O)$ as $\Psi(\beta, g; Z)$ and $\chi(\beta, g; O)$, respectively. The proposed one-step estimator of β

becomes

$$\hat{\beta}_{\text{os}} = \tilde{\beta} + \begin{pmatrix} 0_{p\times p} & -\hat{\Sigma}^{-1}\hat{B} \end{pmatrix} \begin{pmatrix} \hat{\Sigma} & \hat{B} \\ -\hat{B}^\top & \hat{Q} \end{pmatrix}^{-1} \begin{pmatrix} n^{-1}\sum_{i=1}^n \iota(\tilde{\beta}, \tilde{g}; O_i) \\ n^{-1}\sum_{i=1}^n \chi(\tilde{\beta}, \tilde{g}; O_i) \end{pmatrix},$$

where $\hat{B} = n^{-1}\sum_{i=1}^n \left\{\dot{\Psi}_\beta(\tilde{\beta}, \tilde{g}; O_i) - \{\dot{\Psi}_g(\tilde{\beta}, \tilde{g}; O_i)[\tilde{h}_1^*]\}^\top\right\}$ and $\hat{Q} = n^{-1}\sum_{i=1}^n \left\{\Psi(\tilde{\beta}, \tilde{g}; Z_i) - \dot{l}_g(\tilde{\beta}, \tilde{g}; O_i)[\tilde{h}_2^*]\right\}^{\otimes 2}$. The $\tilde{\beta}$, $\tilde{\rho}$, and the plug-in estimators are easy to calculate. The one-step estimator is also easy to calculate as it has a closed form.

3. Asymptotic results

3.1 Asymptotic properties of $(\tilde{\beta}, \tilde{F})$

A desirable theoretical property of the maximum conditional likelihood estimators $\tilde{\beta}$ and $\tilde{F}(\cdot)$ is that they are jointly asymptotically normal. Let $\lambda_0(t)$, $f_0(t)$, and $S_0(t)$ denote the true hazard, density, and survival functions, respectively, of ϵ and β_0 is the true value of β. Our asymptotic normality results are established under the following regularity conditions:

(C1) The range $\mathcal{Z} \subset \mathbb{R}^p$ of Z is compact.

(C2) (a) The support of C contains that of T, and (b) there exists two positive constants $\tau_l < \tau_u$ such that $\text{pr}(\log(X) - Z^\top\beta_0 \geq \tau_l) = 1$ and $\text{pr}(\log(T) - Z^\top\beta_0 > \tau_u) > 0$.

(C3) β_0 is an interior point of a compact set $\mathcal{B} \subset \mathbb{R}^p$.

(C4) The conditional density function of C given Z is uniformly bounded away from zero and infinity.

(C5) If their exists a constant vector υ and a deterministic function $\varrho(\cdot)$ such that $Z^\top \upsilon = \varrho(\epsilon)$ with probability 1, then $\upsilon = 0$ and $\varrho = 0$.

(C6) The information matrices $-\partial^2 E\{\ell_n(\beta, \tilde{F}(\cdot \mid \beta))\}/\partial\beta\partial\beta^\top|_{\beta=\beta_0}$ and $J_0 = E[ZZ^\top \int_{\tau_l}^{\tau_u} Y(t, \beta_0)$ $\exp\{g_0(t)\}\{\dot{g}_0(t)\}^2 dt]$ are positive definite, where $Y(t, \beta) = I(\log(X) - Z^\top\beta \geq t)$, $g_0(t) = \log \lambda_0(t)$ and $\dot{g}_0(t)$ is its derivative.

(C7) Let $\mathcal{G}$ denote the collection of twice differentiable functions on $[\tau_l, \tau_u]$, and for every $g \in \mathcal{G}$, g, its first derivative $\dot{g}$ and second derivative $\ddot{g}$ have bounded variations. The true log hazard function $g_0(\cdot) = \log\{\lambda_0(\cdot)\}$ belongs to $\mathcal{G}$.

(C8) (a) $K(x)$ is a thrice-continuously differentiable distribution function, and $K^{(r)}(\cdot)$, the r-th derivative of $K(\cdot)$, has bounded variations on $(-\infty, \infty)$ for $r = 1, 2, 3$. (b) the smoothing parameter $\sigma = \sigma_n$ satisfies $|K(t/\sigma_n) - I(t > 0)| = o(n^{-1/2})$.

Conditions (C1), (C3) and (C4) are commonly imposed regularity conditions in the literature of censored linear regressions. We adopt Condition (C2)(a) to avoid a lengthy technical discussion on the tail behavior of the failure time. Condition (C2)(b) implies that F_0 belongs to $\mathcal{F} = \{F : [\tau_l, \tau_u] \mapsto [0, 1], F$ is non-decreasing, $F(\tau_l)$ $= 0,\ F(\tau_u) < 1\}$. Condition (C5), also adopted by Zeng and Lin (2007), guarantees the identifiability of β. When Condition (C6) is satisfied, the objective function $E\{\ell_n(\beta, \tilde{F}(\cdot \mid \beta))\}$ is strictly concave in a neighborhood of β_0, which implies that the maximum conditional likelihood estimator $\tilde{\beta}$ is unique with probability tending to one; Qin et al. (2011) and Huang et al. (2015) made similar assumptions. Under Condition (C7), $\lambda_0(t)$ and $f_0(t)$ are differentiable. Let $\dot{\lambda}_0(t)$ and $\dot{f}_0(t)$ be first derivatives of $\lambda_0(t)$ and $f_0(t)$, respectively. Condition (C7) implies that $\dot{\lambda}_0(t)$ and $\dot{f}_0(t)$ are bounded, and that $\int_{\tau_l}^{\tau_u}\{\dot{f}_0(t)/f_0(t)\}^2 f_0(t)dt < \infty$, which are commonly-imposed conditions in censored linear regressions. The function $K(\cdot)$ in Condition (C8) is similar to the primitive functions of the kernel functions used in Zeng and Lin (2007). Condition (C8)(b) guarantees that the difference between $\breve{F}(x \mid \beta)$ and $\tilde{F}(x \mid \beta)$ is $o_p(n^{-1/2})$, so that they have the same limiting distribution.

Theorem 1. *Under conditions (C1)–(C8), the following results hold as* $n \to \infty$*:* *(i)* $\|\tilde{\beta} - \beta_0\| + \sup_{t\in[\tau_l,\tau_u]} |\tilde{F}(t) - F_0(t)| = o_p(1)$*, which implies* $(\tilde{\beta}, \tilde{F})$ *is consistent.* *(ii)* $\sqrt{n}(\tilde{\beta} - \beta_0, \tilde{F} - F_0)$ *converges weakly to a tight mean zero Gaussian process* $\psi_0'\{-\dot{U}_0^{-1}(\mathbb{W})\}$*, where* ψ_0'*,* $\dot{U}_0^{-1}$*, and* $\mathbb{W}$ *are defined in the Supplementary Material;* *(iii)* $\sqrt{n}(\tilde{\beta} - \beta_0)$ *converges in distribution to a p-variate normal distribution with mean* 0 *and variance* Σ^{-1}*, where* Σ *is defined in* (2.16)*, provided* Σ *is non-singular.*

Theorem 1 indicates that $(\tilde{\beta}, \tilde{F})$ is consistent and asymptotically normal. In particular, we find that the asymptotic variance of $\tilde{\beta}$ is equal to the semiparametric efficiency lower bound of β when no auxiliary information is available. See Ritov and Wellner (1988) and Bickel et al. (1989). In other words, the maximum conditional likelihood estimator $\tilde{\beta}$ is asymptotically semiparametric efficient.

3.2 Asymptotic properties of $\hat{\beta}_{\text{os}}$

To investigate the consistency and asymptotic normality of $\hat{\beta}_{\text{os}}$, we make the following additional assumptions.

(D1) Model (2.3) is correctly specified with $\rho_0 > 0$ being the true value of ρ.

(D2) (i) $\Psi(\beta, \rho, F; Z)$ is smooth enough with respect to (β, ρ, F). (ii)There exists a function $K(Z)$ such that $\|\Psi(\beta, \rho, F; Z)\| \leq K(Z)$ and $\mathbb{E}\{K^3(Z)\} < \infty$ for all $\beta \in \mathcal{B}$, $\rho > 0$ and $F \in \mathcal{F}$. (iii) The matrix $E\{\Psi(\beta_0, \rho_0, F_0; Z)^{\otimes 2}\}$ is positive definite.

Theorem 2. *Suppose that conditions (C1)–(C8) and (D1)–(D2) are satisfied. As $n \to \infty$, (i) $\hat{\beta}_{\text{os}}$ is consistent to β_0, and (ii) $n^{1/2}(\hat{\beta}_{\text{os}} - \beta_0)$ converges to a normal distribution with mean zero and covariance $\{\Sigma + BQ^{-1}B^\top - BQ^{-1}A(A^\top Q^{-1}A)^{-1}A^\top Q^{-1}B^\top\}^{-1}$, provided the matrices Σ and $A^\top Q^{-1}A$ are nonsingular. (iii) $\hat{\beta}_{\text{os}}$ is asymptotically more efficient than $\tilde{\beta}$.*

Theorem 2 discloses that the one-step estimator $\hat{\beta}_{\text{os}}$ is asymptotically more efficient than the maximum conditional likelihood $\tilde{\beta}$. In the homogeneous case where $\rho_0 = 1$ is known, the asymptotic covariance of $\hat{\beta}_{\text{os}}$ is $(\Sigma + BQ^{-1}B^\top)^{-1}$, which is no greater than that in the heterogeneous case. A possible interpretation for this finding is that in the homogeneous case, $\rho = 1$ is known and is automatically taken as auxiliary information in the estimation of β. However in the heterogeneous case, ρ is unknown and has to be estimated from data, which brings additional variability.

The proposed maximum likelihood estimator and one-step estimator still work when there exists heterogeneity in the covariate distributions or uncertainty in the auxiliary information. Similar to Theorems 1 and 2, the general theoretical results of our methods still hold when the variability in the aggregated information and heterogeneity in the covariate distributions are taken into consideration. See Section S7 of the Supplementary Material.

According to Huang et al. (2016), when the subgroup survival information is determined only by a subset of covariates, the efficiency gain of the estimated coefficients for other covariates is minimal. Consider a simple case where $Z = (Z_1, Z_2)$ and the subgroups are determined only by Z_1. Auxiliary survival information for the jth subgroup can be expressed as

$$
\begin{aligned}
& E\{\Psi_j(\beta,\rho,g;Z)\} \\
&= \iint I(z_1 \in \Omega_j)\Big[\exp\Big\{-\int I(s \le \log(t_j^*) - z_1\beta_1 - z_2\beta_2)e^{g(s)}\mathrm{d}s\Big\} - \zeta_j^{1/\rho}\Big]\mathrm{d}F_z(z_1,z_2) \\
&= \iint I(z_1 \in \Omega_j)\Big[\exp\Big\{-\int I(s \le \log(t_j^*) - z_1\beta_1 - z_2)e^{g(s)}\mathrm{d}s\Big\} - \zeta_j^{1/\rho}\Big]\mathrm{d}F_z(z_1,z_2/\beta_2).
\end{aligned}
$$

Note that the proposed estimation procedure incorporates the auxiliary information in the marginal likelihood and allows for an arbitrary distribution function F_z of (Z_1, Z_2). After parameterization, our method is equivalent to maximizing the log marginal likelihood $\sum_{i=1}^n \log(p_i^*)$ subject to the constraints $p_i^* \geq 0, \sum_{i=1}^n p_i^* = 1$, and

$$\sum_{i=1}^n p_i^* I(Z_{i1} \in \Omega_j)\Big[\exp\Big\{-\int I(s \leq \log(t_j^*) - Z_{i1}\beta_1 - Z_{i2})e^{g(s)}\mathrm{d}s\Big\} - \zeta_j^{1/\rho}\Big] = 0,$$

where p_i^* is the jump of F_z at $(Z_{i1}, Z_{i2}/\beta_2)$. When we profile out p_i^*, the marginal likelihood becomes $-\sum_{i=1}^n \log\{1+\nu^\top[\exp\{-\int I(s \leq \log(t_j^*) - Z_{i1}\beta_1 - Z_{i2})e^{g(s)}\mathrm{d}s\} - \zeta_j^{1/\rho}]\}$, which does not involve β_2. Therefore, our estimator for β_2 has limited efficacy gain by incorporating the subgroup survival information.

4. Simulation Studies

In this section, we conduct simulations to investigate the finite sample performance of the proposed estimators. The survival time T in the target population is generated from the accelerated failure time model $\log(T) = Z_1\beta_{10} + Z_2\beta_{20} + Z_1Z_2\beta_{30} + \epsilon$, where Z_1 is generated from $N(0, 1.5^2)$, the normal distribution with mean 0 and standard deviation 1.5, and Z_2 is generated from Bernoulli(0.5), the Bernoulli distribution with success probability 0.5. Five distributions are considered for the error term ϵ, a normal distribution, a generalized extreme value distribution, a Weibull distribution, a log-normal distribution, and a chi-squared distribution. The censoring time C is generated from an exponential distribution so that the censoring rate is approximately 20%. We consider both the heterogeneous scenario ($\rho_0 = 0.9$ and unknown) and the homogeneous scenario ($\rho_0 = 1$ and known). Suppose that the t_*–year survival probabilities for the two subgroups $\Omega_1 = \{(Z_1, Z_2) : Z_1 \leq 0, Z_2 = 0\}$

and $\Omega_2 = \{(Z_1, Z_2) : Z_1 > 0, Z_2 = 1\}$, denoted by ζ_1 and ζ_2, are known. We set $\beta_{10} = -1$ and $\beta_{20} = 1.5$. Our specific simulation settings are as follows.

Case I: $\epsilon \sim N(0, 0.5^2)$. We set $t_* = 2$, $\beta_{30} = 0.75$, $\zeta_1 = 0.671$ and $\zeta_2 = 0.838$ when $\rho_0 = 0.9$, and $\zeta_1 = 0.380$ and $\zeta_2 = 0.371$ when $\rho_0 = 1$.

Case II: $\epsilon \sim$ the generalized extreme value distribution with location 0, scale 0.5, and shape 0.1. We set $t_* = 4$, $\beta_{30} = 0.75$. $\zeta_1 = 0.515$ and $\zeta_2 = 0.548$ when $\rho_0 = 0.9$, and $\zeta_1 = 0.491$, $\zeta_2 = 0.514$ when $\rho_0 = 1$.

Case III: $\epsilon \sim$ the Weibull distribution with shape 1.3 and scale 0.8. We set $t_* = 4$, $\beta_{30} = 0.5$. $\zeta_1 = 0.668$ and $\zeta_2 = 0.660$ when $\rho_0 = 0.9$, and $\zeta_1 = 0.649$ and $\zeta_2 = 0.638$ when $\rho_0 = 1$.

Case IV: $\epsilon \sim$ the log-normal distribution with mean -0.5 and standard deviation 0.9 in the log scale. We set $t_* = 4$, $\beta_{30} = 0.25$. $\zeta_1 = 0.684$ and $\zeta_2 = 0.548$ when $\rho_0 = 0.9$, and $\zeta_1 = 0.665$ and $\zeta_2 = 0.523$ when $\rho_0 = 1$.

Case V: $\epsilon \sim$ the chi-squared distribution with 1 degree of freedom. We set $t_* = 4$, $\beta_{30} = 0.75$. $\zeta_1 = 0.654$ and $\zeta_2 = 0.752$ when $\rho_0 = 0.9$, and $\zeta_1 = 0.632$ and $\zeta_2 = 0.731$ when $\rho_0 = 1$.

In each case, we generate 1000 random samples of sample size $n = 100$. Based on each random sample, we calculate the proposed maximum conditional likelihood estimator $\tilde{\beta}$ without incorporating auxiliary information and the proposed one-step estimator $\hat{\beta}_{\mathrm{os}}$, which incorporates auxiliary information. We use $\hat{\beta}_{\mathrm{os,homo}}$ and $\hat{\beta}_{\mathrm{os,hete}}$ to denote $\hat{\beta}_{\mathrm{os}}$ in the homogeneous and heterogeneous scenarios, respectively. In our method, the kernel function K is chosen to be the standard normal distribution function, and the smoothing parameter σ is set to $csn^{-1/3}$, where c is a constant, s is the sample standard deviation of $\log(X) - Z^{\top}\tilde{\beta}$ among all subjects. For comparison, we also take into consideration the weighted log-rank estimator ($\hat{\beta}_{\mathrm{L}}$) with unit weight and the generalized method of moments estimators in the presence ($\hat{\beta}_{\mathrm{G1}}$) and absence ($\hat{\beta}_{\mathrm{G0}}$) of auxiliary information (Sheng et al., 2020). The estimator $\hat{\beta}_{\mathrm{L}}$ is calculated using the function `aftsrr()` in the R package `aftgee`.

For a generic point estimator of β, we calculate its empirical bias (Bias), empirical standard deviation (SD), and the average of its standard error (SE) estimates. Table 1 presents 1000 times the Bias, SD and SE of the six estimators under comparison in the heterogeneous scenario, and the corresponding coverage probability (CP) of the corresponding Wald-type confidence region.

First of all, the proposed estimators ($\hat{\beta}_{\mathrm{os}}$ and $\tilde{\beta}$) and $\hat{\beta}_{\mathrm{L}}$ have negligible biases, whereas Sheng et al. (2020)'s estimators ($\hat{\beta}_{\mathrm{G1}}$ and $\hat{\beta}_{\mathrm{G0}}$) usually have much larger biases. The one-step estimators $\hat{\beta}_{\mathrm{os,homo}}$ and $\hat{\beta}_{\mathrm{os,hete}}$ exhibit nearly the same good finite-sample performance. This is probably because the scenario $\rho = 0.9$ here is close to the homogeneous setting. Second, compared with Sheng et al. (2020)'s estimators, our estimators usually have much smaller SDs in Cases III, IV, and V, although their SDs are comparable in Cases I and II. The estimator $\hat{\beta}_{\mathrm{L}}$ often has too much fluctuation, as its SDs corresponding to β_2 and β_3 are much larger than the other four estimators in all cases except Case I. Third, the SDs and SEs of our estimators usually match well and their CPs are close to the nominal level 95% in most cases. However, those of Sheng et al. (2020)'s estimators do not match well. This together with their big biases makes their CPs much less than 95%. For example, in Case I, the CPs corresponding to $\hat{\beta}_{\text{os}}$ and $\tilde{\beta}$ are around 95%, however those corresponding to $\hat{\beta}_{\text{G1}}$ and $\hat{\beta}_{\text{G0}}$ are only 81% or 82%, which are far from the nominal level. It is worth mentioning that in the implementation of Sheng et al. (2020)'s method in our simulation study, nonsingular matrices arise quite a few times.

Regarding the proposed estimators, they all have negligible biases, and $\hat{\beta}_{\text{os}}$ has smaller SDs than $\tilde{\beta}$ in most cases. In particular, in Cases I and II, the SD reductions of $\hat{\beta}_{\text{os}}$ against $\tilde{\beta}$ corresponding to β_2 can be as large as 14%. This can be regarded as the efficiency gain of incorporating auxiliary information.

Table 2 displays the simulation results of the proposed one-step estimator and the extended generalized method of moments (GMM) estimator (Sheng et al., 2020) for ρ. Compared with the GMM estimator, the proposed estimator has clearly more favorable performance: it still has negligible biases and smaller SDs in all cases, and its SEs are generally very close to its empirical SDs, making the corresponding Wald-type interval often have close-to-nominal CPs. However, the GMM estimator has big biases and much larger SDs, and its empirical SDs and SEs do not match well, leading to under-coverages of the corresponding Wald-type interval.

The simulation results under the homogeneous scenario are shown in Table S1 in the Supplementary Material. Our general findings are the same as those from Table 1. Comparing the results of the proposed estimators in these two tables, we find that the estimated standard errors in the heterogeneous case are generally a little bit larger than those in the homogeneous case. This is in accordance with our theoretical findings. We have also conducted additional simulations to study the proposed estimators in terms of computational efficiency, estimation efficiency, bias and variance trade-off and robustness to the nonconstancy of ρ. The simulation results are generally favorable to our methods. See Section S6 in the Supplementary Material.

In summary, the proposed estimators are nearly unbiased and are often more reliable than popular competitors such as $\hat{\beta}_{\mathrm{L}}$ and Sheng et al. (2020)'s estimators ($\hat{\beta}_{\mathrm{G1}}$ and $\hat{\beta}_{\mathrm{G0}}$). The corresponding Wald-type confidence intervals are often more accurate than their competitors. More importantly, the performance of our method can indeed be improved by incorporating auxiliary information.

5. A Real Data Example

In this section, we apply the proposed estimation methods to analyze real data from a chemotherapy study for Stage III colon cancer. The data set is available in the R package `survival`. In the study, patients diagnosed with stage III colon cancer were enrolled between March 1984 and October 1987. There were 929 subjects in the study, and the subjects were randomized such that 315, 310, and 304 patients received observation (Obs), levamisole alone (Lev), and levamisole combined with fluorouracil (Lev+5FU) treatments, respectively. The patients were followed for up to 9 years for the outcomes of cancer recurrence and death. For the purpose of illustration, following Gao and Chan (2023), we modeled death using the accelerated failure time model and associated the survival time with the treatments, gender, and diagnosis age. The model is $\log(T) = Z_1\beta_1 + Z_2\beta_2 + Z_3\beta_3 + Z_4\beta_4 + \epsilon$, where $Z_1 = 1$ for male and $Z_1 = 0$ for female, Z_2 is diagnosis age, Z_3 is 1 for treatment levamisole alone and 0 otherwise, and $Z_4 = 1$ for treatment levamisole combined with fluorouracil and 0 otherwise. The observed survival time ranged from 0.06 to 9.12 years with mean 4.58 and median 5.41. There were 484 males and 445 females, and the minimum, median, and maximum diagnosis ages were 18, 61, and 85, respectively. The censoring rate was approximately 51.3%.

The Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute collects and publishes cancer incidence and survival data from population-based cancer registries covering approximately 34.6% of the U.S. population. The SEER Cancer Statistics Review reports the most recent cancer incidence, mortality, survival, prevalence, and lifetime risk statistics annually.

We analyze the data from the chemotherapy study combined with the 5-year gender-specific survival information reported in SEER. Based on the SEER Cancer Statistics Review, 1973-1993 (National Cancer Institute, 1997), the 5-year survival rates among regional colon cancer patients who were diagnosed from 1986 to 1992 are 66.7% for males and 66.6% for females (Gao and Chan, 2023). The populations in the chemotherapy study and the SEER Program may be different, however, the conditional effect of the covariates may be more generalizable.

We analyze the data using the six estimators considered in the simulation section, where the proposed one-step estimators for the heterogeneous case and the homogeneous case are both implemented. We choose the smooth parameter σ to be $0.287sn^{-1/3}$, where s is the standard deviation of $\log(X) - Z^\top\hat{\beta}_{\mathrm{L}}$ among all subjects. The analysis results are given in Table 3. The estimates of ρ based on the proposed approach and the generalized method of moments approach are almost the same. Their values together with the accompanying standard errors imply that $\rho \neq 0$ at the 5% significance level, or equivalently, the individual-level data and the auxiliary aggregate data come from different populations. As the standard error of the proposed estimator for ρ is smaller, the Wald-type test based on our estimator produces a smaller p-value than that based on $\hat{\beta}_{\mathrm{G1}}$, namely, our method provides stronger evidence for $\rho \neq 0$. For the estimation of β, the standard errors of the proposed three estimators and the two generalized methods of moments estimators from Sheng et al. (2020) are all smaller than that of the log-rank estimator, indicating efficiency improvements over the latter. Compared with the proposed estimators with and without auxiliary information, the coefficient estimate of gender has accuracy improvement when auxiliary information is incorporated. For the estimation of β_4, in

---

line with the log-rank estimator, the proposed three estimators show that the treatment effect of levamisole combined with fluorouracil has a significantly positive effect on the survival time. The smaller p-values of the proposed estimators reflect obvious efficiency gain in this case. The estimation results of the proposed estimators for the heterogeneous and homogeneous cases are very close to each other, which suggests the robustness of the proposed framework. The proposed estimators for $\beta_1, \beta_2, \beta_3, \beta_4$ have smaller standard errors than the generalized method of moments estimators.

6. Discussion

Our method allows the individual-level data and the external auxiliary information to be incompatible and accommodate the incompatibility by (2.3). We assume that the sample size m used to derive the external information is much larger than that n of the individual-level data and thus variability in the auxiliary information is ignorable. In Section S7 of the Supplementary Material, we extend the current framework to account for heterogeneity in covariate distributions and uncertainty in external aggregate information. We assume the covariate distributions of the two populations satisfy a density ratio model, and take the variability of external information into consideration when m and n are comparable.

For ease of discussion, we assume that all covariates are time-independent although time-dependent covariates are also commonly seen. In the case of no auxiliary information, the proposed expectation-maximization algorithm readily accommodates time-dependent covariates. Specifically, following Zeng and Lin (2007), we consider an extension of model (2.1) with $\epsilon = \ln \int_0^T \exp\{-\beta^\top Z(t)\}dt$, where $Z(\cdot)$ is a vector of time-dependent covariates. Our expectation-maximization algorithm proceeds by re-defining $e_i(\beta) = \log\{\int_0^{X_i} \exp\{-\beta^\top Z_i(t)\}dt\}$ for fixed β.

In the construction of our one-step estimator of β, we take the proposed maximum conditional likelihood estimator $\tilde{\beta}$ as an initial estimator of β. An advantage of the expectation-maximization algorithm is that we can simultaneously obtain $(\tilde{\beta}, \tilde{F})$ (or $(\tilde{\beta}, \tilde{g})$). The consistency of $\tilde{F}$ depends on the assumption that the support of C contains that of T. If this assumption is violated, $\tilde{F}$ is no longer a valid distribution function estimate for ϵ. For example, if the maximum point of the support of C, say τ_1, is smaller than that of T, τ, then the proposed method estimates the conditional distribution function of T given $T \leq \tau_1$, namely, $F(t)/F(\tau_1)$. In this case, we recommend using a different initial estimator such as the sieve estimator of Ding and Nan (2011), or we may estimate g by using the kernel smoothing techniques of Lin and Chen (2013). We leave this problem for future research.

Acknowledgments

This work is partly supported by the National Key R&D Program of China (2021YFA1000100, 2021YFA1000101), the National Natural Science Foundation of China (12471253, 72331005, 12171157), and the Natural Science Foundation of Shanghai (23JS1400500). The authors would like to thank the editor, associate editor, and two reviewers for their insightful comments which have helped improve the manuscript substantially.

Supplementary Material

Supplementary Material includes notations, the asymptotic results of $\hat{\beta}_{\text{au}}$, proofs of Theorems 1-2, the relationship between $(\hat{\beta}_{\text{au}}, \hat{\rho}_{\text{au}})$ and $(\tilde{\beta}, \tilde{\rho})$, additional simulation studies, the heterogeneous case with heterogeneity in covariate distributions and uncertainty in external information.

Appendix

Expectation-Maximization algorithm on the conditional likelihood

Instead of observing $e_{1i}(\beta) = \log(T_i) - Z_i^\top \beta$, we observe $e_i(\beta) = \min\{e_{1i}(\beta), e_{2i}(\beta)\}$, where $e_{2i}(\beta) = \log(C_i) - Z_i^\top \beta$. Write the observed $e_{1i}(\beta)$ as $\varepsilon_i(\beta)$ and the censored and thus unobserved $e_{1i}(\beta)$ as $\varepsilon_i^*(\beta)$. The complete data are represented by $(\varepsilon_i(\beta), \delta_i = 1)$ and $(\varepsilon_i^*(\beta), \delta_i = 0)$, $i = 1, \ldots, n$. Recall that $F(x \mid \beta) = \sum_{i=1}^n p_i(\beta) I(e_i(\beta) \le x)$. The complete-data log conditional likelihood is

$$\ell_c(\beta, F) = \sum_{i=1}^n \Big[\delta_i \log p_i(\beta) + (1 - \delta_i) \sum_{j=1}^n I(\varepsilon_i^*(\beta) = e_j(\beta)) \log p_j(\beta)\Big].$$

Let $\mathcal{D}(\beta) = \{e_i(\beta), i = 1, \ldots, n\}$. It can be seen that

$$\begin{aligned} E\{I(\varepsilon_i^*(\beta) = e_j(\beta)) \mid \mathcal{D}(\beta), \delta_i = 0\} &= \text{pr}\{\varepsilon_i^*(\beta) = e_j(\beta) \mid \mathcal{D}(\beta), \delta_i = 0\} \\ &= \text{pr}\{e_{1i}(\beta) = e_j(\beta) \mid \mathcal{D}(\beta), e_{1i}(\beta) > e_i(\beta)\} \\ &= \frac{\text{pr}(e_{1i}(\beta) = e_j(\beta) \mid \mathcal{D}(\beta)) I(e_{1i}(\beta) > e_i(\beta))}{\text{pr}(e_{1i}(\beta) > e_i(\beta) \mid \mathcal{D}(\beta))} \\ &= \frac{p_j(\beta) I(e_j(\beta) > e_i(\beta))}{\sum_{v=1}^n p_v(\beta) I(e_v(\beta) > e_i(\beta))}. \end{aligned}$$

Given $p_j^{(u)}(\beta)$'s in the u-iteration, in the E-step of the $(u+1)$th cycle, the expectation of complete data log conditional likelihood given observed data is

$$E\{\ell_c(\beta, F) \mid \mathcal{D}(\beta), p_1^{(u)}(\beta) \ldots, p_n^{(u)}(\beta)\} = \sum_{j=1}^n w_j^{(u)}(\beta) \log p_j(\beta),$$

where

$$w_j^{(u)}(\beta) = \delta_j + \sum_{i=1}^n (1 - \delta_i) \frac{p_j^{(u)}(\beta) I(e_j(\beta) > e_i(\beta))}{\sum_{v=1}^n p_v^{(u)}(\beta) I(e_v(\beta) > e_i(\beta))}.$$

The M-step is to maximize $E\{\ell_c(\beta, F) \mid \mathcal{D}(\beta), p_1^{(u)}(\beta) \ldots, p_n^{(u)}(\beta)\}$ under the constraints (2.10), which gives

$$p_j^{(u+1)}(\beta) = \frac{w_j^{(u)}(\beta)}{\sum_{l=1}^n w_l^{(u)}(\beta)} = \frac{1}{n} w_j^{(u)}(\beta), \quad j = 1, \ldots, n.$$

References

Bickel, P. J., C. A. Klaassen, Y. Ritov, and J. A. Wellner (1989). *Efficient and adaptive estimation for semiparametric models*. Baltimore: John Hopkins University Press.

Buckley, J. and I. James (1979). Linear regression with censored data. *Biometrika 66*(3), 429–436.

Cheng, Y.-J., Y.-C. Liu, C.-Y. Tsai, and C.-Y. Huang (2023). Semiparametric estimation of the transformation model by leveraging external aggregate data in the presence of population heterogeneity. *Biometrics 79*(3), 1996–2009.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological) 34*(2), 187–202.

Ding, J., J. Li, Y. Han, I. W. McKeague, and X. Wang (2023). Fitting additive risk models using auxiliary information. *Statistics in Medicine 42*(6), 894–916.

Ding, Y. and B. Nan (2011). A sieve M-theorem for bundled parameters in semiparametric models, with application to the efficient estimation in a linear model for censored data. *Annals of Statistics 39*(6), 3032–3061.

Gao, F. and K. Chan (2023). Noniterative adjustment to regression estimators with population-based auxiliary information for semiparametric models. *Biometrics 79*(1), 140–150.

Han, B., I. Van Keilegom, and X. Wang (2022). Semiparametric estimation of the nonmixture cure model with auxiliary survival information. *Biometrics 78*(2), 448–459.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica 50*(4), 1029–1054.

He, J., H. Li, S. Zhang, and X. Duan (2019). Additive hazards model with auxiliary subgroup survival information. *Lifetime Data Analysis 25*, 128–149.

Horowitz, J. L. (1992). A smoothed maximum score estimator for the binary response model. *Econometrica 60*, 505–531.

Huang, C.-Y., J. Ning, and J. Qin (2015). Semiparametric likelihood inference for left-truncated and right-censored data. *Biostatistics 16*(4), 785–798.

Huang, C.-Y. and J. Qin (2020). A unified approach for synthesizing population-level covariate effect information in semiparametric estimation with survival data. *Statistics in Medicine 39*(10), 1573–1590.

Huang, C.-Y., J. Qin, and H.-T. Tsai (2016). Efficient estimation of the Cox model with auxiliary subgroup survival information. *Journal of the American Statistical Association 111*(514), 787–799.

Jin, Z., D. Y. Lin, L. J. Wei, and Z. Ying (2003). Rank-based inference for the accelerated failure time model. *Biometrika 90*(2), 341–353.

Jin, Z., D. Y. Lin, and Z. Ying (2006). On least-squares regression with censored data. *Biometrika 93*(1), 147–161.

Lai, T. L. and Z. Ying (1991). Large sample theory of a modified Buckley-James estimator for regression analysis with censored data. *The Annals of Statistics 19*, 1370–1402.

Lin, Y. and K. Chen (2013). Efficient estimation of the censored linear regression model. *Biometrika 100*(2), 525–530.

Owen, A. B. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics 18*, 90–120.

Prentice, R. L. (1978). Linear rank tests with right censored data. *Biometrika 65*(1), 167–179.

Qin, J. and J. Lawless (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics 22*(1), 300–325.

Qin, J., Y. Liu, and P. Li (2022). A selective review of statistical methods using calibration information from similar studies. *Statistical Theory and Related Fields 6*(3), 175–190.

Qin, J., J. Ning, H. Liu, and Y. Shen (2011). Maximum likelihood estimations and EM algorithms with length-biased data. *Journal of the American Statistical Association 106*(496), 1434–1449.

Ritov, Y. (1990). Estimation in a linear regression model with censored data. *The Annals of Statistics 18*, 303–328.

Ritov, Y. and J. A. Wellner (1988). Censoring, martingales, and the Cox model. *Contemporary Mathematics 80*, 191–219.

Shang, W. (2022). Statistical inference for Cox model under case-cohort design with subgroup survival information. *Journal of the Korean Statistical Society 51*(3), 884–926.

Shang, W. and X. Wang (2017). The generalized moment estimation of the additive–multiplicative hazard model with auxiliary survival information. *Computational Statistics & Data Analysis 112*, 154–169.

Shang, W. and C. Wu (2023). More effective estimation for additive hazards model in generalized case-cohort study. *Communications in Statistics–Simulation and Computation 52*(11), 5345–5370.

Sheng, Y., Y. Sun, D. Deng, and C.-Y. Huang (2020). Censored linear regression in the presence or absence of auxiliary survival information. *Biometrics 76*(3), 734–745.

Sheng, Y., Y. Sun, C.-Y. Huang, and M.-O. Kim (2021). Synthesizing external aggregated information in the penalized Cox regression under population heterogeneity. *Statistics in Medicine 40*(23), 4915–4930.

Su, P.-F., J. Zhong, Y.-C. Liu, T.-H. Lin, and H.-T. Ou (2023). Efficient estimation of a Cox model when integrating the subgroup incidence rate information. *Journal of Applied Statistics 50*(10), 2151–2170.

Tsiatis, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *The Annals of Statistics 18*, 354–372.

Vardi, Y. (1989). Multiplicative censoring, renewal processes, deconvolution and decreasing density: nonparametric estimation. *Biometrika 76*(4), 751–761.

Wei, L.-J., Z. Ying, and D. Y. Lin (1990). Linear regression analysis of censored survival data based on rank tests. *Biometrika 77*(4), 845–851.

Yang, S. (1997). Extended weighted log-rank estimating functions in censored regression. *Journal of the American Statistical Association 92*(439), 977–984.

Ying, Z. (1993). A large sample study of rank estimation for censored regression data. *The Annals of Statistics 21*, 76–99.

Zeng, D. and D. Lin (2007). Efficient estimation for the accelerated failure time model. *Journal of the American Statistical Association 102*(480), 1387–1396.

Zhou, M. (2005). Empirical likelihood analysis of the rank estimator for the censored accelerated failure time model. *Biometrika 92*(2), 492–498.

Huijuan Ma; Manli Cheng; Yukun Liu; KLATASDS-MOE, School of Statistics and Academy of Statistics and Interdisciplinary Sciences,

East China Normal University, 200062, Shanghai, China

E-mail: hjma@fem.ecnu.edu.cn; cmanli176@163.com; ykliu@sfs.ecnu.edu.cn (The corresponding author)

Donglin Zeng, Department of Biostatistics, University of Michigan, 48109, Michigan, U.S.A.

E-mail: dzeng@umich.edu

Yong Zhou, KLATASDS-MOE, School of Statistics and Academy of Statistics and Interdisciplinary Sciences,

East China Normal University, 200062, Shanghai, China

E-mail: yzhou@fem.ecnu.edu.cn

Table 1: Simulation results for β under the heterogeneous scenario

		β_1				β_2				β_3			
Case	Est	Bias	SD	SE	CP	Bias	SD	SE	CP	Bias	SD	SE	CP
I	$\hat{\beta}_{\text{os,homo}}$	-1	62	65	93.7	0	112	109	94.6	-1	83	94	95.9
	$\hat{\beta}_{\text{os,hete}}$	0	62	65	94.1	3	111	109	94.4	-1	83	94	96.1
	$\tilde{\beta}$	0	63	67	94.5	2	125	129	95.3	-1	84	97	96.4
	$\hat{\beta}_{\text{L}}$	-4	63	52	89.0	2	120	110	92.3	2	84	74	91.6
	$\hat{\beta}_{\text{G1}}$	-16	67	48	85.8	9	105	87	88.3	26	91	61	81.1
	$\hat{\beta}_{\text{G0}}$	-11	67	51	86.7	13	120	106	90.8	20	96	67	82.0
II	$\hat{\beta}_{\text{os,homo}}$	-4	63	68	93.6	5	115	117	94.7	4	90	102	95.9
	$\hat{\beta}_{\text{os,hete}}$	-4	63	68	93.4	7	115	116	94.2	4	90	102	95.7
	$\tilde{\beta}$	-4	64	72	93.5	4	131	139	94.9	4	90	103	95.9
	$\hat{\beta}_{\text{L}}$	-1	86	74	90.2	-7	175	158	92.1	3	122	106	90.4
	$\hat{\beta}_{\text{G1}}$	-12	66	53	88.5	11	106	90	88.1	23	90	68	84.7
	$\hat{\beta}_{\text{G0}}$	-7	71	56	89.5	2	130	115	91.6	16	102	77	87.5
III	$\hat{\beta}_{\text{os,homo}}$	-3	47	51	93.6	6	83	91	96.0	1	67	77	96.0
	$\hat{\beta}_{\text{os,hete}}$	-3	47	51	93.5	6	83	90	95.6	1	67	77	95.8
	$\tilde{\beta}$	-3	48	52	93.6	4	86	100	96.3	1	67	78	95.4
	$\hat{\beta}_{\text{L}}$	-1	71	61	90.2	-1	145	130	93.0	3	100	89	91.0
	$\hat{\beta}_{\text{G1}}$	-7	55	43	90.0	8	103	72	84.1	10	83	56	82.5
	$\hat{\beta}_{\text{G0}}$	-4	54	45	92.5	2	104	91	92.1	10	83	62	86.8
IV	$\hat{\beta}_{\text{os,homo}}$	-2	41	46	92.3	5	76	91	96.4	-1	63	74	95.2
	$\hat{\beta}_{\text{os,hete}}$	-2	41	46	92.3	3	76	90	96.0	-1	63	74	95.2
	$\tilde{\beta}$	-1	41	47	92.6	5	77	98	96.7	0	62	75	95.4
	$\hat{\beta}_{\text{L}}$	-3	89	80	93.5	-2	174	170	95.0	-2	130	119	93.3
	$\hat{\beta}_{\text{G1}}$	-9	61	43	87.3	22	122	74	82.0	5	103	62	82.0
	$\hat{\beta}_{\text{G0}}$	-10	68	46	88.5	8	109	92	92.2	16	93	67	85.9
V	$\hat{\beta}_{\text{os,homo}}$	-2	38	38	93.7	16	100	70	94.8	4	54	57	95.9
	$\hat{\beta}_{\text{os,hete}}$	-2	38	38	93.7	16	101	70	94.7	4	54	57	95.8
	$\tilde{\beta}$	-2	38	38	94.1	10	112	82	94.5	4	54	59	95.8
	$\hat{\beta}_{\text{L}}$	3	116	123	97.4	-9	228	242	97.3	1	161	171	97.5
	$\hat{\beta}_{\text{G1}}$	-15	124	48	90.8	13	528	81	90.2	33	437	66	85.7
	$\hat{\beta}_{\text{G0}}$	-16	124	49	92.3	-5	131	94	93.1	29	175	71	88.5

Table 2: Simulation results for ρ under the heterogeneous scenario

		Case I	Case II	Case III	Case IV	Case V
Proposed	Bias	0.002	0.013	0.024	0.006	0.017
	SD	0.258	0.202	0.276	0.253	0.249
	SE	0.265	0.203	0.275	0.260	0.240
	CP	92.0	93.5	93.1	92.1	95.2
GMM	Bias	0.101	0.054	0.096	0.085	0.098
	SD	0.334	0.235	0.385	0.356	0.563
	SE	0.261	0.197	0.218	0.187	0.239
	CP	93.4	95.2	86.0	81.2	93.2

Proposed, the proposed one-step estimator that incorporates the auxiliary information; GMM, the generalized method of moments estimator incorporating the auxiliary information (Sheng et al., 2020). Bias, empirical bias ($\times 1000$); SD, empirical standard deviation ($\times 1000$); SE, estimated standard error ($\times 1000$); CP, empirical coverage probability.

Table 3: Real data analysis of colon cancer study

		$\hat{\beta}_{\mathrm{os,hete}}$	$\hat{\beta}_{\mathrm{os,homo}}$	$\tilde{\beta}$	$\hat{\beta}_{\mathrm{L}}$	$\hat{\beta}_{\mathrm{G1}}$	$\hat{\beta}_{\mathrm{G0}}$
Gender(β_1)	Est	0.0159	0.0073	0.0043	0.0082	0.0148	0.1129
	SE	0.0082	0.0091	0.0097	0.1818	0.0328	0.1111
	p-value	0.051	0.424	0.661	0.964	0.652	0.309
Diagnosis age(β_2)	Est	-0.0027	-0.0027	-0.0027	-0.0028	0.0012	-0.0017
	SE	0.0004	0.0004	0.0004	0.0095	0.0045	0.0046
	p-value	< 0.001	< 0.001	< 0.001	0.772	0.788	0.704
Lev(β_3)	Est	0.0411	0.0403	0.04	0.0286	0.0807	-0.0199
	SE	0.0114	0.0114	0.0114	0.2394	0.1387	0.129
	p-value	< 0.001	< 0.001	< 0.001	0.905	0.561	0.878
Lev+5FU(β_4)	Est	0.5626	0.5623	0.5622	0.5367	0.5051	0.2599
	SE	0.0128	0.0128	0.0128	0.184	0.1285	0.1357
	p-value	< 0.001	< 0.001	< 0.001	0.004	< 0.001	0.055
ρ	Est	0.7258	-	-	-	0.7028	-
	SE	0.0433	-	-	-	0.0506	-
	p-value	< 0.001	-	-	-	< 0.001	-

Lev, levamisole; Lev+5FU, levamisole combined with fluorouracil. Est, the estimator; SE, the estimated standard error. $\hat{\beta}_{\mathrm{os,hete}}$, the proposed one-step estimator incorporating auxiliary information under the heterogeneous case; $\hat{\beta}_{\mathrm{os,homo}}$, the proposed one-step estimator incorporating auxiliary information under the homogeneous case; $\tilde{\beta}$, the maximum conditional likelihood estimator without auxiliary information; $\hat{\beta}_{\mathrm{L}}$, the weighted log-rank estimator with unit weight; $\hat{\beta}_{\mathrm{G1}}$, the extended generalized method of moments estimator incorporating the auxiliary information (Sheng et al., 2020); $\hat{\beta}_{\mathrm{G0}}$, the generalized method of moments estimator without the auxiliary information (Sheng et al., 2020).