Statistics and Its Interface Volume 8 (2015) 367–377

Regression analysis of interval-censored failure time data with the additive hazards model in the presence of informative censoring

Shishun Zhao, Tao Hu, Ling Ma, Peijie Wang, and Jianguo Sun*

Regression analysis of interval-censored failure time data has recently attracted a great deal of attention and many procedures have been developed (Chen et al., 2012; Finkelstein, 1986; Sun, 2006; Sun and Li, 2013). However, most of the established procedures are for noninformative censoring, meaning that the censoring mechanism is independent of the underlying failure time of interest. In this paper, we discuss a more general situation where the censoring mechanism and the failure time of interest may be related and a sieve maximum likelihood estimation procedure is proposed for such data arising from the additive hazards model. In the method, the copula model is employed to model the correlation. The asymptotic properties of the resulting estimators are established and a simulation study is conducted, which indicates that the presented approach works well for the situations considered. An illustrative example is also provided.

Keywords and phrases: Additive hazards model, Copula models, Informative censoring, Regression analysis.

1. INTRODUCTION

Statistical analysis of interval-censored failure time data has recently attracted a great deal of attention (Chen et al., 2012; Finkelstein, 1986; Sun, 2006; Sun and Li, 2013). By interval-censored data, we mean that instead of being known exactly, the failure time of interest is observed only to belong to a window or an interval. We will have an exact observation if the interval reduces to a single point and a right-censored observation if the interval includes infinity. In other words, interval-censored data include right-censored failure time data as a special case (Kalbfleisch and Prentice, 2002). Amomg others, one field that often produces interval-censored data is medical follow-up or clinical studies (Chen et al., 2012; Finkelstein, 1986).

Many procedures have been developed for regression analysis of interval-censored failure time data (Betensky et al. 2001; Chen and Sun 2010; Goggins and Finkelstein 2000; Lin et al., 1998; Wang et al., 2010). However, most of the existing methods only apply to the situation where the censoring mechanism that generates censoring intervals is independent of the failure time of interest, which is often referred to as noninformative censoring. As pointed out by many authors, sometimes this may not be true (Betensky and Finkelstein, 2002; Finkelstein et al., 2002; Huang and Wolfe, 2002; Zhang et al., 2007). For example, in the case of right-censored data, the failure time of interest and the censoring time may be related and Huang and Wolfe (2002) discussed regression analysis of such data under the proportional hazards model. Finkelstein et al. (2002) and Zhang et al. (2007) considered the interval-censored data in which the censoring intervals may be correlated with the failure time of interest. The former investigated the nonparametric estimation problem, while the latter considered the regression problem. In this paper, we will discuss the same problem as that in Zhang et al. (2007), but employ the more general copula model approach (Hougaard, 2000; Nelsen, 2006; Zheng and Klein, 1995).

Among many regression models for failure time data, one of the most commonly used models is the additive hazards model (Chen and Sun, 2010; Li et al., 2012; Lin et al., 1998; Lin and Ying, 1994; Martinussen and Scheike, 2002; Tong et al., 2012; Zhou and Sun, 2003). For example, one of the early references on this model was given by Lin and Ying (1994), who discussed the fitting of the model to right-censored data. Following them, Lin et al. (1998) and Martinussen and Scheike (2002) investigated the same problem except for current status data and developed some estimating equation-based approaches. More recently, Li et al. (2012) and Tong et al. (2012) studied the use of the model for regression analysis of clustered interval-censored data and bivariate current status data, respectively. In the following, we will focus on regression analysis of interval-censored data arising from the additive hazards model when there exists informative censoring.

To present the proposed method, we will first in Section 2 describe the type of interval-censored data or censoring mechanism considered here along with some notation and models. In particular, we will assume that the relationship between the failure time of interest and the censoring mechanism can be described by the correlation between the failure time and the length of a censoring interval. Note that

*Corresponding author.

this would be the case in, for example, a medical follow-up study as in this situation, patients may tend to visit clinical centers less or more often than the pre-specified schedule depending on their disease status. More comments on this are given below. The copula model will be used to model the correlation. For inference, we will present a sieve maximum likelihood estimation procedure based on I-spline functions in Section 3 (Ramsay, 1988; Lu et al. 2007) and the asymptotic properties of the proposed estimators are established. A simulation study is performed in Section 4 and the obtained results indicate that the proposed method works well in practical situations. Section 5 provides an illustrative example and Section 6 contains some concluding remarks.

2. INTERVAL-CENSORED DATA AND INFORMATIVE CENSORING

Consider a failure time study that consists of n independent subjects and gives only interval-censored data characterized by $\{ L_i < T_i \leq R_i;\, i = 1, ..., n \}$. Here T_i denotes the failure time of interest but not observed and $(L_i, R_i]$ represents the interval to which T_i is observed to belong. As mentioned above, many procedures have been developed for the analysis of such interval-censored data, but most of them assume that L_i and R_i are independent of T_i completely or given covariates and this may not be true in practice. The latter situation is usually referred to as informative censoring (Sun, 2006). In the following, we consider a situation where L_i and R_i may be related to T_i, but they are independent given the interval length $W_i = R_i - L_i$. In other words, the relationship between T_i and $(L_i, R_i]$ can be described by the relationship between T_i and W_i.

A motivating example for the informative censoring described above is clinical trials or medical follow-up studies. In these situations, some pre-specified observation or follow-up times are usually given before the study, but it is wel-known that most patients would not follow these time points for their clinical visits or observations. One case that can often occur is that the patients may tend to pay less or more visits than the pre-specified visits depending on their health or disease status. A natural way to describe this is clearly to employ the interval lengths W_i's. Note that an equivalent approach for describing the informative censoring here is by the relationship between T_i and L_i, which may be more natural given the fact that it is often the case that $(L_i, R_i]$ depends on T_i through L_i. On the other hand, it is more convenient to model the relationship between T_i and W_i.

More specifically, we will assume that

$$P(T_i \leq t | L_i, R_i, Z_i) = P(T_i \leq t | W_i, Z_i)$$

given the vector of covariates Z_i. To describe the effects of covariates of the T_i's and W_i's, we will assume that the hazard function of T_i has the form

$$\lambda(t|Z_i) = \lambda_{10}(t) + Z_i'\beta \,, \tag{1}$$

and W_i follows the proportional hazards model given by

$$\lambda(w|Z_i) = \lambda_{20}(w) \exp(Z_i'\gamma) \,. \tag{2}$$

In the above, both $\lambda_{10}(t)$ and $\lambda_{20}(w)$ are unspecified baseline hazard functions, and β and γ denote the vectors of regression parameters. That is, the T_i's follow the additive hazards model marginally (Lin and Ying, 1994).

Let F_T and F_W denote the marginal distribution of the T_i's and W_i's, respectively, and F their joint distribution given covariates. Then it follows from the Theorem 2.3.3 of Nelsen (2006) that there exists a copula function $C_\alpha(u, v)$ defined on $I^2 = [0, 1] \times [0, 1]$ with $C_\alpha(u, 0) = C_\alpha(0, v) = 0$, $C_\alpha(u, 1) = u$ and $C_\alpha(1, v) = v$ such that

$$F(t, w) = C_\alpha((F_T(t), F_W(w)) \,.$$

In the above, the parameter α represents the association between the T_i's and W_i's. Note that the copula model is commonly used to describe the correlation between variables and among others, one advantage is that it allows one to model the correlation and the marginal distribution separately (Hougaard, 2000; Zheng and Klein, 1995). By following the conditional inversion idea (Section 2.8 and Section 3.1, Nelsen, 2006), we have

$$P(T \leq t | W = w, Z) = \frac{\partial C_\alpha(u, v)}{\partial v}\Big|_{u=F_T(t), v=F_W(w)} \,,$$

which will be denoted by $m_\alpha(F_T(t), F_W(w))$ for simplicity.

In the following, it will be supposed that the main goal is to estimate the regression parameters β and γ. By following Zheng and Klein (1995) and others, we will assume that the copula function and α are known. More comments on this are given below.

3. SIEVE MAXIMUM LIKELIHOOD ESTIMATION

Define $\Lambda_1(t) = \int_0^t \lambda_{10}(s)ds$, $\Lambda_2(w) = \int_0^w \lambda_{20}(s)ds$, and $\theta = (\beta, \gamma, \Lambda_1, \Lambda_2)$. Let f_W denote the marginal density function of the W_i's given covariates. Then under the models above, we have

$$F_T(t) = 1 - \exp\{-\Lambda_1(t) - Z'\beta t\} \,,$$
$$F_W(w) = 1 - \exp\{-\Lambda_2(w)\exp(\gamma' Z)\} \,,$$

and

$$f_W(w) = \exp\{-\Lambda_2(w)\exp(\gamma' Z)\}\, \lambda_{20}(w)\, \exp(\gamma' Z) \,.$$

Furthermore define $\delta_i = I(R_i < \infty)$, $i = 1, \cdots, n$ and let ζ denote longest follow-up time. Then the likelihood of θ has the form

$$L(\theta)$$
$$= \prod_{i=1}^{n} \Big\{ [m_\alpha(F_T(R_i), F_W(W_i))$$

$$- m_\alpha(F_T(L_i), F_W(W_i))]f_W(W_i)\Big\}^{\delta_i}$$
$$\times \Big\{1 - F_W(\zeta - L_i) - F_T(L_i)$$
$$+ C_\alpha(F_T(L_i), F_W(\zeta - L_i))\Big\}^{1-\delta_i}.$$

The derivation of the likelihood function above is sketched in Appendix B.

For estimation of β and γ or θ, a natural way is clearly to maximize the likelihood function $L(\theta)$. On the other hand, this may not be easy due to the dimensions of $\Lambda_1(t)$ and $\Lambda_2(t)$. For this, following Huang and Rossini (1997) and others, we will employ the sieve approach. Specifically, let $\psi = (\Lambda_1, \Lambda_2)$ and define the sieve space

$$\Theta_n = \Big\{\theta_n = (\beta, \gamma, \psi_n) : \psi_n = (\Lambda_{1n}(t), \Lambda_{2n}(t))\Big\} = \mathcal{B} \otimes \mathcal{M}_n^1 \otimes \mathcal{M}_n^2$$

for θ. In the above,

$$\mathcal{B} = \{(\beta', \gamma')' \in R^{2p}, \|\beta\| + \|\gamma\| \leq M\},$$
$$\mathcal{M}_n^1 = \{\Lambda_{1n} : \Lambda_{1n}(t) = \sum_{j=1}^{m+k_n} \xi_j I_{j1}(t), \quad \xi_j \geq 0, j = 1, \ldots, m + k_n, t \in [l, u]\},$$

and

$$\mathcal{M}_n^2 = \{\Lambda_{2n} : \Lambda_{2n}(t) = \sum_{j=1}^{m+k_n} \eta_j I_{j2}(t), \quad \eta_j \geq 0, j = 1, \ldots, m + k_n, t \in [0, \zeta - l]\}.$$

Here p denotes the dimension of Z_i, the I_{j1}'s and I_{j2}'s are I-spline base functions, and m and $k_n = o(n^\nu)$ represent the order and the number of interior knots of the functions, respectively, with $0 < \nu < 0.5$. Then it is natural to estimate θ by maximizing the log likelihood function $l_n(\theta) = \log L(\theta)$ over the sieve space Θ_n.

Define the estimator $\hat{\theta}_n = (\hat{\beta}_n, \hat{\gamma}_n, \hat{\Lambda}_{1n}, \hat{\Lambda}_{2n}) = \arg\min_{\Theta_n} l_n(\theta)$. To establish the asymptotic properties of $\hat{\theta}_n$, let $\theta_0 = (\beta_0, \gamma_0, \Lambda_{10}, \Lambda_{20})$ denote the true value of θ. Also for a vector a and a function f, let $\|a\|$ denote the Euclidean norm and $\|f\|_\infty = \sup_t |f(t)|$, the supremum norm. For a random variable X being distributed according to the probability measure P, define $\|f(X)\|_2 = (\int f^2 dP)^{1/2}$, the $L^2(P)$ norm. The following three theorems give the consistency and asymptotic normality of the proposed estimator.

Theorem 1. *Suppose that the conditions A1–A4 described in Appendix A hold. Then as* $n \to \infty$, *we have that* $\|\hat{\beta}_n - \beta_0\| \to 0$, $\|\hat{\gamma}_n - \gamma_0\| \to 0$, $\|\hat{\Lambda}_{1n} - \Lambda_{10}\|_2 \longrightarrow 0$, *and* $\|\hat{\Lambda}_{2n} - \Lambda_{20}\|_2 \longrightarrow 0$ *almost surely.*

Theorem 2. *Also suppose that the conditions A1–A4 described in Appendix A hold. Then as* $n \to \infty$, *we have that*

$$\|\hat{\Lambda}_{1n} - \Lambda_{10}\|_2 + \|\hat{\Lambda}_{2n} - \Lambda_{20}\|_2 = O_p(n^{-(1-\nu)/2} + n^{-r\nu}),$$

where r *is defined in condition A4 described in Appendix A.*

Theorem 3. *Suppose that the conditions A1–A5 described in Appendix A hold and* $r > 2$. *Then as* $n \to \infty$, *we have*

$$n^{1/2}\left((\hat{\beta}_n - \beta_0)', (\hat{\gamma}_n - \gamma_0)'\right)' \to N(0, \Sigma)$$

in distribution, where Σ *is given in Appendix A. Furthermore* $\hat{\beta}_n$ *and* $\hat{\gamma}_n$ *are semiparametrically efficient.*

The proofs of the theorems above are sketched in Appendix A. To estimate the covariance matrix Σ, we suggest to adopt the common and straightforward approach that uses the inverse of the observed information matrix based on the sieve likelihood function. Note that this approach could be computationally intensive as it involves the inversion of a potentially high-dimensional and possibly ill-conditioned matrix. However, the simulation study given in the next section indicates that it works well in general, especially when m and k_n are not too large.

4. A SIMULATION STUDY

A simulation study was performed to evaluate the estimation procedure proposed in the previous sections. In the study, it was assumed that the covariate Z_i's follow the Bernoulli distribution with the success probability of 0.5. To generate the failure time T_i and the censoring interval $(L_i, R_i]$, we first generated two independent random numbers u_i and c_i from the uniform distribution over $(0, 1)$. Then define the number v_i to be the solution to the equation $c_i = \partial C_\alpha(u, v)/\partial u|_{u=u_i, v=v_i}$ for a given copula function $C_\alpha(u, v)$ and furthermore define $T_i = t_i$ and $W_i = w_i$, where t_i and w_i denote the solutions to the equations $F_T(t_i) = u_i$ and $F_W(w_i) = v_i$, respectively. Here we took $\lambda_{10}(t) = 1.5$ and $\lambda_{20}(w) = 0.5$. At the last, we define L_i to be the largest number in $\{0, W_i, 2W_i, 3W_i, \ldots\}$ that is smaller than the minimum of T_i and ζ and $R_i = L_i + W_i$. Here ζ is a constant taken to give a proper percentage of right-censored observations (PRC).

For the informative censoring, we considered two copula models and they are the Gumbel and Frank models given by

$$C_\alpha(u, v) = \begin{cases} \exp\{-[(-\log u)^\alpha + (-\log v)^\alpha]^{1/\alpha}\}, \alpha \geq 1, \\ \log_\alpha\{1 + (\alpha^u - 1)(\alpha^v - 1)/(\alpha - 1)\}, \alpha > 0, \alpha \neq 1, \end{cases}$$

respectively. Note that for different copula models, the association parameter α has different ranges and thus it is common to employ the Kendall's τ to measure the association between the T_i's and W_i's in general. For the Gumbel copula, the relationship is given by $\tau = 1 - 1/\alpha$ and we have $\tau = 1 + 4x^{-1}\{D_1(x) - 1\}$ for the Frank copula, where $x = -\log\alpha$ and $D_1(x) = x^{-1}\int_0^x t(e^t - 1)^{-1} dt$. The results given below are based on $n = 200$ and 500 replications with the quadratic splines with 3 interior knots. More specifi-

Table 1. Results on estimation of regression parameters based on the simulated data under the Gumbel model and with $\beta_0 = 0$

				$\hat{\beta}_n$				$\hat{\gamma}_n$		
PRC	γ_0	τ	Bias	SSE	SEE	CP	Bias	SSE	SEE	CP
20%	0.0	0.05	0.0059	0.3088	0.2942	0.950	-0.0135	0.1575	0.1587	0.954
		0.25	-0.0009	0.3274	0.3146	0.964	-0.0078	0.1592	0.1587	0.950
		0.50	-0.0144	0.4037	0.3612	0.959	-0.0045	0.1590	0.1587	0.950
	0.2	0.05	0.0019	0.3036	0.2870	0.958	-0.0104	0.1554	0.1589	0.952
		0.25	-0.0056	0.3083	0.3046	0.958	-0.0092	0.1568	0.1587	0.956
		0.50	-0.0415	0.3584	0.3345	0.957	-0.0071	0.1593	0.1587	0.942
	0.4	0.05	-0.0030	0.2976	0.2820	0.952	-0.0103	0.1584	0.1597	0.958
		0.25	-0.0123	0.3016	0.2948	0.962	-0.0107	0.1564	0.1593	0.964
		0.50	-0.0468	0.3387	0.3222	0.957	-0.0101	0.1616	0.1592	0.946
40%	0.0	0.05	0.0050	0.3210	0.3178	0.956	-0.0036	0.1758	0.1834	0.954
		0.25	0.0013	0.3453	0.3484	0.968	-0.0040	0.1756	0.1832	0.960
		0.50	-0.0004	0.4173	0.4015	0.976	0.00003	0.1823	0.1831	0.956
	0.2	0.05	-0.0009	0.3227	0.3143	0.964	-0.0026	0.1777	0.1835	0.954
		0.25	-0.0098	0.3351	0.3424	0.956	-0.0096	0.1796	0.1836	0.960
		0.50	-0.0349	0.3728	0.3844	0.982	-0.0058	0.1819	0.1839	0.952
	0.4	0.05	-0.0102	0.3239	0.3117	0.950	0.0009	0.1811	0.1848	0.956
		0.25	-0.0246	0.3348	0.3406	0.956	-0.0152	0.1849	0.1848	0.952
		0.50	-0.0274	0.3528	0.3707	0.966	-0.0132	0.1917	0.1849	0.942

cally, we approximate $\Lambda_1(t)$ and $\Lambda_2(w)$ by $\sum_{j=1}^{6} \xi_j I_{j1}(t)$ and $\sum_{j=1}^{6} \eta_j I_{j2}(w)$, respectively. Then we maximize the log likelihood function $l_n(\theta)$ over the sieve space Θ_n or the parameters β, γ, ξ_j's and η_j's by using the Newton-type algorithm through the R function *nlm*. For the knot selection, we used the $0.25, 0.5, 0.75$ quantiles of the pooled set of all L_i's and the uncensored R_i's for estimation of Λ_1, while for estimation of Λ_2, we used $0.25, 0.5, 0.75$ quantiles of the pooled set of the W_i's from the non-right-censored subjects and $(\zeta - L_i)$'s from the right-censored subjects.

Table 1 presents the results on estimation of β and γ based on the simulated data generated under the Gumbel model with $\beta_0 = 0$, $\gamma_0 = 0, 0.2$ or 0.4, and $\tau = 0.05, 0.25$ or 0.5. Here we considered two percentages for right-censored observations, 20% and 40%. The table includes the estimated bias given by the average of the estimators minus the true value (Bias), the sample standard deviation (SSE) of the estimators, the average of the estimated standard errors (SEE) and the 95% empirical coverage probability (CP). The results given in Table 2 were obtained similarly as with Table 1 except $\beta_0 = 0.4$. Tables 3 and 4 gives the results obtained under the Frank model with $\tau = -0.5, -0.25, 0.25$ and 0.5, and the other set-ups being the same as with Tables 1 and 2. One can see from these tables that the proposed sieve maximum likelihood estimation procedure seems to work well for the situations considered. In particular, the estimator seems to be unbiased and the variance estimation seems to be reasonable. Also as expected, the estimation is more efficient when there is less right censoring although the difference is not large.

Note that one question of practical interest is the robustness of the proposed estimation procedure to the selection of copula models. To investigate this, Table 5 presents the results on estimation of regression parameters β and γ based on the simulated data generated under the set-up used for Table 2 but obtained as those in Table 4. That is, we generated the data from the Gumbel model but assumed the Frank model instead. One can see that the estimation seems to be reasonable and similar to the results given in Table 2 and suggests that the method seems to be robust. We also considered a few other set-ups and obtained similar conclusions.

5. AN ILLUSTRATIVE EXAMPLE

In this section, for the illustration, we apply the proposed methodology to a well-known set of interval-censored data on breast cancer patients that has been discussed by many authors (Sun, 2006). The data set consists of 94 patients in two treatment groups, radiation therapy alone (46) and radiation therapy plus adjuvant chemotherapy (48). In the study, one variable of interest is the time until the appearance of breast retraction, and the patients were asked to visit clinical centers at pre-specified time points. However, as expected, most of them had their own observation times. Hence only interval-censored data are available for the time to breast retraction. One objective of the study is to compare the effects of the two treatments on the time to breast retraction.

For the analysis, define T_i to be the time to breast retraction for patient i and $Z_i = 0$ if the patient was given radiotherapy alone and 1 otherwise. Following the simulation study described in the previous section, we will use the quadratic I-spline functions for the sieve spaces and considered the use of both Gumbel and Frank models. For the number of knots, we considered several choices including

Table 2. *Results on estimation of regression parameters based on the simulated data under the Gumbel model and with* $\beta_0 = 0.4$

				$\hat{\beta}_n$				$\hat{\gamma}_n$		
PRC	γ_0	τ	Bias	SSE	SEE	CP	Bias	SSE	SEE	CP
20%	0.0	0.05	0.0365	0.3643	0.3535	0.956	-0.0058	0.1560	0.1587	0.952
		0.25	0.0410	0.4098	0.3833	0.962	0.0065	0.1614	0.1585	0.952
		0.50	0.0298	0.4944	0.4495	0.966	0.0060	0.1614	0.1581	0.952
	0.2	0.05	0.0273	0.3492	0.3400	0.956	-0.0007	0.1574	0.1590	0.948
		0.25	0.0199	0.3691	0.3632	0.964	-0.0002	0.1568	0.1589	0.960
		0.50	0.0160	0.4550	0.4227	0.960	-0.0012	0.1594	0.1590	0.946
	0.4	0.05	0.0263	0.3439	0.3292	0.956	0.0007	0.1583	0.1598	0.962
		0.25	0.0164	0.3529	0.3477	0.966	-0.0008	0.1590	0.1596	0.952
		0.50	-0.0114	0.3891	0.3860	0.972	-0.0015	0.1612	0.1597	0.944
40%	0.0	0.05	0.0264	0.3748	0.3740	0.964	0.0281	0.1772	0.1835	0.956
		0.25	0.0416	0.4244	0.4193	0.972	0.0199	0.1765	0.1833	0.964
		0.50	0.0403	0.5041	0.4816	0.976	0.0217	0.1819	0.1830	0.962
	0.2	0.05	0.0184	0.3609	0.3650	0.968	0.0336	0.1810	0.1839	0.952
		0.25	0.0134	0.3917	0.4005	0.970	0.0215	0.1782	0.1838	0.958
		0.50	0.0143	0.4785	0.4620	0.976	0.0142	0.1824	0.1837	0.950
	0.4	0.05	0.0106	0.3621	0.3575	0.964	0.0391	0.1787	0.1853	0.964
		0.25	0.0103	0.3947	0.3872	0.966	0.0293	0.1900	0.1853	0.954
		0.50	-0.0208	0.4211	0.4348	0.972	0.0128	0.1834	0.1848	0.952

Table 3. *Results on estimation of regression parameters based on the simulated data under the Frank model and with* $\beta_0 = 0$

				$\hat{\beta}_n$				$\hat{\gamma}_n$		
PRC	γ_0	τ	Bias	SSE	SEE	CP	Bias	SSE	SEE	CP
20%	0.0	-0.50	0.0303	0.2444	0.2232	0.936	-0.0207	0.1437	0.1535	0.956
		-0.25	0.0148	0.2786	0.2583	0.952	-0.0190	0.1485	0.1570	0.964
		0.25	0.0021	0.3450	0.3246	0.948	-0.0078	0.1587	0.1592	0.952
		0.50	-0.0124	0.3848	0.3688	0.966	-0.0054	0.1604	0.1589	0.944
	0.2	-0.50	0.0360	0.2401	0.2212	0.924	-0.0166	0.1432	0.1531	0.964
		-0.25	0.0178	0.2734	0.2555	0.950	-0.0144	0.1500	0.1570	0.962
		0.25	-0.0144	0.3217	0.3123	0.962	-0.0112	0.1577	0.1596	0.950
		0.50	-0.0408	0.3829	0.3535	0.970	-0.0114	0.1619	0.1593	0.946
	0.4	-0.50	0.0417	0.2477	0.2202	0.918	-0.0102	0.1448	0.1534	0.962
		-0.25	0.0167	0.2695	0.2524	0.956	-0.0090	0.1503	0.1578	0.962
		0.25	-0.0261	0.3223	0.3031	0.954	-0.0085	0.1608	0.1605	0.952
		0.50	-0.0366	0.3576	0.3290	0.958	-0.0077	0.1654	0.1602	0.944
40%	0.0	-0.50	0.0405	0.2435	0.2325	0.944	-0.0267	0.1684	0.1748	0.952
		-0.25	0.0169	0.2823	0.2705	0.954	-0.0209	0.1744	0.1804	0.966
		0.25	-0.0022	0.3700	0.3579	0.946	-0.0042	0.1777	0.1841	0.966
		0.50	-0.0061	0.4195	0.4181	0.974	0.0018	0.1816	0.1831	0.954
	0.2	-0.50	0.0464	0.2425	0.2314	0.946	-0.0136	0.1648	0.1739	0.954
		-0.25	0.0179	0.2847	0.2686	0.954	-0.0071	0.1719	0.1801	0.966
		0.25	-0.0151	0.3557	0.3490	0.944	-0.0131	0.1828	0.1847	0.952
		0.50	-0.0443	0.4154	0.3995	0.976	-0.0147	0.1829	0.1833	0.950
	0.4	-0.50	0.0428	0.2514	0.2315	0.946	0.0115	0.1819	0.1744	0.940
		-0.25	0.0180	0.2854	0.2681	0.950	0.0083	0.1731	0.1812	0.970
		0.25	-0.0446	0.3483	0.3436	0.954	-0.0193	0.1837	0.1860	0.954
		0.50	-0.0554	0.3935	0.3861	0.960	-0.0305	0.1884	0.1845	0.952

$k_n = 1, 2, 3, 4$ and 5 and used the same method as in the previous section for the knot selection. Table 6 gives the estimated treatment effects with $k_n = 3$ and 4 along with the estimated standard errors and the p-values for testing no treatment effect for various values of τ. They suggest that there seems no significant treatment effect no matter the correlation between the failure time of interest and the censoring interval.

Table 4. Results on estimation of regression parameters based on the simulated data under the Frank model and with $\beta_0 = 0.4$

				$\hat{\beta}_n$				$\hat{\gamma}_n$		
PRC	γ_0	τ	Bias	SSE	SEE	CP	Bias	SSE	SEE	CP
20%	0.0	-0.50	0.0224	0.2636	0.2593	0.948	-0.0120	0.1563	0.1541	0.944
		-0.25	0.0313	0.3205	0.3033	0.952	-0.0040	0.1713	0.1576	0.928
		0.25	0.0259	0.4043	0.3887	0.962	0.0028	0.1544	0.1588	0.948
		0.50	0.0150	0.5172	0.4772	0.980	0.0072	0.1669	0.1585	0.934
	0.2	-0.50	0.0222	0.2783	0.2587	0.944	0.0046	0.1545	0.1541	0.940
		-0.25	0.0367	0.3142	0.2972	0.960	-0.0093	0.1495	0.1574	0.966
		0.25	0.0373	0.3898	0.3769	0.964	0.0049	0.1593	0.1594	0.952
		0.50	0.0012	0.4463	0.4301	0.956	0.0050	0.1625	0.1591	0.944
	0.4	-0.50	0.0301	0.2701	0.2561	0.948	0.0019	0.1574	0.1546	0.956
		-0.25	0.0428	0.3174	0.2937	0.944	-0.0035	0.1516	0.1581	0.954
		0.25	0.0085	0.3680	0.3570	0.958	0.0064	0.1622	0.1604	0.946
		0.50	-0.0217	0.4420	0.3949	0.948	-0.0002	0.1668	0.1599	0.934
40%	0.0	-0.50	0.0285	0.2881	0.2713	0.944	0.0185	0.1802	0.1762	0.940
		-0.25	0.0408	0.3285	0.3143	0.946	0.0055	0.1721	0.1808	0.962
		0.25	0.0439	0.4496	0.4310	0.964	0.0286	0.1763	0.1839	0.956
		0.50	0.0414	0.5732	0.5271	0.966	0.0065	0.1885	0.1830	0.954
	0.2	-0.50	0.0333	0.2798	0.2684	0.940	0.0340	0.1810	0.1759	0.936
		-0.25	0.0368	0.3211	0.3091	0.964	0.0227	0.1750	0.1812	0.958
		0.25	0.0097	0.4237	0.4113	0.968	0.0260	0.1845	0.1846	0.954
		0.50	-0.0153	0.4820	0.4808	0.974	0.0110	0.1828	0.1835	0.956
	0.4	-0.50	0.0469	0.2891	0.2670	0.934	0.0470	0.1765	0.1766	0.940
		-0.25	0.0415	0.3281	0.3069	0.954	0.0432	0.1792	0.1826	0.948
		0.25	-0.0251	0.3956	0.3953	0.960	0.0237	0.1816	0.1862	0.956
		0.50	-0.0402	0.4724	0.4481	0.962	-0.0120	0.1925	0.1843	0.934

Table 5. Results on estimation of regression parameters based on the simulated data under the Gumbel but estimated under the Frank model and with $\beta_0 = 0.4$

				$\hat{\beta}_n$				$\hat{\gamma}_n$		
PRC	γ_0	τ	Bias	SSE	SEE	CP	Bias	SSE	SEE	CP
20%	0.0	0.05	0.0499	0.3758	0.3625	0.956	-0.0052	0.1561	0.1588	0.948
		0.25	0.0836	0.4521	0.4220	0.962	0.0080	0.1623	0.1587	0.950
		0.50	0.0538	0.5182	0.4744	0.960	0.0073	0.1617	0.1580	0.952
	0.2	0.05	0.0437	0.3607	0.3482	0.954	0.0002	0.1579	0.1591	0.948
		0.25	0.0734	0.4099	0.4012	0.964	0.0019	0.1582	0.1592	0.952
		0.50	0.0444	0.4838	0.4487	0.964	-0.0035	0.1592	0.1584	0.950
	0.4	0.05	0.0442	0.3559	0.3370	0.950	0.0020	0.1590	0.1600	0.962
		0.25	0.0853	0.3916	0.3839	0.954	0.0018	0.1608	0.1600	0.952
		0.50	0.0497	0.4317	0.4148	0.966	-0.0048	0.1620	0.1593	0.944
40%	0.0	0.05	0.0374	0.3849	0.3797	0.964	0.0295	0.1776	0.1837	0.954
		0.25	0.0781	0.4683	0.4456	0.960	0.0252	0.1779	0.1838	0.964
		0.50	0.0463	0.5176	0.4850	0.968	0.0263	0.1820	0.1823	0.952
	0.2	0.05	0.0319	0.3706	0.3697	0.964	0.0352	0.1817	0.1842	0.952
		0.25	0.0592	0.4293	0.4209	0.954	0.0263	0.1796	0.1844	0.956
		0.50	0.0269	0.4919	0.4612	0.980	0.0141	0.1826	0.1833	0.954
	0.4	0.05	0.0256	0.3712	0.3615	0.956	0.0409	0.1795	0.1856	0.964
		0.25	0.0720	0.4266	0.4001	0.944	0.0334	0.1918	0.1860	0.946
		0.50	0.0143	0.4394	0.4319	0.974	0.0089	0.1843	0.1843	0.948

6. CONCLUDING REMARKS

In the previous sections, we discussed regression analysis of interval-censored failure time data in the presence of informative censoring and a sieve maximum likelihood estimation procedure was developed. As discussed before, interval-censored data occur in many fields and also many procedures have been developed for their analysis. But most of them are for the case of noninformative censoring, especially for regression analysis. The proposed procedure can

Table 6. Estimated treatment effects for the breast cancer data

Model	τ	$\hat{\beta}_n$	SEE	p-value	$\hat{\gamma}_n$	SEE	p-value
				$k_n = 3$			
Gumbel	0	0.01289	0.00730	0.078	0.6993	0.2769	0.0116
	0.25	0.00931	0.00723	0.198	0.6232	0.2817	0.0270
	0.50	0.00568	0.00643	0.377	0.4859	0.2838	0.0869
	0.61	0.00450	0.00600	0.452	0.4294	0.2870	0.1346
	0.65	0.00421	0.00582	0.470	0.4135	0.2880	0.1511
Frank	0	0.01289	0.00730	0.078	0.6993	0.2769	0.0116
	0.25	0.00889	0.00653	0.174	0.6135	0.2727	0.0244
	0.50	0.00587	0.00557	0.292	0.4594	0.2693	0.0880
	0.62	0.00427	0.00505	0.397	0.3676	0.2708	0.1747
	0.65	0.00385	0.00491	0.432	0.3437	0.2724	0.2070
				$k_n = 4$			
Gumbel	0	0.01286	0.00745	0.084	0.7093	0.2772	0.0105
	0.25	0.00897	0.00731	0.220	0.6281	0.2820	0.0259
	0.50	0.00492	0.00645	0.446	0.4748	0.2852	0.0959
	0.62	0.00360	0.00599	0.548	0.4084	0.2875	0.1555
	0.65	0.00329	0.00588	0.575	0.3894	0.2881	0.1765
Frank	0	0.01286	0.00745	0.084	0.7093	0.2772	0.0105
	0.25	0.00854	0.00658	0.194	0.6185	0.2724	0.0232
	0.50	0.00540	0.00559	0.334	0.4580	0.2693	0.0890
	0.63	0.00380	0.00505	0.451	0.3620	0.2709	0.1815
	0.65	0.00339	0.00491	0.490	0.3323	0.2709	0.2200

apply to various types of informative censoring as the copula model is very flexible and includes many choices. On the other hand, one may ask if there exists a method to determine which specific copula model is appropriate for a given problem. The same can be asked about the association parameter α. As mentioned above, in general, they are not identifiable unless there exists some extra information and similar situations exist in many contexts. A simple example is how to choose a proper regression model among a class of models such as linear transformation models for a given set of failure time data.

Note that in the proposed approach, we have assumed the additive hazards model and the proportional hazards model for the T_i and W_i, respectively. It should be straightforward to apply the idea discussed above to the situation where the T_i or W_i may follow some other models such as the linear transformation model. However, the implementation or the maximization may be quite different depending on the model and the same is true for the derivation of the asymptotic properties of the resulting estimators of regression parameters.

It is easy to see that compared to the existing regression methods that base the analysis on frailty models or latent variables, the proposed approach has an advantage that it applies to more general situations. In addition, the determination of the proposed estimators is relatively easy as most of the frailty models-based procedures have to rely on EM or some iterative algorithms. Note that among three variables L_i, R_i and W_i, one only needs to deal with two of them to address informative censoring. Also note that in the proposed approach, it has been assumed that the informative censoring can be described by the correlation between the failure time of interest and the length of censoring intervals. Although this could cover many situations such as medical follow-up studies, sometimes the informative censoring may involve both L_i and R_i instead of just W_i. It is apparent that this latter situation is much more complicated than that considered in this paper and a new estimation procedure is thus needed.

ACKNOWLEDGEMENTS

The authors wish to thank the guest editor and the two reviewers for their constructive and helpful comments and suggestions. This work was partly supported by the Humanities and Social Science Research Project of Ministry of Education of P.R. China (11YJAZH125) to the first author, the NSFC of P.R. China (11371062) to the second author, and NSF and NIH grants to the last author.

APPENDIX A. PROOFS OF THEOREMS 1–3

This appendix will sketch the proofs for the asymptotic properties of the proposed estimators. For this, we need the following regularity conditions.

(A1) The covariate Z_i's have a bounded support.
(A2) The copula function $C(\cdot,\cdot)$ has a bounded first order partial derivatives and both partial derivatives are Lipschitz.
(A3) For $\theta^j = (\beta^j, \gamma^j, \Lambda_1^j, \Lambda_2^j)$, $j = 1, 2$, define the distance $d^2(\theta^1, \theta^2) = \|\beta^1-\beta^2\|_2^2+\|\gamma^1-\gamma^2\|_2^2+\|\Lambda_1^1-\Lambda_1^2\|_2^2+\|\Lambda_2^1-\Lambda_2^2\|_2^2$. Then we have that $\inf_{d(\theta,\theta_0)<\epsilon} Pl(\theta, X) > Pl(\theta_0, X)$.

(A4) The mth derivative of $\Lambda_k(\cdot)$, denoted by $\Lambda_k^{(m)}(\cdot)$, is Holder continuous with the exponent η, i.e., $|\Lambda_k^{(m)}(t_1) - \Lambda_k^{(m)}(t_2)| \leq M|t_1 - t_2|^\eta$ for some $\eta \in (0,1]$ and for all $t_1, t_2 \in (l,u)$, $k = 1,2$ and M is some constant. Denote $r = m + \eta$.

(A5) The matrix $E(S_\vartheta S_\vartheta')$ is finite and positive definite, where $\vartheta = (\beta', \gamma')'$ and S_ϑ is defined in the proof of Theorem 3.

Proof of Theorem 1

To establish the consistency using the empirical process theory (van der Vaart and Wellner, 1996), we consider a class of functions $\mathcal{L}_n$ defined by

$$\{l(\theta, X) : \theta \in \Theta_n\}.$$

For any $\theta^1 = (\beta^1, \gamma^1, \Lambda_1^1, \Lambda_2^1)$, $\theta^2 = (\beta^2, \gamma^2, \Lambda_1^2, \Lambda_2^2) \in \Theta_n$, we can easily obtain

$$\begin{aligned}
&|l(\theta^1, X) - l(\theta^2, X)| \\
&\leq K\big(\|\beta^1 - \beta^2\| + \|\gamma^1 - \gamma^2\| + \|\Lambda_1^1 - \Lambda_1^2\|_\infty \\
&\quad + \|\Lambda_2^1 - \Lambda_2^2\|_\infty\big)
\end{aligned}$$

using Taylor's series expansion under conditions (A1) and (A2).

Denote $p_m = 2p + 2(m + k_n)$. By the calculation of van der Vart and Wellner(1996)(p.94), we have

$$\begin{aligned}
&N\Big(\epsilon, \mathcal{L}_n, L_1(P_n)\Big) \\
&\leq N\Big(\frac{\epsilon}{3M}, \mathcal{B}, \|\cdot\|\Big) \cdot N\Big(\frac{\epsilon}{3M_n}, \mathcal{M}_n^1, L_\infty\Big) \\
&\quad \times N\Big(\frac{\epsilon}{3M_n}, \mathcal{M}_n^2, L_\infty\Big) \\
&\leq \Big(\frac{9M}{\epsilon}\Big)^{2p} \cdot \Big(\frac{9M_n^2}{\epsilon}\Big)^{m+k_n} \cdot \Big(\frac{9M_n^2}{\epsilon}\Big)^{m+k_n} \\
&\leq K M^{2p} M_n^{4(m+k_n)} \epsilon^{-p_m}.
\end{aligned}$$

Applying the inequality (31) in Pollard (1984) (p.31), in probability, we have

$$\sup_{\theta \in \Theta_n} \big|P_n l(\theta, X) - P l(\theta, X)\big| \to 0. \tag{A.1}$$

Let $M(\theta, X) = -l(\theta, X)$ and

$$\begin{aligned}
\zeta_{1n} &= \sup_{\theta \in \Theta_n} |P_n M(\theta, X) - PM(\theta, X)|, \\
\zeta_{2n} &= P_n M(\theta_0, X) - PM(\theta_0, X).
\end{aligned}$$

Denote $K_\epsilon = \{\theta : d(\theta, \theta_0) \geq \epsilon, \theta \in \Theta_n\}$. Then we have that

$$\begin{aligned}
&\inf_{K_\epsilon} PM(\theta, X) \\
&\quad = \inf_{K_\epsilon} \Big\{PM(\theta, X) - P_n M(\theta, X) + P_n M(\theta, X)\Big\} \\
&\quad \leq \zeta_{1n} + \inf_{K_\epsilon} P_n M(\theta, X). \qquad \text{(A.2)}
\end{aligned}$$

If $\hat{\theta}_n \in K_\epsilon$, we have

$$\begin{aligned}
&\inf_{K_\epsilon} P_n M(\theta, X) = P_n M(\hat{\theta}_n, X) \\
&\qquad \leq P_n M(\theta_0, X) = \zeta_{2n} + PM(\theta_0, X). \qquad \text{(A.3)}
\end{aligned}$$

By condition (A3), we obtain that $\inf_{K_\epsilon} PM(\theta, X) - PM(\theta_0, X) = \delta_\epsilon > 0$.

By (A.2) and (A.3), we have

$$\inf_{K_\epsilon} PM(\theta, X) \leq \zeta_{1n} + \zeta_{2n} + PM(\theta_0, X) = \zeta_n + PM(\theta_0, X)$$

with $\zeta_n = \zeta_{1n} + \zeta_{2n}$. Hence, we can get that $\zeta_n \geq \delta_\epsilon$. Furthermore, we have $\{\hat{\theta}_n \in K_\epsilon\} \subseteq \{\zeta_n \geq \delta_\epsilon\}$. By (A.1) and Strong Law of Large Numbers, we have $\zeta_{1n} = o(1)$ almost surely, $\zeta_{2n} = o(1)$ almost surely. Therefore, by $\cup_{k=1}^\infty \cap_{n=k}^\infty \{\hat{\theta}_n \in K_\epsilon\} \subseteq \cup_{k=1}^\infty \cap_{n=k}^\infty \{\zeta_n \geq \delta_\epsilon\}$, we complete the proof.

Proof of Theorem 2

To show the convergence rate, for any $\eta > 0$, define the class $\mathcal{F}_\eta = \{l(\theta_{n0}, X) - l(\theta, X) : \theta \in \Theta_n, d(\theta, \theta_{n0}) \leq \eta\}$ with $\theta_{n0} = (\beta_0, \gamma_0, \Lambda_{1n0}, \Lambda_{2n0})$. Following the calculation of Shen and Wong (1994) (p.597), we can establish that $\log N_{[]}(\varepsilon, \mathcal{F}_\eta, \|\cdot\|_2) \leq CN \log(\eta/\varepsilon)$ with $N = 2(m + k_n)$. Moreover, some algebraic calculations lead to $\|l(\theta_{n0}, X) - l(\theta, X)\|_2^2 \leq C\eta^2$ for any $l(\theta_{n0}, X) - l(\theta, X) \in \mathcal{F}_\eta$. Therefore, by Lemma 3.4.2 of van der Vaart and Wellner (1996), we obtain

$$\begin{aligned}
&E_P \|n^{1/2}(P_n - P)\|_{\mathcal{F}_\eta} \\
&\qquad \leq C J_\eta(\varepsilon, \mathcal{F}_\eta, \|\cdot\|_2)\Big\{1 + \frac{J_\eta(\varepsilon, \mathcal{F}_\eta, \|\cdot\|_2)}{\eta^2 n^{1/2}}\Big\}, \qquad \text{(A.4)}
\end{aligned}$$

where $J_\eta(\varepsilon, \mathcal{F}_\eta, \|\cdot\|_2) = \int_0^\eta \{1 + \log N_{[]}(\varepsilon, \mathcal{F}_\eta, \|\cdot\|_2)\}^{1/2} d\varepsilon \leq CN^{1/2}\eta$. The right-hand side of (A.4) yields $\phi_n(\eta) = C(N^{1/2}\eta + N/n^{1/2})$. It is easy to see that $\phi_n(\eta)/\eta$ decreasing in η, and $r_n^2 \phi_n(1/r_n) = r_n N^{1/2} + r_n^2 N/n^{1/2} < 2n^{1/2}$, where $r_n = N^{-1/2} n^{1/2} = n^{(1-\nu)/2}$ with $0 < \nu < 0.5$. Hence $n^{(1-\nu)/2} d(\hat{\theta}, \theta_{n0}) = O_P(1)$ by Theorem 3.2.5 of van der Vaart and Wellner (1996). This, together with $d(\theta_{n0}, \theta_0) = O_p(n^{-r\nu})$ (Lemma A1 in Lu et al. 2007) yields that $d(\hat{\theta}, \theta_0) = O_p(n^{-(1-\nu)/2} + n^{-r\nu})$. The choice of $\nu = 1/(1 + 2r)$ yields the rate of convergence of $d(\hat{\theta}_n, \theta_0) = O_p(n^{-\frac{r}{1+2r}})$.

Proof of Theorem 3

Denote V as the linear span of $\Theta_0 - \theta_0$, where θ_0 denotes the true value of $\theta = (\theta, \gamma, \psi)$ and Θ_0 denote the true parameter space. Let $l(\theta, W)$ be the log-likelihood for a sample of size one and $\delta_n = (n^{-(1-\nu)/2} + n^{-r\nu})$. For any $\theta \in \{\theta \in \Theta_0 : \|\theta - \theta_0\| = O(\delta_n)\}$, define the first order directional derivative of $l(\theta, X)$ at the direction $v \in V$ as

$$\dot{l}(\theta, X)[v] = \frac{dl(\theta + sv, X)}{ds}\Big|_{s=0}, \tag{A.5}$$

and the second order directional derivative as

$$\ddot{l}(\theta, X)[v, \tilde{v}] = \frac{d^2 l(\theta + sv + \tilde{s}\tilde{v}, X)}{d\tilde{s}ds}\Big|_{s=0}\Big|_{\tilde{s}=0} = \frac{d\dot{l}(\theta + \tilde{s}\tilde{v}, X)}{d\tilde{s}}\Big|_{\tilde{s}=0}.$$

Also define the Fisher inner product on the space V as

$$< v, \tilde{v} >= P\Big\{\dot{l}(\theta, X)[v]\dot{l}(\theta, X)[\tilde{v}]\Big\}$$

and the Fisher norm for $v \in V$ as $\|v\|^{1/2} =< v, v >$. Let $\bar{V}$ be the closed linear span of V under the Fisher norm. Then $(\bar{V}, \|\cdot\|)$ is a Hilbert space.

Define the smooth functional of θ as

$$\gamma(\theta) = b_1'\beta + b_2'\gamma,$$

where $b = (b_1', b_2')'$ is any vector of $2p$ dimension with $\|b\| \leq 1$. For any $v \in V$, we denote

$$\dot{\gamma}(\theta_0)[v] = \frac{d\gamma(\theta_0 + sv)}{ds}\Big|_{s=0} = r(v)$$

whenever the right hand-side limit is well defined. Note that $\gamma(\theta) - \gamma(\theta_0) = \dot{\gamma}(\theta_0)[\theta - \theta_0]$. It follows by the Riesz representation theorem that, there exists $v^* \in \bar{V}$ such that $\dot{\gamma}(\theta_0)[v] =< v^*, v >$ for all $v \in \bar{V}$ and $\|v^*\|^2 = \|\dot{\gamma}(\theta_0)\|$.

Let ε_n be any positive sequence satisfying $\varepsilon_n = o(n^{-1/2})$. For any $v^* \in \Theta_0$, by (A4), Corollary 6.21 of Schumaker (1981) (p.227), there exists $\Pi_n v^* \in \Theta_n$ such that $\|\Pi_n v^* - v^*\| = o(1)$ and $\delta_n\|\Pi_n v^* - v^*\| = o(n^{-1/2})$. Also define $g[\theta - \theta_0, X] = l(\theta, X) - l(\theta_0, X) - \dot{l}(\theta, X)[\theta - \theta_0]$. Then by definition of $\hat{\theta}$, we have

$$\begin{aligned}
0 &\leq P_n[l(\hat{\theta}, W) - l(\hat{\theta} \pm \varepsilon_n\Pi_n v^*, W)] \\
&= (P_n - P)[l(\hat{\theta}, W) - l(\hat{\theta} \pm \varepsilon_n\Pi_n v^*, W)] \\
&\quad + P[l(\hat{\theta}, W) - l(\hat{\theta} \pm \varepsilon_n\Pi_n v^*, W)] \\
&= \pm\varepsilon_n P_n\dot{l}(\theta, W)[\Pi_n v^*] \\
&\quad + (P_n - P)\Big\{g[\hat{\theta} - \theta_0, W] - g[\hat{\theta} \pm \varepsilon_n\Pi_n v^* - \theta_0, W]\Big\} \\
&\quad + P\Big\{g[\hat{\theta} - \theta_0, W] - g[\hat{\theta} \pm \varepsilon_n\Pi_n v^* - \theta_0, W]\Big\} \\
&= \mp\varepsilon_n P_n\dot{l}(\theta; W)[v^*] \pm \varepsilon_n P_n\dot{l}(\theta, W)[\Pi_n v^* - v^*] \\
&\quad + (P_n - P)\Big\{g[\hat{\theta} - \theta_0, W] - g[\hat{\theta} \pm \varepsilon_n\Pi_n v^* - \theta_0, W]\Big\} \\
&\quad + P\Big\{g[\hat{\theta} - \theta_0, W] - g[\hat{\theta} \pm \varepsilon_n\Pi_n v^* - \theta_0, W]\Big\} \\
&:= \mp\varepsilon_n P_n\dot{l}(\theta, W)[v^*] + I_1 + I_2 + I_3.
\end{aligned}$$

For I_1, it follows from Conditions (A1)–(A2), Chebyshev inequality and $\|\Pi_n v^* - v^*\| = o(1)$ that $I_1 = \varepsilon_n \times o_p(n^{-1/2})$. For I_2, we have

$$\begin{aligned}
I_2 &= (P_n - P)\Big\{l(\hat{\theta}, W) \\
&\quad - l(\hat{\theta} \pm \varepsilon_n\Pi_n v^*, W) \pm \varepsilon_n\dot{l}(\theta_0, W)[\Pi_n v^*]\Big\} \\
&= \mp\varepsilon_n(P_n - P)\Big\{\dot{l}(\tilde{\theta}, W) - \dot{l}(\theta_0, W)[\Pi_n v^*]\Big\},
\end{aligned}$$

where $\tilde{\theta}$ lies between $\hat{\theta}$ and $\hat{\theta} \pm \varepsilon_n\Pi_n v^*$. By Theorem 2.8.3 in of van der Vaart and Wellner (1996), we know that $\{\dot{l}(\theta; W)[\Pi_n v^*] : \|\theta - \theta_0\| = O(\delta_n)\}$ is Donsker class. Therefore, by Theorem 2.11.23 of van der Vaart and Wellner (1996), we have $I_2 = \varepsilon_n \times o_p(n^{-1/2})$.

For I_3, note that

$$\begin{aligned}
&P(g[\theta - \theta_0, W]) = P\{l(\theta, W) - l(\theta_0, W) - \dot{l}(\theta_0, W[\theta - \theta_0])\} \\
&= 2^{-1}P\{\ddot{l}(\tilde{\theta}, W)[\theta - \theta_0, \theta - \theta_0] - \ddot{l}(\theta_0, W)[\theta - \theta_0, \theta - \theta_0]\} \\
&\quad + 2^{-1}P\{\ddot{l}(\theta_0, W)[\theta - \theta_0, \theta - \theta_0]\} \\
&= 2^{-1}P\{\ddot{l}(\theta_0, W)[\theta - \theta_0, \theta - \theta_0]\} + \varepsilon_n \times o_p(n^{-1/2})
\end{aligned}$$

where $\tilde{\theta}$ lies between θ_0 and θ and the last equation is due to Taylor expansion, (A1)–(A2) and $r > 2$. Therefore,

$$\begin{aligned}
I_3 &= -2^{-1}\{\|\hat{\theta} - \theta_0\|^2 - \|\hat{\theta} \pm \varepsilon_n\Pi_n v^* - \theta_0\|^2\} \\
&\quad + \varepsilon_n \times o_p(n^{-1/2}) \\
&= \pm\varepsilon_n < \hat{\theta} - \theta_0, \Pi_n v^* > +2^{-1}\|\varepsilon_n\Pi_n v^*\|^2 \\
&\quad + \varepsilon_n \times o_p(n^{-1/2}) \\
&= \pm\varepsilon_n < \hat{\theta} - \theta_0, v^* > +2^{-1}\|\varepsilon_n\Pi_n v^*\|^2 + \varepsilon_n \times o_p(n^{-1/2}) \\
&= \pm\varepsilon_n < \hat{\theta} - \theta_0, v^* > +\varepsilon_n \times o_p(n^{-1/2})
\end{aligned}$$

where the last equality holds since $\delta_n\|\Pi_n v^* - v^*\| = o(n^{-1/2})$, Cauchy-Schwartz inequality, and $\|\Pi_n v^*\|^2 \to \|v^*\|^2$. Combing the above facts, together with $P\dot{l}(\theta_0, W[v^*]) = 0$, we can establish that

$$\begin{aligned}
0 &\leq P_n\{l(\hat{\theta}, W) - l(\hat{\theta} \pm \varepsilon_n\Pi_n v^*, W)\} \\
&= \mp\varepsilon_n P_n\dot{l}(\theta_0, W)[v^*] \pm \varepsilon_n < \hat{\theta} - \theta_0, v^* > \\
&\quad + \varepsilon_n \times o_p(n^{-1/2}) \\
&= \mp\varepsilon_n(P_n - P)\{\dot{l}(\theta_0, W)[v^*]\} \pm \varepsilon_n < \hat{\theta} \\
&\quad - \theta_0, v^* > +\varepsilon_n \times o_p(n^{-1/2}).
\end{aligned}$$

Therefore, we obtain $\sqrt{n} < \hat{\theta} - \theta_0, v^* >= \sqrt{n}(P_n - P)\{\dot{l}(\theta_0, W)[v^*]\} + o_p(1) \to N(0, \|v^*\|^2)$, where the asymptotic normality is guaranteed by Central limits Theorem and the the asymptotic variance being equal to $\|v^*\|^2 = \|\dot{l}(\theta_0, W)[v^*]\|^2$. This implies $n^{1/2}(\gamma(\hat{\theta}) - \gamma(\theta_0)) = n^{1/2} < \hat{\theta} - \theta_0, v^* > +o_p(1) \to N(0, \|v^*\|^2)$ in distribution. The semiparametric efficiency can be established by applying Theorem 4 in Shen (1997).

For each component ϑ_q, $q = 1, 2, \cdots, 2p$, we denote by $\psi_q^* = (b_{1q}^*, b_{2q}^*)$ the solution to

$$\inf_{\psi_q^*} E\Big\{l_\vartheta \cdot e_q - l_{b_1^*}[b_{1q}^*] - l_{b_2^*}[b_{2q}^*]\Big\}^2.$$

where $l_\vartheta = (l'_\beta, l'_\gamma)'$, $l_{b_1^*}[b_{1q}^*]$ and $l_{b_2^*}[b_{1q}^*]$ are defined similar to (A.5). Now let $\psi^* = (\psi_1^*, \cdots, \psi_q^*)$. By the calculations of Chen et al. (2006), we have $\|v^*\|^2 = \|\dot\gamma(\theta_0)\| = \sup_{v\in\bar V:\|v\|>0} \frac{|\dot\gamma(\theta_0)[v]|}{\|v\|} = b'\Sigma b$, where $\Sigma = E(S_\vartheta S'_\vartheta)$, $S_\vartheta = \{l_\vartheta - l_{b_1^*} b_1^* - l_{b_2^*} b_2^*\}$. Now, since $b'((\hat\beta - \beta_0)', (\hat\gamma - \gamma_0)') =< \hat\theta - \theta_0, v^* >$, the conclusion of the theorem follows by the Cramér-Wold device.

APPENDIX B. THE DERIVATION OF THE LIKELIHOOD FUNCTION $L(\theta)$

To derive the likelihood function $L(\theta)$, note that given covariates, if $\delta = 1$, we have

$$\begin{aligned} & P[L < T < R, W \in (w, w+dw)] \\ = \; & P[L < T < R | W \in (w, w+dw)] P[W \in (w, w+dw)] \\ = \; & P[L < T < R | W \in (w, w+dw)] f_W(w) dw \\ = \; & [m_\alpha(F_T(R), F_W(w)) - m_\alpha(F_T(L), F_W(w))] f_W(w) dw. \end{aligned}$$

Furthermore, for $\delta = 0$, we have $R = \infty$ and $P(T > L, W > \zeta - L) = 1 - F_W(\zeta - L) - F_T(L) + C_\alpha(F_T(L), F_W(\zeta - L))$. It then follows that the likelihood function of θ based on a single observation (δ, L, W, Z) has the form

$$\Big\{[m_\alpha(F_T(R), F_W(W)) - m_\alpha(F_T(L), F_W(W))] f_W(W)\Big\}^\delta \times \Big\{1 - F_W(\zeta - L) - F_T(L) + C_\alpha(F_T(L), F_W(\zeta - L))\Big\}^{1-\delta}.$$

Hence the likelihood function of θ based on an i.i.d. sample $(\delta_i, L_i, W_i, Z_i)$ has the form

$$\begin{aligned} & L(\theta) \\ & = \prod_{i=1}^n \Big\{[m_\alpha(F_T(R_i), F_W(W_i)) \\ & \quad - m_\alpha(F_T(L_i), F_W(W_i))] f_W(W_i)\Big\}^{\delta_i} \\ & \times \Big\{1 - F_W(\zeta - L_i) - F_T(L_i) \\ & \quad + C_\alpha(F_T(L_i), F_W(\zeta - L_i))\Big\}^{1-\delta_i}. \end{aligned}$$

Received 26 February 2014

REFERENCES

Betensky, R. A. and Finkelstein, D. M. (2002). Testing for dependence between failure time and visit compliance with interval-censored data. *Biometrics*, **58**, 58–63. MR1891043

Betensky, R. A, Rabinowitz, D. and Tsiatis, A. A. (2001). Computationally simple accelerated failure time regression for interval censored data. *Biometrika*, **88**, 703–711. MR1859403

Chen, D. G. D., Sun, J. and Peace, K. E. (2012). *Interval-censored time-to-event data: methods and applications.* Chapman & Hall/CRC. MR3053014

Chen, L. and Sun, J. (2010). A multiple imputation approach to the analysis of interval-censored failure time data with the additive hazards model. *Computational Statistics & Data Analysis*, **54**, 1109–1116. MR2580942

Chen, X., Fan, Y. and Tsyrennikov, V. (2006). Efficient Estimation of Semiparametric Multivariate Copula Models. *Journal of the American Statistical Association*, **101**, 1228–1240. MR2328309

Finkelstein, D. M. (1986). A proportional hazards model for interval-censored failure time data. *Biometrics*, **42**, 845–854. MR0872963

Finkelstein, D. M., Goggins, W.B. and Schoenfeld, D. A. (2002). Analysis of failure time data with dependent interval censoring. *Biometrics*, **58**, 298–304. MR1908169

Goggins, W. B. and Finkelstein, D. M. (2000). A proportional hazards model for multivariate interval-censored failure time data. *Biometrics*, **56**, 940–943.

Hougaard, P. (2000). *Analysis of multivariate survival data.* Springer-Verlag: New York. MR1777022

Huang, J. and Rossini, A. J. (1997). Sieve estimation for the proportional odds failure-time regression model with interval censoring. *Journal of the American Statistical Association*, **92**, 960–967. MR1482126

Huang, X. and Wolfe, R. A. (2002). A frailty model for informative censoring. *Biometrics*, **58**, 510–520. MR1925548

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The statistical analysis of failure time data.* Second edition, John Wiley: New York. MR1924807

Li, J., Wang, C. and Sun, J. (2012). Regression analysis of clustered interval-censored failure time data with the additive hazards model. *Journal of Nonparametric Statistics*, **24**, 1041–1050. MR2995491

Lin, D. Y., Oakes, D. and Ying, Z. (1998). Additive hazards regression with current status data. *Biometrika*, **85**, 289–298. MR1649115

Lin, D. Y. and Ying, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika*, **81**, 61–71. MR1279656

Lu, M., Zhang, Y. and Huang, J. (2007). Estimation of the mean function with panel count data using monotone polynomial splines. *Biometrika*, **94**, 705–718. MR2410018

Martinussen, T. and Scheike, T. H. (2002). Efficient estimation in additive hazards regression with current status data. *Biometrika*, **89**, 649–658. MR1929169

Nelsen, R. B. (2006). *An introduction to copulas*, Second Edition. Springer: New York. MR2197664

Pollard, D. (1984). *Convergence of stochastic processes.* Springer-Verlag: New York. MR0762984

Ramsay, J. O. (1988). Monotone regression splines in action. *Statistical Science*, **3**, 425–441.

Schumaker, L. L. (1981). *Spline functions: basic theory.* John Wiley: New York. MR0606200

Shen, X. (1997). On methods of sieves and penalization. *The Annals of Statistics*, **25**, 2555–2591. MR1604416

Shen, X. and Wong, W. H. (1994). Convergence rate of sieve estimates. *The Annals of Statistics*, **22**, 580–615. MR1292531

Sun, J. (2006). *The statistical analysis of interval-censored failure time data.* Springer: New York. MR2287318

Sun, J. and Li, J. (2013). *Interval Censoring. Handbook of Survival Analysis*, Editors: Klein, J. P., van Houwelingen, H. C., Ibrahim, J. G. and Scheike, T. H., Chapman & Hall/CRC, 369–390. MR3053014

Tong, X., Hu, T. and Sun, J. (2012). Efficient estimation for additive hazards regression with bivariate current status data. *Science China Mathematics*, **55**, 763–774. MR2903461

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes.* Springer-Verlag: New York. MR1385671

Wang, L., Sun, J. and Tong, X. (2010). Regression analysis of case II interval-censored failure time data with the additive hazards model. *Statistica Sinica*, **20**, 1709–1723. MR2777342

ZHENG, M. and KLEIN, J. P. (1995). Estimates of marginal survival for dependent competing risk based on an assumed copula. *Biometrika*, **82**, 127–138. MR1332844

ZHANG, Z., SUN, L., SUN, J. and FINKELSTEIN, D. M. (2007). Regression Analysis of Failure Time Data with Informative Interval Censoring. *Statistics in Medicine*, **26**, 2533–2546. MR2361363

ZHOU, X. and SUN, L. (2003). Additive hazards regression with missing censoring information. *Statistica Sinica*, **13**, 1237–1257. MR2026071

Shishun Zhao
College of Mathematics
Jilin University
Changchun 130012
P.R. China
E-mail address: zhaoss@jlu.edu.cn

Tao Hu
School of Mathematical Sciences & BCMIIS
Capital Normal University
Beijing 100048
P.R. China
E-mail address: hutaomath@foxmail.com

Ling Ma
Department of Statistics
University of Missouri
146 Middlebush Hall, MO 65211
USA
E-mail address: lingma@missouri.edu

Peijie Wang
College of Mathematics
Jilin University
Changchun 130012
P.R. China
E-mail address: wangpj19890114@163.com

Jianguo Sun
Department of Statistics
University of Missouri
146 Middlebush Hall, MO 65211
USA
E-mail address: sunj@missouri.edu