Ann Inst Stat Math (2012) 64:1–25
DOI 10.1007/s10463-010-0301-x

Semiparametric efficient inferences for lifetime regression model with time-dependent covariates

Hsieh Fushing

Received: 5 December 2008 / Revised: 17 December 2009 / Published online: 11 June 2010

Abstract Through a threshold equation, we propose a time-transformed accelerated failure time (AFT) model with time-dependent covariate history in survival analysis. This model contains a general class of semiparametric lifetime regression models, including AFT with identical time-scale and a wide spectrum of Cox's hazard regression models and their frailty variants. We first construct the semiparametric efficient statistical inferences on the AFT model with identical time-scale. The theoretical semiparametric Fisher information bound is explicitly derived under right-censored data setting. And the overidentified estimating equation (OEE) approach based on two martingale processes is shown to achieve this semiparametric efficiency bound. Extensions of the semiparametric efficient statistical inferences to the time-transformed AFT versions are also discussed. We also conclude that most log-rank estimating equations would suffer severe information loss primarily caused by wiggling pattern of the baseline hazard function, while the OEE approach can alleviate the damaging effects. A simulated biological life history example is numerically studied.

Keywords Accelerated failure time model · Log-rank estimating equation · Martingale central limit theorem · Overidentified estimating equation (OEE) approach · Transformation model

This research was supported in part by the NSF under Grant No:BCS 0527766 and HSD 0826844, and by the NIH under Grant No: 1R01AG025218-01A2.

H. Fushing (✉)
Department of Statistics, Mathematical Sciences Building,
UC Davis, Davis, CA 95616, USA
e-mail: fushing@wald.ucdavis.edu

1 Introduction

The primary goal of survival analysis is to foster better understanding into the dynamics under study by inferring the relationship between the event of interest and the continuous progress or development preceding the occurrence of the event. For instance, scientists investigate AIDS dynamics from HIV infection to the diagnosis of AIDS disease, or dyadic dynamics from marriage to divorce, or aging dynamics from birth through death. Along the temporal axis of the dynamic system, the continuous measurements leading to the designated event is generically called time-dependent covariate history. Although this covariate history is an indisputable manifestation of what is responsible for the event, two rather distinct viewpoints regarding its effects are taken in statistical literature.

The first viewpoint is that it has instantaneous impacts to make the event of interest more (or less) likely to occur at any point in time (Aalen and Gjessing 2001). This perspective is well captured in the classic Cox (1972) hazard regression models and its variants with or without frailty, see Andersen et al. (1993). The second viewpoint is that the covariate history only influences the underlying mechanisms in such a way that it gradually speeds up (or slows down) the event's occurrence in the future. The second viewpoint seems to better seize the developmental characteristics of biological and physical dynamics. The semiparametric AFT model with time-dependent covariate proposed in Cox and Oakes (1984) genuinely captures this viewpoint.

Nowadays these two viewpoints still widely separate researchers in different areas of sciences. For example, researchers in industry reliability and engineering make heavy use of parametric time-independent version of AFT model, see Escobar and Meeker (1992) and reference therein, while majority of researchers in biology and medicine and social sciences apply Cox's hazard regression model. In fact there is a bit of concern that the great popularity of hazard regression model might have been mainly attribute to the simplicity of the partial likelihood approach for statistical inferences, rather than its appropriateness of scientific interpretation.

One way of addressing this concern is to construct a unified framework that simultaneously accommodate these two sharply contrasting effects of time-dependent covariate history. This unified framework would provide a platform for scientists who can probe various modeling conditions and link them with biological and physical mechanisms. By doing such explorations, scientists are likely to achieve better statistical modeling and to arrive with proper and accurate interpretations. In this paper, we construct such a unified modeling framework, called time-transformed AFT model, through the following threshold equation:

$$U = \int_0^T e^{\beta' Z(t)} \, \mathrm{d}\Lambda_0(t) \tag{1}$$

where T and $Z(t)$ are the observable survival time and time-dependent covariate, respectively, on time scale t. And $\Lambda_0(t) = \int_0^t \lambda_0(s)\mathrm{d}s$ is a monotone, but unknown transformation on the time-scale t. Here U is more than a pure positive random noise. It denotes an unobservable threshold that represent a key random biological quantity associated with the dynamics. It is observable only when $Z(t) = 0$ for all t. Its unknown distribution is denoted by $F(u) = \int_0^u f(v)\mathrm{d}v$ and its cumulative hazard function by $\Lambda_U(u) = \int_0^u \lambda_U(v)\mathrm{d}v$ on a hypothetical time scale u.

From mechanistic perspective, the acceleration factor $e^{\beta' Z(t)}\lambda_0(t)$ in Eq. (1) is modeling the rate of using up the threshold quantity U to lead to the observable event at time T. This threshold type of equation is ubiquitous in biological and physical systems.

From statistical modeling perspective, it is an extension of semiparametric linear regression models with capacity of accommodating time-dependent covariates. To see this, let the covariate $Z(t) \equiv Z$ be time-independent; then Eq. (1) becomes a classic semiparametric linear regression model with an unknown transformation on the dependent variable as $H(T) = -\beta' Z + \log U$, where the monotone transformation $H(.) = \log\{\Lambda_0(.)\}$ and $\log U$ is the error term, see also Horowitz (1996). This simple framework indeed includes: the accelerated failure time (AFT) model, or so-called linear regression with censored data (when H is known, see e.g. Tsiatis 1990; Wei et al. 1990; Ritov 1990; Lai and Ying 1991; Ying 1993; Hsieh 1997), Cox's proportional hazard model (when U is exponential distributed, see e.g. Cox 1972), frailty model (when $U = U_1/U_2$ and U_1 is exponential distributed, while U_2 has a frailty distribution, see e.g. Vaupel et al. 1979; Heckman and Singer 1984; Hougaard 1986; Nielsen et al. 1992), and transformation models (when U is known distributed according to Normal, Logistic or other distributions, see e.g. Box and Cox 1964; Bickel and Doksum 1981; Dabrowska and Doksum 1988; Hsieh 1995, 1996a; Murhpy et al. 1997).

Further by denoting the covariate history up to time t as $\bar{Z}(t) = \{z(s) : 0 < s < t\}$, the cumulative hazard rate of the observed survival time T under model (1) is calculated as:

$$\Lambda(t; \beta, \bar{Z}) = \Lambda_U \left[\int_0^t e^{\beta' Z(s)} \lambda_0(s)\, \mathrm{d}s \right]. \tag{2}$$

Again, if $\Lambda_U(t) \equiv t$, then we have Cox's model with time-dependent covariate. Therefore, from Eqs. (1) and (2), all the time-dependent extended versions of aforementioned models with time-independent covariate are unified under one single framework of the time-transformed AFT model. In this way the two kinds of effect of time-dependent covariate history discussed above is accommodated in one unified framework.

It is noted that the simplest model in the class specified by Eq. (1) is the following identical time-scale version as:

$$U = \int_0^T e^{\beta' Z(s)} \mathrm{d}s, \tag{3}$$

with $\Lambda_0(t) \equiv t$. This simple AFT model indeed plays a rather fundamental role in the following way. The unknown transformation $\mathrm{d}\Lambda_0(t) = \lambda(t)\mathrm{d}t$ in Eq. (1) can be absorbed into the accelerating factor. Hence the time-transformed AFT model is re-expressed as:

$$U = \int_0^T e^{\beta' Z(t) + g(t)} \mathrm{d}t \tag{4}$$

with $g(t) = \log \lambda_0(t)$. Furthermore, if $g(t) = \sum a_j \varphi_j(t)$ with $\{\varphi_j(.), j = 1, \dots n\}$ being a known basis of functions on $[0, \infty)$, then we arrive at an equation as:

$$U = \int_0^T e^{\beta' Z(t) + \sum a_i \psi_i(t)} \mathrm{d}t. \tag{5}$$

This fact very importantly indicate that modeling equations (1) and (2) and all transformation models and their variants can be rewritten or approximated by the identical time-scale version of AFT in Eq. (3). Thus theoretical and computational developments for statistical inferences on the time-transformed AFT model of Eq. (1) can be very well achieved by the less complex developments for the simpler equation (3). This distinct perspective offers a resolution to statistical inferences in semiparametric transformation models, which is still plunged by well-known theoretical and computational difficulties, see Bickel and Doksum (1981), Dabrowska and Doksum (1988), and Murhpy et al. (1997).

The identical time-scale version of AFT model in (3) was first proposed in Cox and Oakes (1984). Its statistical inference was first pioneered in Robins and Tsiatis (1992) using log-rank estimating equations. They indicated that their estimating equations are efficient when the baseline hazard $\lambda_U(u)$ is constant. They also speculated that, under general setting, the semiparametric efficiency could be achieved by combining increasing number of estimating equations as sample size n increases, see also Lin and Ying (1995). However, Hsieh (1997) pointed out that log-rank estimating equation could have severe information loss primarily due to the possible wiggling patterns of $\lambda_U(u)$ even under the time-independent covariate setting. To avoid such a information loss, the weighting function built into the log-rank estimating equation must adapt to precise locations of sign-changes of the ratio $\lambda'_U(u)/\lambda_U(u)$. This is not an easy task. Certainly this information loss problem remains for log-rank estimating equations under the time-dependent covariate setting with even more profound computational and theoretical complexities. Thus so far semiparametric efficient inferences for this model (3) are still missing in survival analysis literature.

In this paper, we provide a comprehensive account of semiparametric efficient inferences on model (3). The semiparametric Fisher information bound is explicitly calculated and the efficient scoring function is proved in Sect. 2. The computational weakness of log-rank estimating equation, specifically called local confounding, is demonstrated in Sect. 3. In Sect. 4, the overidentified estimating equation (OEE) approach, see Hsieh (1997, 2001), is applied to achieve the semiparametric efficiency bound. Several extensions to general settings are also briefly mentioned. In Sect. 5, a numerical example of evaluating the effect of reproductive costs on female's survival is studied. The purpose of this numerical study is twofold. First, it illustrates applicability of (1) and (3) in one central issue of life history theory in aging biology. Second, several regression parameters estimators are compared to manifest the effect of local confounding on log-rank estimating equations. Discussions of related issues are collected in the last section. All proofs of theoretical results and regularity conditions are stated in "Appendix".

2 Fisher information bound in AFT model with right censored data

In this section, the efficient score function and the semiparametric Fisher information bound are calculated for the regression parameter β under the AFT model in Eq. (3) with right censored data. For expositional simplicity, let β be a real value parameter from here on. Results of inferences for a vector value parameter can be similarly derived.

Let $(X_i, \delta_i, \bar{Z}_i(X_i)), i = 1, \ldots, n$ be n independent copies of observable random vectors with possibly censored survival time $X_i = (T_i \wedge C_i)$ and the censoring indicator $\delta_i = 1$ if $X_i = T_i$, and 0 otherwise, and $\bar{Z}_i(X_i)$ the covariate history only observed up to X_i. Here the uncensored survival time T_i is generated by solving the Eq. (3) given U_i and the covariate history $\bar{Z}_i(t) = \{z_i(s) : 0 < s < t\}$ up to $t \le T_i$. We further assume that $C_i, \ i = 1, \ldots, n$, are independent and identically distributed, and, for each i, C_i is independent of T_i given $\bar{Z}_i(t)$ for any t. This identically distribution condition is made only for simplicity in asymptotic arguments. It could be further relaxed.

The following notations are used throughout this paper:

$$\psi\{\bar{Z}_i(t), \beta\} = \int_0^t e^{\beta Z_i(s)} \mathrm{d}s,$$
$$U_i(\beta) = \psi\{\bar{Z}_i(T_i), \beta\}, \quad V_i(\beta) = \psi\{\bar{Z}_i(X_i), \beta\}.$$

That is, if T_i is not censored, then the otherwise uncomputable baseline survival time $U_i(\beta)$ is equal to $V_i(\beta)$. And from here on the subscript U for the baseline hazard rate $\lambda_U(u)$ is suppressed, and is denoted simply by $\lambda(u)$.

Furthermore, let β_0 denote the true value of β. Then $U_i(\beta_0) = U_i$ and $V_i(\beta_0) = V_i$. These two simple observations give rise to the heuristic idea that if β takes the value β_0 in the $\psi(.)$ transformation, then the hazard rate of $V_i(\beta_0)$ would be equal to the baseline hazard rate, $\lambda(u)$, and would be independent of the covariate history of $Z_i(t)$ up to $\psi^{-1}(X_i, \bar{Z}_i, \beta_0)$; that is, $\bar{Z}_i(\psi^{-1}(X_i, \bar{Z}_i(X_i), \beta_0))$, see Robins and Tsiatis (1992) for more detailed causal interpretations regarding this independence.

To setup an estimating equation, we calculate the hazard rate of the transformed variable $V_i(\beta)$ for any given β, instead of that of X_i. Under the AFT model, the survival probability of $U_i(\beta)$ is calculated as

$$Pr\{U_i(\beta) > u\} = Pr\{T_i > \psi^{-1}(u, \bar{Z}_i, \beta)\},$$
$$= Pr[U_i > \psi\{\bar{Z}_i(\psi^{-1}(u, \bar{Z}_i, \beta)), \beta_0\}]$$
$$= e^{-\Lambda(\psi\{\bar{Z}_i(\psi^{-1}(u,\bar{Z}_i,\beta)),\beta_0\})}.$$

Further denote the corresponding hazard rate $\partial\Lambda(\psi\{\bar{Z}_i(\psi^{-1}(u, \bar{Z}_i, \beta)), \beta_0\})/\partial u$ by $\lambda_{i,\beta}(u)$ with

$$\lambda_{i,\beta}(u) = \lambda(\psi\{\bar{Z}_i(\psi^{-1}(u, \bar{Z}_i, \beta)), \beta_0\}) \exp\{(\beta_0 - \beta) Z_i(\psi^{-1}(u, \bar{Z}_i, \beta))\}. \quad (6)$$

In Robins and Tsiatis (1992), a scoring function for the ith subject is defined on the baseline time scale, u, as $G_i(u,\beta) = g[\bar{Z}_i\{\psi^{-1}(u,\bar{Z}_i,\beta)\}]$, where $g(.)$ is a real-valued function of covariate history up to the corresponding real time $t = \psi^{-1}(u,\bar{Z}_i,\beta)$. Typical examples of $g\{\bar{Z}_i(t)\}$ might include cumulative exposure, $\int_0^t Z_i(u)\mathrm{d}u$, or its current value, $Z(t)$. In this paper, we mainly consider the following two scoring functions:

$$R_i(u,\beta) = \int_0^{\psi^{-1}(u,\bar{Z}_i,\beta)} Z_i(v)\exp\{\beta Z_i(v)\}\mathrm{d}v$$
$$Z_i(u,\beta) = Z_i\{\psi^{-1}(u,\bar{Z}_i,\beta)\}.$$

For convenience, define:

$$Y(u,\beta) = \sum_1^n Y_i(u,\beta), \quad Y^K(u,\beta) = \sum_1^n K_i(u,\beta)Y_i(u,\beta),$$

where $K(u,\beta)$ is any possible scoring function, and $Y_j(u,\beta) = I\{V_j(\beta) \geq u\}$ indicates whether the jth individual is still at risk in terms of its estimated baseline survival time $V_j(\beta)$.

With a chosen scoring function, $G_i(u,\beta)$, the log-rank estimating equation considered in Robins and Tsiatis (1992) is $S(\beta) = 0$ with

$$S(\beta) = \sum_1^n \Delta_i[G_i\{V_i(\beta),\beta\} - G^{av}\{V_i(\beta),\beta\}],$$
$$G^{av}\{u,\beta\} = \frac{Y^G(u,\beta)}{Y(u,\beta)}$$

being the average scores values among those individuals who are still at risk at time u; that is, for those j, $Y_j(u,\beta) = 1$.

A more convenient expression of the above log-rank estimating equation is in terms of a counting process representation. That is, let $N_i(u;\beta)$ be the counting process for the i-th individual defined as: $N_i(u;\beta) = I\{V_i(\beta) \leq u, \delta_i = 1\}$, with its random intensity process derived from (6) as $\lambda_{i,\beta}(u,\beta)Y_i(u,\beta)$. Then a log-rank estimating equation truncated at a prespecified time point, u, is

$$S(u,\beta) = \sum_1^n \int_0^u \{G_i(v,\beta) - G^{av}(v,\beta)\}\mathrm{d}N_i(v;\beta). \tag{7}$$

Therefore, the estimating equation $S(\beta)$ originally proposed in Robins and Tsiatis (1992) is equal to $S(\infty,\beta)$ without imposing a finite truncation. Below the efficient score function, says G_{op} is derived and the detailed proof is given in "Appendix".

Theorem 1 *Under the AFTM model specified in* (3) *with right censored data, the efficient score function is calculated as*

$$\int_0^{\infty} I_{\{Q_i(\beta)\geq v\}}[G_{OP,i}(v,\beta)-\mathcal{E}[G_{OP,i}(v,\beta)|V_i(\beta)\geq u]]\mathrm{d}M_i(v),$$
$$=\int_0^{\infty}[G_{OP,i}(v,\beta)-\mathcal{E}[G_{OP,i}(v,\beta)|V_i(\beta)\geq u]]\mathrm{d}M_{i,uc}(v),$$

where

$$Q_i(\beta)=\int_0^{C_i}\exp\{\beta Z_i(v)\}\mathrm{d}v,\quad G_{OP,i}(u,\beta)=R_i(u,\beta)\frac{\lambda'(u)}{\lambda(u)}+Z_i(u,\beta),$$
$$\mathcal{E}[G_{OP,i}(u,\beta)|V_i(\beta)>u]=\mathcal{E}[R_i(u,\beta)|V_i(\beta)>u]\frac{\lambda'(u)}{\lambda(u)}$$
$$+\mathcal{E}[Z_i(u,\beta)|V_i(\beta)>u],$$

and martingales

$$M_i(u)=I_{\{U_i\leq u\}}-\int_0^u I_{\{U_i\geq v\}}\mathrm{d}\Lambda(v),\quad M_{i,uc}(u)=I_{\{V_i\leq u,\delta_i=1\}}-\int_0^u I_{\{V_i\geq v\}}\mathrm{d}\Lambda(v).$$

It is noted that Theorem 1 contains the finite-sum version of the time-transformed AFT model (5) by including the functions $\varphi_j(.)$ as components of multivariate time-dependent covariate $Z(t)$.

With the efficient score function given in Theorem 1, the semiparametric Fisher information of β contained in the data of the ith subject is, again by martingale calculations,

$$I_i(\beta_0)=\mathcal{E}\int_0^{\infty} I_{\{Q_i\geq u\}}[G_{OP,i}(v,\beta_0)-\mathcal{E}[G_{OP,i}(v,\beta_0)|Q_i\geq u]]^2 I_{\{U_i\geq v\}}\mathrm{d}\Lambda(v),$$
$$=\mathcal{E}\int_0^{\infty} I_{\{Q_i\geq u\}}[G_{OP,i}(v,\beta_0)-\mathcal{E}[G_{OP,i}(v,\beta_0)|Q_i\geq u]]^2\mathrm{d}F(v),$$
$$=\int_0^{\infty}\mathcal{E}[1_{\{Q_i\geq u\}}]\{G_{OP,i}(v,\beta_0)-\mathcal{E}[G_{OP,i}(v,\beta_0)|V_i(\beta_0)\geq u]\}^2\mathrm{d}F(v).$$

Furthermore, this semiparametric Fisher information bound suggests the efficient estimating equation is

$$0=\sum_1^n\int_0^{\infty}[G_{OP,i}(v,\beta)-G^{av}_{OP,i}(v,\beta)]\mathrm{d}N_i(v;\beta).$$

This estimating equation seems not practically feasible because $G_{OP,i}(u,\beta)$ involves with the ratio of unknown baseline hazard and its derivative, $\lambda'(u)/\lambda(u)$.

Very naively smoothing techniques, such as in Ramlau-Hansen (1983) and many others, could be applied for estimating this ratio, and then plug it in to construct an approximate efficient estimating equation.

Unfortunately the problem generated by the presence of this ratio has a subtlety that could not be resolved nicely by this simple-minded strategy. A clear picture of this subtlety in computation is given in the next section. An brief and intuitive idea of it is that, due to the sensitivity in bandwidth selection in estimating the ratio $\lambda'(u)/\lambda(u)$, it is very difficult to make the estimated ratio and the true ratio to have synchronous sign changes. Without such synchronization, the information of the regression parameter β is deemed to lose. And the severity of information loss essentially depends on the structure of $\lambda(u)$. This phenomenon is called local confounding in Hsieh (1997).

3 Effects of local confounding on log-rank estimating equations

In this section, we discuss how the phenomenon of local confounding could possibly arise, and analytically show its effect on the log-rank estimating equations. Heuristic comments regarding to smoothing techniques to this phenomenon and a remedy are also discussed.

Recall the log-rank estimating equation (7) and consider following approximations of the difference $S(u;\beta) - S(u;\beta_0)$ given β within a close vicinity of β_0:

$$
\begin{aligned}
&S(u;\beta) - S(u;\beta_0)\\
&\approx \sum_1^n \int_0^u \{G_i(v,\beta_0) - G^{av}(v,\beta_0)\}[\lambda_{i,\beta}(v) - \lambda(v)]Y_i(v,\beta_0)\mathrm{d}v,\\
&\approx (\beta-\beta_0)\sum_1^n \int_0^u \sum_1^n \{G_i(v,\beta_0) - G^{av}(v,\beta_0)\}\{\mathrm{d}\lambda_{i,\beta_0}(v)/\mathrm{d}\beta\}Y_i(v,\beta_0)\mathrm{d}v,\\
&\approx n(\beta-\beta_0)[1/n\int_0^v \sum_1^n \{G_i(v,\beta_0) - G^{av}(v,\beta_0)\}\\
&\quad \times\{R_i(v,\beta_0)\lambda'(v) + Z_i(v,\beta_0)\lambda(v)\}Y_i(v,\beta_0)\mathrm{d}v],
\end{aligned}
$$

where both $R_i(u,\beta)$ and $Z_i(u,\beta)$ are defined in the last section. The first approximating equation is obtained by canceling off two nearly identical zero-mean martingales, see also Tsiatis (1990). And the remaining approximations are so called the locally asymptotic linearities. They are rigorously stated in the following theorem under the regularity conditions on bounded variations of the covariates $z_i(t)$ and on tail behaviors of the baseline density function $f(u)$ given in "Appendix".

Theorem 2 *Let*

$$
\bar{M}_1(u;\beta) = \sum_1^n \int_0^u \left\{R_i(v,\beta) - \frac{\mathcal{E}Y^R(v,\beta)}{\mathcal{E}Y(v,\beta)}\right\} \mathrm{d}\mathcal{E}N_i(v,\beta),
$$

$$\bar{M}_2(u;\beta) = \sum_1^n \int_0^u \left\{ Z_i(v,\beta) - \frac{\mathcal{E}Y^Z(v,\beta)}{\mathcal{E}Y(v,\beta)} \right\} \mathrm{d}\mathcal{E}N_i(v,\beta),$$

$$A_{1,n}(u) = n^{-1} \sum_1^n \int_0^u \left\{ R_i(v,\beta) - \frac{\mathcal{E}Y^R(v,\beta)}{\mathcal{E}Y(v,\beta)} \right\} \times \left\{ Z_i(v,\beta) + \frac{\lambda'(v)}{\lambda(v)} R_i(v,\beta) \right\} L_i(v,\beta)\,\mathrm{d}F(v),$$

$$A_{2,n}(u) = n^{-1} \sum_1^n \int_0^u \left\{ Z_i(v,\beta) - \frac{\mathcal{E}Y^Z(v,\beta)}{\mathcal{E}Y(v,\beta)} \right\} \times \left\{ Z_i(v,\beta) + \frac{\lambda'(v)}{\lambda(v)} R_i(v,\beta) \right\} L_i(v,\beta)\,\mathrm{d}F(v),$$

with $L_i(v,\beta) = Pr\{Q_i(\beta) > v\}$.

Under regularity conditions (C1) to (C4), the martingale processes $M_i(u;\beta),\ i = 1,2$, *has the following strong approximations which hold uniformly for* $u \in [0, u_{k(n)}]$ *with* $u_{k(n)}$ *going to* ∞ *in a rather slow rate, such as* $\log n^\eta$ *for some positive* η.

(i) *For any fixed* $B > 0$ *and* $\varepsilon > 0$

$$\sup_{|\beta-\beta_0|\le B} \left\{ |M_i(u;\beta) - \bar{M}_i(u;\beta) \right\} = o(n^\varepsilon) \text{ almost surely;}$$

(ii) *and for each* j *and any positive sequence* $d_n \to 0$ *almost surely*

$$\sup_{|\beta-\beta_0|\le d_n} \left\{ |\Delta_j M_i(u;\beta) - \Delta_j M_i(u;\beta_0) - n\Delta_j A_{i,n}(u)(\beta-\beta_0)|/(n^{1/2} + n|\beta-\beta_0|) \right\} = o(n^\varepsilon)$$

almost surely (*or with great probability*).

Further denote, for any scoring function G and K,

$$A_K^G(u,\beta) = \lim 1/n \int_0^u \sum_1^n \{G_i(v,\beta) - G^{av}(v,\beta)\} K_i(v,\beta) Y_i(v,\beta)\mathrm{d}v,$$

and

$$\rho_G(u;\beta,\lambda) = \int_0^u [A_R^G(v,\beta)\lambda'(v) + A_Z^G(v,\beta)\lambda(v)]\mathrm{d}v.$$

Then, with Theorem 2 and the approximations given before it, we have the locally asymptotic linearity expressed as:

$$\begin{aligned} S(u;\beta) - S(u;\beta_0) &\approx n[\beta-\beta_0] \int_0^u [A_R^G(v,\beta_0)\lambda'(v) + A_Z^G(v,\beta_0)\lambda(v)]\,\mathrm{d}v, \\ &= n[\beta-\beta_0]\rho_G(u;\beta_0,\lambda). \end{aligned} \tag{8}$$

Let $\hat{\beta}_u^{\text{TLN}}$ be the β solving the truncated long-rank estimating equation $S(u; \beta) = 0$. Since the equation $S(u; \beta) = 0$ is monotonic in β, $\hat{\beta}_u^{\text{TLN}}$ is unique in the sense that the set of generalized solutions is convex, see also Fygenson and Ritov (1994). With the locally asymptotic linearity (8), Rebolledo's martingale central limit theorem on $S(u; \beta_0)$ renders that $\sqrt{n}(\hat{\beta}_u^{\text{TLN}} - \beta_0)$ converges in distribution to a zero mean normal distribution with variance $\sigma_G^2(u)$:

$$\sigma_G^2(u) = \{\rho_G(u; \beta_0, \lambda)\}^{-2} H_G(u),$$
$$H_G(u) = \lim_{n\to\infty} 1/n \int_0^u \sum_1^n \{G_i(v, \beta_0) - G^{av}(v, \beta_0)\}^2 Y_i(v, \beta_0)\lambda(v)\,\mathrm{d}v.$$

It is essential to note that $(\rho_G(u; \beta_0, \lambda))^2$ is not necessary increasing in u, while $H_G(u)$ is surely increasing. Therefore the asymptotic variance $\sigma_G^2(u)$ might not be decreasing in u. This phenomenon implies that, by making the truncating time u larger to incorporate more uncensored data points into a log-rank estimating equation, we might end up losing more information. This is counter-intuitive. For this reason, the original proposal of untruncated log-rank estimating equation $S(\infty, \beta) = 0$ in Robins and Tsiatis (1992) could be very inefficient, when $\lambda(u)$ indeed wiggles up and down along its time scale u. This phenomenon called local confounding. Such phenomenon was studied in Hsieh (1997) for censored linear regression with time-independent covariates.

Apparently the non-monotonicity of function $(\rho_G(u; \beta, \lambda))^2$ is caused by either the presence of the derivative function $\lambda'(u)$, or the time-dependent interaction terms between the scoring functions $G(.)$ and $R(.)$ and $Z(.)$. As for eliminating the possible effects coming from latter interactions, we simply choose the scoring function $G(.)$ either $R(.)$ or $Z(.)$.

But, no matter which single scoring function is chosen, the local confounding still could have its effects. Since the ratio $\lambda'(u)/\lambda(u)$ is built in the locally asymptotic linearity (8). A simple-minded strategy toward this problem is to apply the smoothing techniques; for example, in Ramlau-Hansen (1983), for estimating $\lambda(u)$ as well as its derivative $\lambda'(u)$, and then plug in an estimated ratio into the efficient score to result an estimated efficient score equation. However, in order for this plug-in estimating equation to work against the local confounding, the estimated and the true ratios must have synchronized sign changes. Without such synchronization, the effects of local confounding is deemed to cause information loss.

It is known that hardly such guarantee could be granted from any smoothing technique to achieve this synchronous behavior. This difficulty resides on two types of sensitivity: one is a reasonably robust initial estimate of β; the other is a suitable bandwidth. The first choice is in general an issue in the semiparametric literature. And the second choice is clearly crucial. Since, on one hand, under-smoothing would automatically cause information loss due to the frequent sign changes of the estimated $\lambda'(u)$; on the other hand, over-smoothing would miss the true sign changes of $\lambda'(u)$ and result in information loss as well.

Below a resolution for local confounding is proposed. Consider an log-rank estimating equation $S(u; \beta_0)$ constructed from a scoring function G. A reliable initial estimate of β and then an improved version are derived via a simple methodology with capability to alleviate the effect of local confounding. The heuristic idea is that we partition the entire concerned time interval into several sub-intervals. Practically we might be able to contain all possible effects of local confounding within just a few subintervals, while the function $\rho_G(u; \beta_0, \lambda)$ is monotonic in the rest of subintervals. These extra computations would put a premium on the efficiency of estimating β as illustrated below.

With $S(u; \beta_0)$ being a martingale process, we define a process as

$$M(u; \beta) = n^{-1/2} \sum_1^n \int_0^u \{G_i(v, \beta) - G^{av}(v, \beta)\} \,\mathrm{d}N_i(v; \beta).$$

Further, for a chosen set of time points $\{u_1, u_2, \ldots, u_k\}$ with $u_0 = 0$ and $u_{k+1} = \infty$, denote the increments by $\Delta_i K = K(u_i) - K(u_{i-1})$ for $i = 1, \ldots, k+1$, where K can be any function, matrix of functions, or martingale process. Again by applying Rebolledo's central limit theorem, see Andersen and Gill (1982), and the locally asymptotic linearity (8), we have a system of $k+1$ regression equations:

$$\begin{aligned} \Delta_i M(u; \beta) &= M(u_i; \beta) - M(u_{i-1}; \beta) \\ &\approx \Delta_i M(u; \beta_0) + N^{1/2}(\beta - \beta_0)\Delta_i \rho_G(u; \beta_0, \lambda), \\ \Delta_i M(u; \beta_0) &\overset{d}{\approx} N(0, \Delta_i H_G(u)). \end{aligned}$$

Let an initial estimate be constructed as $\hat{\beta}_0 = \arg_\beta \inf \sum_1^{k+1} \{\Delta_i M(u, \beta)\}^2$. Then, with $\hat{\beta}_0$, the functions $H_G(u)$ and $\Lambda(u)$ are estimated. In turn, the one-step estimate of β is derived as

$$\hat{\beta}_{G,k} = \arg_\beta \inf \sum_1^{k+1} \{\Delta_i M(u, \beta)\}^2 / \Delta_i \hat{H}_G(u),$$

so is the following asymptotic result:

$$\begin{aligned} n^{-1/2}(\hat{\beta}_{G,k} - \beta_0) &\overset{d}{\approx} N(0, \sigma^2_{G,k}), \\ \sigma^{-2}_{G,k} &= \sum_{i=1}^{k+1} (\Delta_i \rho_G(u; \beta_0, \lambda))^2 / \Delta_i H_G(u). \end{aligned}$$

If a truncated time point $u^*(\leq \infty)$ is previously taken, that is, $u_i < u^*, \ i = 1, \ldots, k$, and k is large, then we have

$$\begin{aligned} \sigma^{-2}_{G,k} &\approx \int_0^{u^*} (\mathrm{d}\rho_G(v; \beta_0, \lambda))/\mathrm{d}v)^2 / \{\mathrm{d}H_G(v)/\mathrm{d}v\} \,\mathrm{d}v, \\ &= \int_0^{u^*} [\{A^G_R(v, \beta_0)\lambda'(v) + A^G_Z(v, \beta_0)\lambda(v)\}^2 / \{A^G_G(v, \beta_0)\lambda(v)\}] \,\mathrm{d}v, \end{aligned}$$

$$\geq \left[\int_0^{u^*} \{A_R^G(v,\beta_0)\lambda'(v) + A_Z^G(v,\beta_0)\lambda(v)\}\,\mathrm{d}v \right]^2 \Big/ \left[\int_0^{u^*} \{A_G^G(v,\beta_0)\lambda(v)\}\,\mathrm{d}v \right],$$
$$= \{\rho_G(u^*)\}^2 / H_G(u^*) = \sigma_G^{-2}(u^*).$$

where the above inequality follows the Cauchy–Schwarz inequality, and the last equality denotes the inverse of the asymptotic variance of the estimator $\hat{\beta}_{u^*}^{\mathrm{TLN}}$, which solves the log-rank estimating equation $0 = S(u^*;\beta) = n^{1/2} M(u^*;\beta)$.

The above derivations clearly illustrate a methodology to reduce the effect of local confounding and, at the same time, to secure the information contained in the data to a certain extent. This methodology is indeed the backbone of the OEE proposed and discussed in the next section.

4 The OEE approach

In this section, the OEE approach is applied to AFT model (3) simultaneously based on the following two processes constructed through scoring functions $R_i(u,\beta)$ and $Z_i(u,\beta)$, respectively:

$$M_1(u;\beta) = n^{-1/2} \sum_1^n \int_0^u \{R_i(v,\beta) - R^{av}(v,\beta)\}\,\mathrm{d}N_i(v;\beta),$$

and

$$M_2(u;\beta) = n^{-1/2} \sum_1^n \{Z_i(v,\beta) - Z^{av}(v,\beta)\}\,\mathrm{d}N_i(v;\beta).$$

Explicitly this inference approach for β has neither involvement of the unknown derivative $\lambda'(u)$, nor its estimate. As for $\lambda(u)$, it appears only in the predictable covariance processes of the above two martingales. However, we would show that the OEE approach indeed achieves the semiparametric Fisher information bound calculated in Sect. 2. It is noted that the above system of martingale processes can be correspondingly extended to accommodate functions $\varphi_j(.)$ as components of multivariate $Z(t)$ in the finite-sum version of the time-transformed AFT model (5).

For the time being, let our discussion of the OEE approach be restricted on a finite and fixed interval $[0, u^*]$. Such a finite truncation would theoretically free us from dealing with complicated tail conditions on the baseline hazard function $\lambda(u)$. Practically, from the perspective of local confounding, it would be unrealistic for any inference procedure to extract information beyond a large truncating time point if $\lambda(u)$ has a wiggling tail as in most real world problems. It is noted that the regularity conditions C(1)–C(4), given in "Appendix", are needed for the argument with u^* being depending on n and going to ∞ in whatever slow rate. So that they are slightly more than what is needed with a finite u^* to ensure the validity of Rebolledo's central limit theorem for a local square integrable martingale.

By applying Rebolledo's central limit theorem, martingale processes $M_1(\cdot;\beta)$ and $M_2(\cdot;\beta)$, evaluated at β_0, converge jointly in $D[0, u^*]$ to the continuous Gaussian martingales $W_1(\cdot)$ and $W_2(\cdot)$, respectively, with $W_1(0) = W_2(0) = 0$ and a 2×2 matrix of covariance functions

$$cov[\{W_1(u_1), W_2(u_2)\}, \{W_1(v_1), W_2(v_2)\}^T] = H(u_1 \wedge v_1, u_2 \wedge v_2; \beta_0, \lambda)$$

for all positive u_i, v_i $(\leq u^*)$, where

$$H(u, v; \beta, \lambda) = \begin{bmatrix} H_{11}(u; \beta, \lambda) & H_{12}(u \wedge v; \beta, \lambda) \\ H_{21}(u \wedge v; \beta, \lambda) & H_{22}(v; \beta, \lambda) \end{bmatrix},$$

with

$$H_{11}(u; \beta, \lambda) = \int_0^u A_R^R(v, \beta)\lambda(v)\,\mathrm{d}v,$$
$$H_{12}(u; \beta, \lambda) = H_{21}(u; \beta, \lambda) = \int_0^u A_Z^R(v, \beta)\lambda(v)\,\mathrm{d}v,$$

and

$$H_{22}(u; \beta, \lambda) = \int_0^u A_Z^Z(v, \beta)\lambda(v)\,\mathrm{d}v.$$

Again consider a k-vector of ordered cutoff points $\{u_1, u_2, \ldots, u_k\}^T$, with $u_0 = 0$ and $u_{k+1} = u^*$. Then the above weak convergence of $\{M_1(u; \beta_0), M_2(u; \beta_0)\}$ render the following approximations as:

$$\{\Delta_i M_1(t; \beta_0), \Delta_i M_2(t; \beta_0)\} \overset{d}{\approx} N\{0, \Delta_i H(t, t; \beta_0, \lambda)\}. \tag{9}$$

Similarly, for any u in $[0, u^*]$, the locally asymptotic linearities (8) pertaining to $M_1(u; \beta)$ and $M_2(u; \beta)$ can also be derived. Hence we have the following system of regression equations with approximated Gaussian errors: for $j=1, 2$, $i=1, \ldots, k+1$,

$$\Delta_i M_j(u; \beta) = \Delta_i M_j(u; \beta_0) + n^{1/2}(\beta - \beta_0)\Delta_i \rho_j(u; \beta_0, \lambda), \tag{10}$$

where

$$\rho_1(u; \beta, \lambda) = \int_0^u \{A_R^R(v, \beta)\lambda'(v) + A_Z^R(v, \beta)\lambda(v)\}\,\mathrm{d}v,$$

and

$$\rho_2(u; \beta, \lambda) = \int_0^u \{A_Z^R(v, \beta)\lambda'(v) + A_Z^Z(v, \beta)\lambda(v)\}\,\mathrm{d}v.$$

Then, based on the above approximated regression setting, an initial estimate $\hat{\beta}_0$ can either be calculated as $\arg_\beta \inf \sum_1^{k+1}\{\Delta_j M_1(t; \beta)\}^2 + \{\Delta_j M_2(t; \beta)\}^2$, or simply as $\arg_\beta \inf \sum_1^{k+1}\{\Delta_j M_1(u; \beta)\}^2$. Clearly both $\hat{\beta}_0$ are root-n-consistent.

Further, with $\hat{\beta}_0$, the function $H_{12}(u;\beta,\lambda)$ is consistently estimated by $\hat{H}_{12}(u) = \int_0^u \hat{A}_Z^R(v,\hat{\beta}_0)\,\mathrm{d}\hat{\Lambda}(v)$, where

$$\hat{A}_Z^R(u,\beta) = \frac{1}{n}\sum_{i=1}^{n}\{R_i(u,\beta) - R^{av}(u,\beta)\}\{Z_i(u,\beta) - Z^{av}(u,\beta)\}Y_i(u,\beta),$$

and

$$\hat{\Lambda}(u) = \sum_{i=1}^{n}\int_0^u \frac{\mathrm{d}N_i(v,\hat{\beta}_0)}{\sum_{j=1}^{n} Y_j(v,\hat{\beta}_0)}.$$

Here $\hat{\Lambda}(u)$ is Nelson–Aalen estimate of the cumulative hazard function $\Lambda(u) = \int_0^u \lambda(v)\mathrm{d}u$ based on calculated baseline failure times $\{U_i(\hat{\beta}_0)\}$. Likewise $H_{11}(u;\beta,\lambda)$ and $H_{22}(u;\beta,\lambda)$ are estimated.

Finally, with $\hat{\beta}_0$ and the 2×2 matrix, $\hat{H}$, our OEE estimate of β, denoted by $\hat{\beta}_k^{\mathrm{OEE}}$, is calculated as $\hat{\beta}_k^{\mathrm{OEE}} = \arg_\beta \inf \chi^2_{k+1}(\beta)$, where

$$\chi^2_{k+1}(\beta) = \sum_{j=1}^{k+1}\{\Delta_j M_1(u;\beta), \Delta_j M_2(u;\beta)\}\{\Delta_j \hat{H}(u,u)\}^{-1} \times\{\Delta_j M_1(u;\beta), \Delta_j M_2(u;\beta)\}^T.$$

By standard asymptotic arguments, we have the following result:

$$\sqrt{n}(\hat{\beta}_k^{\mathrm{OEE}} - \beta_0) \overset{d}{\approx} N(0,\sigma_k^2),$$
$$\sigma_k^{-2} = \sum_{j=1}^{k+1}\{\Delta_j\rho_1(u), \Delta_j\rho_2(u)\}\Delta_j H(u,u)^{-1}\{\Delta_j\rho_1(u), \Delta_j\rho_2(u)\}^T.$$

Theoretically, if k is chosen to grow with n in a rate, even as slow as $\log n$, to make the corresponding mesh $\max_{1\le i\le k(n)}(u_i - u_{i-1})$ shrink to zero as n goes to ∞, then the semiparametric Fisher information bound is obtained as a limit of $\sigma^{-2}_{k(n)}$. This is,

$$\sigma^{-2}_{k(n)} \approx \lim_{n\to\infty} 1/n \sum_{1}^{n}\int_0^{u^*}\{G_{OP,i}(v,\beta_0) - G^{av}_{OP}(v,\beta_0)\}^2 Y_i(v,\beta_0)\,\mathrm{d}v,$$
$$= \int_0^{u^*}[A_R^R\{\lambda'(v)\}^2 + 2A_Z^R\lambda'(v)\lambda(v) + A_Z^Z\{\lambda(v)\}^2]/\lambda(v)\,\mathrm{d}v,$$

where the efficient scoring function recalled from Sect. 2 is

$$G_{OP,i}(u,\beta) = \mathrm{d}\log\{\lambda_{i,\beta}\}/\mathrm{d}\beta, = R_i(u,\beta)\lambda'(u)/\lambda(u) + Z_i(u,\beta).$$

In this sense our OEE estimate $\hat{\beta}^{\mathrm{OEE}}_{k(n)}$ of β achieves the information bound given any finite truncation at u^*.

In fact, with slightly modification on the above asymptotic argument, we could allow the finite truncation $u^*(n)$ to be depending sample size n and go to ∞ in rather slow rate. By doing so, the OEE approach can asymptotically achieve the semiparametric Fisher information bound calculated in Theorem 1. This result is summarized in the next theorem.

Theorem 3 *Under the conditions of Theorem 2, let* α *and* ζ *be positive constants and take* $k_0 = \alpha + \zeta$ *and* $k(n) = \{\log n\}^{k_0}$. *Then mesh* $\max_{1\le i\le k(n)}(u_i - u_{i-1}) = O(\{\log n\}^{-\zeta})$ *with* $u_{k(n)} = O(\{logn\}^{\alpha})$. *Now the estimate,* $\hat{\beta}^{\text{OEE}}_{k(n)}$, *is consistent and asymptotically normal; that is*

$$\sqrt{n}(\hat{\beta}^{\text{OEE}}_{k(n)} - \beta_0) \overset{d}{\approx} N(0, \sigma^2_\infty),$$

where

$$\sigma_\infty^{-2} = \lim_{n\to\infty} 1/n \sum_{i=1}^{n} \int_0^\infty \left\{ \left[Z_i(v,\beta) - \frac{\mathcal{E}Y^Z(v,\beta)}{\mathcal{E}Y(v,\beta)} \right] + \frac{\lambda'(v)}{\lambda(v)} \left[R_i(v,\beta) - \frac{\mathcal{E}Y^R(v,\beta)}{\mathcal{E}Y(v,\beta)} \right] \right\}^2 L_i(v,\beta)\, \mathrm{d}F(v).$$

Finally, as a by-product, the statistic $\chi^2_{k+1}(\hat{\beta}^{\text{OEE}}_k)$ is in fact a Chi-squared (χ^2) statistic for testing the AFT model assumption in (3). Specifically, under the null hypothesis, this statistic is asymptotically χ^2 distributed with degree of freedom $2K + 1$.

Remark 1 Theorem 3 also works the finite-sum version of the time-transformed AFT model (5) without extra arguments needed. However, for infinite-sum version, which is equivalent to Eq. (1), we need extra approximation arguments. Based on some smoothness assumptions on Λ_0, we grow the finite-sum at a certain rate that is regulated by the nonparametric converging rate of approximating Λ_0. For finite sample computations, ideally we should choose a reasonable basis $\{\varphi_j(.), j = 1, \ldots n\}$ that can facilitate an efficient approximation on Λ_0.

Remark 2 The OEE approach can be easily adapted for the time-dependent structure of $\beta(t)$ by imposing piece-wise constant on every subinterval involved in the development of the approach, and by solving individually equations in (10).

5 A simulated biological example

The goal of the numerical study reported in this section is twofold. First, to illustrate that the basic lifetime regression model, the identical time scale AFT model (3), can be used for modeling reproductive cost on female animal's survival time. Second, to numerically exhibit the effects of the phenomenon of local confounding.

Let $\tilde{m} = \{m(i), i = 1, 2, \ldots M\}$ denote the age-specific fecundity schedule experienced by a female individual who is capable of reproducing more than once. Here $m(i)$ is the number of off-springs reproduced at the age i. The M is taken to be the default of the highest attainable age, while the reproductive span is limited within [1, 10]. That is, $m(i) = 0$, for all age i greater than 10. Also we take $m(0) = 0$ to indicate that the age at sexual maturity not at 0, but at age 1.

For an individual who experiences the reproductive schedule specified by $\tilde{m}$, her observable survival time is T and the corresponding (age-specific) hazard function, denoted by $\lambda(t; \beta, \tilde{m})$, is calculated based on the identical time scale AFT model (3). Let the zero-vector $\tilde{m}_0 = \{0, \ldots, 0\}$ denote the case of having no reproduction throughout a female's whole life span. With $\tilde{m}_0$, the individual is thought of investing all her vital resource and energy reserve for somatic (bodily) repair and maintenance, not on reproduction. Hence, her random survival time corresponds to the baseline survival time U with the baseline hazard rate $\lambda(t)$ to be defined on [0, M].

Instead of using $\tilde{m}$ directly as the covariate history, we derive its impact function via the shock model, see Griffiths (1988), for the covariate $Z(t)$ as follows. Let

$$Z(s) = \sum_{i=1}^{M} m_i K(s-i); \quad K(s) = \beta_1 I_{[0,1]}(v) + \beta_\infty I_{[0,\infty]}(v),$$

where $I_{[a,b]}$ is the indicator function on interval $[a, b]$. Here the kernel function $K(s)$ is the impact waveform caused by a single reproduction. Further $K(s)$ is a combination of two types of reproductive cost: one is the immediate effect with size β_1; and the other is the ever-lasting damaging effect with size β_∞. Evaluating the latter effect is in fact a central concern in theories of life history and aging. Therefore, the time-dependent covariate history $\bar{Z}(T) = \{Z(t)|0 < t < T\}$ considered here is the linear combination of waves of impact functions caused by the $\tilde{m}$.

Then the hazard function of T is calculated via Eq. (2) as

$$\Lambda(t; \theta, \tilde{m}) = \Lambda\left[\int_0^t e^{\beta Z(s)}(s)\, \mathrm{d}s\right].$$

Therefore, the novelty of this biological example is that the lifetime of a female is shorten from her baseline survival time U by a rate of $e^{\beta Z(t)}$, which is characterized by the composed impacts of reproductive costs underlying her fecundity schedule $\tilde{m}$.

It is interesting to note that, if the above hazard rate plays the role of age-specific mortality on a particular population, then immediately, we are in a significantly different position from that of classic Life Table analysis, which basically relies on the key assumption of independence between age-specific mortality and fecundity. Hence, the novelty of this modeling also bears several evolutionary implications through the sensitivity of Malthusian parameter, so-called the force of natural selection, in theories of life-history and aging and in demography, as reported Hsieh (2003).

To the end of this section, we report results of estimating β from three simulated scenarios under the above setting. For the first two scenarios, the baseline hazard function is taken as:

$$\begin{aligned}\lambda(u) &= 0, \quad 0 < u \le 1,\\ &= b(u-1)^a, \quad 1 < u \le 6,\\ &= \frac{c}{1+(u-6)}, \quad 6 < u,\end{aligned}$$

where constants a, b are chosen such that the survival probability $Pr(U > 6) = 0.5$, and c is chosen such that λ_U is continuous at age 6. Therefore, $\lambda(t)$ is strictly increasing on [1, 6] and is strictly decreasing on $(6, \infty)$.

The reproduction schedule $\tilde{m}$ is generated as follows. For each i, $m(i)$ takes values 0 or 1. The starting $m(1)$ is always generated from a fair coin, then $m(i+1)$ is generated following a Markov process of Bernoulli random variable with transition probabilities specified by $Pr(m(i+1) = 1|m(i) = 1) = p_1$ and $Pr(m(i+1) = 0|m(i) = 0) = p_0$. No censoring is considered throughout this simulation study.

For illustrating purpose two models are separately considered: one with $\beta_1 = 1$ and $\beta_\infty = 0$; and the other with $\beta_\infty = 1$ and $\beta_1 = 0$. Such considerations would facilitate our evaluations of the effects of local confounding on estimating β_∞ and β_1 with identical $\tilde{m}$.

We divide the time scale of u into three pieces:[0, 6], [6, 10] and $(10, \infty)$, that is, $k = 1$ and $u_1 = 6$ and $u_2 = 10 = u^*$ as in Sect. 4. Then we compare the performance in terms of mean squared error (MSE) of the following four estimates of β:

$$\hat{\beta}_0 = \arg_\beta \inf \sum_{j=1}^{k+1} \{\Delta_j M_1(t;\beta)\}^2,$$

$$\hat{\beta}_k^{\text{OEE}} = \arg_\beta \inf \sum_{1}^{k+1} \{\Delta_j M_1(t;\beta), \Delta_j M_2(t;\beta)\}\{\Delta_j \hat{H}(t,t)\}^{-1} \times \{\Delta_j M_1(t;\beta), \Delta_j M_2(t;\beta)\}^T,$$

$$\hat{\beta}^{\text{LN}} = \arg_\beta \inf\{M_1(\infty;\beta)\}^2, \quad \hat{\beta}^{\text{TLN}} = \arg_\beta \inf\{M_1(10;\beta)\}^2$$

Here $\hat{\beta}_0$ is the initial estimate, and would be compared with the log-rank estimate $\hat{\beta}^{\text{LN}}$ solving the untruncated estimating equation $S(\infty, \beta) = 0$ in Sect. 3, as used in Robins and Tsiatis (1992), and the truncated one $\hat{\beta}^{\text{TLN}}$ in regarding to local confounding.

All results reported here are based on 500 replications. Each replication is consisting of a sample of size 100. As the first scenario, both p_1 and p_0 for reproductive schedule are taken to be 0.5. This gives rise to a rather varying on-and-off pattern of short-term impact effect. All four estimators are equally unbiased, but rather different in MSEs. For both short and long-term effects, in Table 1, $\hat{\beta}_k^{\text{OEE}}$ consistently out per-

Table 1 Comparing the four estimates: $\hat{\beta}_0$, $\hat{\beta}_k^{\text{OEE}}$, $\hat{\beta}^{\text{LN}}$ and $\hat{\beta}^{\text{TLN}}$, in terms of MSEs under varying reproductive schedules

Estimates	Short-term effect	(β_1)	Long-term effect	(β_∞)
	Sample mean	MSE	Sample mean	MSE
$\hat{\beta}_0$	0.987	18.498×10^{-4}	1.005	2.964×10^{-4}
$\hat{\beta}_k^{\text{OEE}}$	1.001	6.870×10^{-4}	1.004	1.186×10^{-4}
$\hat{\beta}^{\text{LN}}$	1.008	3.343×10^{-2}	1.002	2.226×10^{-2}
$\hat{\beta}^{\text{TLN}}$	1.006	3.313×10^{-2}	0.997	2.015×10^{-2}

Table 2 Comparing the four estimates: $\hat{\beta}_0$, $\hat{\beta}_k^{\text{OEE}}$, $\hat{\beta}^{\text{LN}}$ and $\hat{\beta}^{\text{TLN}}$, in terms of MSEs under non-varying reproductive schedules

Estimates	Short-term effect	(β_1)
	Sample mean	MSE
$\hat{\beta}_0$	1.007	3.140×10^{-4}
$\hat{\beta}_k^{\text{OEE}}$	1.008	2.302×10^{-4}
$\hat{\beta}^{\text{LN}}$	1.010	1.997×10^{-2}
$\hat{\beta}^{\text{TLN}}$	1.008	1.851×10^{-2}

Table 3 Comparing the four estimates: $\hat{\beta}_0$, $\hat{\beta}_k^{\text{OEE}}$, $\hat{\beta}^{\text{LN}}$ and $\hat{\beta}^{\text{TLN}}$, in terms of MSEs under varying reproductive schedules with a smooth baseline hazard function

Estimates	Long-term effect	(β_∞)
	Sample mean	MSE
$\hat{\beta}_0$	0.995	1.622×10^{-3}
$\hat{\beta}_k^{\text{OEE}}$	1.004	1.633×10^{-3}
$\hat{\beta}^{\text{LN}}$	1.001	3.509×10^{-2}
$\hat{\beta}^{\text{TLN}}$	1.015	4.428×10^{-2}

forms all other three estimates of β. Even $\hat{\beta}_0$ works much better than $\hat{\beta}^{\text{LN}}$ and $\hat{\beta}^{\text{TLN}}$. It is noted that, among these four competing estimates, only the un-truncated log-rank estimate $\hat{\beta}^{\text{LN}}$ makes use of the 10% of data (falling into $(10, \infty)$). But still it performs no better than the truncated log-rank estimate $\hat{\beta}^{\text{TLN}}$. This clearly and exactly exhibits the effect of local confounding.

From reported MSEs in Table 1, we see that the amount of information of parameter $\beta_\infty(=1)$ is five times more than that of $\beta_1(=1)$ given the same reproduction schedules. This numerical result suggests that the long-term effect is indeed estimable in most real data sets even when its value is much smaller than the short-term effect.

In the second scenario, we take $p_1 = p_0 = 0.9$ to give rise to reproductive schedules not so varying as that considered in Table 1. From Table 2, the MSEs of the four estimates show that more information of the short-term effect is contained in the data with stable and frequent reproductions, while keeping the same ranking pattern as seen in Table 1.

Finally, the baseline hazard function is changed into a very smooth and slowly increasing function as:

$$\lambda(u) = 0, \quad 0 < u \leq 1,$$
$$= b(u-1)^a, \quad 1 < u.$$

Since there is no cusp at 6 in $\lambda(u)$. Therefore, no effect of local confounding would be expected. From Table 3, $\hat{\beta}_0$ and $\hat{\beta}_k^{\text{OEE}}$ are still significantly out-perform the two log-rank estimates. This result further confirms the practical advantage of the OEE approach on extracting more information given that it is free of the necessity of choosing among weighting functions. Again it is noted that the untruncated log-rank estimate $\hat{\beta}^{\text{LN}}$ makes use of 20–30% of data falling in $[10, \infty)$ in this scenario. This is the reason that MSEs are higher in Table 3 than that in Tables 1 and 2. The reason behind the equal performance of $\hat{\beta}_0$ and $\hat{\beta}_k^{\text{OEE}}$ is that the scoring function $R(u, \beta)$ only contributes a limited amount of information about the regression parameter β_∞ due to the derivative $\lambda'_U(u)$ being drastically decreasing to zero.

As a practical suggestion for real data analysis with OEE approach, the numerical results reported in this section indicates that is essential to plot the estimated cumulative baseline hazard function $\hat{\Lambda}(t)$, and then to look for reflection points where the concave upward pattern turns into downward, or vise versa. And a partition on the time scale is then accordingly made. As demonstrated in Table 3, even when $\hat{\Lambda}(t)$ looks consistently concave either upward or downward, it is still worthy of making a partition for applying the OEE approach.

6 Discussion

In part I of this paper, results of semiparametric efficient estimation via OEE approach achieving the Fisher information bound are derived and proved under the identical time-scale version of AFT model. The statistical inference approach based on a system of martingale equations is simple and practical for real world applications. And the model is illustrated with biological plausibility, and might be taken as the most fundamental lifetime regression model with time-dependent covariate history. Therefore the model and its comprehensive accounts of semiparametric statistical inferences are important in their own right.

Certainly the real world biological, or social research mostly takes a much more complex form than the one discussed in Sect. 5. As demonstrated in Sect. 1, the time-transformed AFT model via threshold equation (1) exactly exhibits the wide spectrum of modeling capacity being necessary for an ideal lifetime regression model in survival analysis. Especially, a glimpse of its significance is seen as that Cox's hazard regression and transformation models and most of their variants are included as its special cases.

Although the semiparametric efficient estimation pertaining to this time-transformed AFT model involved with greater complexity than that have derived in this paper, the light of feasibility of OEE approach is shed through its applicability on the finite-sum version of model (5). The full and rigorous account of development can be achieved with more detailed approximation arguments under some smoothness conditions on $\Lambda_0(.)$.

From computational and information conserving perspectives, the phenomenon of local confounding is not only a reasonable, but necessary concern in making statistical inferences via lifetime regression analysis. It is worthy emphasizing this point by looking the interesting example reported in Vaupel and Carey (1993), where a mixture of 12 Gompertz mortality curves are needed to model the wiggle mortality curve calculated from the survival times of 1.2 million medflies reported in Carey et al. (1992).

We end this discussion by pointing out how the identical time-scale AFT model accommodates various aspects of heterogeneity. To evaluate measurable sub-populational heterogeneity, a heteroscedastic extension of model (3) via a power transformation on U can be formulated as, see Hsieh (1996b)

$$U^{1/\alpha(X)} = \int_0^T e^{\beta' Z(s)}\,\mathrm{d}s,$$

where $\alpha(0) = 1$ and X is a time-independent environmental condition with each of its values specifying a subpopulation. The backbone of the above model is the now classic heteroscedastic semiparametric linear regression model

$$\log T = -\beta Z + 1/\alpha(X) \log U,$$

when $Z(t)$ is time-independent.

As for unmeasurable individual heterogeneity, as demonstrated in Sect. 1, the AFT models in (3) could adapt frailty adjustment in someway better than the proportional hazard model. Keiding et al. (1998) gave an interesting discussion on this perspective under a setting with time-independent covariates.

Appendix

First, the four regularity conditions needed in Theorems 2 and 3 are listed, and the proofs.

[Regularity conditions]

(C1) Each covariate process, Z_i, is uniformly bounded for all $i = 1, \ldots, n$.
From (C1), we obtain the unique Jordan decomposition $Z_i(t) = Z_i(0)+Z_i^+(t)-Z_i^-(t)$, where both $Z_i^+(t)$ and $Z_i^-(t)$ are increasing functions with $Z_i(0) = 0$; while $Z_i^+(u,\beta) = Z_i^+(\psi^{-1}(u,\beta))$ and $Z_i^-(u,\beta) = Z_i^-(\psi^{-1}(u,\beta))$.

(C2) There are $\eta_0 > 0$ and $\kappa_0 > 0$ such that

$$\sup_{|u-v|+|b-\beta|<n^{-\kappa_0}} n^{-1}\sum_1^n |Z_i^+(u,\beta) - Z_i^+(v,b)| = O(n^{-1/2-\eta_0}).$$

And for $0 < d_n \to 0$, there exists $\varepsilon > 0$ such that

$$\sup_{|u-v|+|b-\beta|<d_n} n^{-1}\sum_1^n |Z_i^+(u,\beta) - Z_i^+(v,b)| = o(\max\{d_n^{\varepsilon_0}, n^{-\varepsilon_0}\}).$$

The same conditions also hold for $Z_i^-(u,\beta)$.

(C3) The baseline density, f, and its derivative, f', are bounded, and

$$\int_0^\infty \left[\frac{f'(u)}{f(u)}\right]^2 f(u)\,\mathrm{d}u < \infty, \quad \int_0^\infty u^{\theta_0} f(u)\,\mathrm{d}u < \infty, \quad \text{for some } \theta_0 > 0.$$

(C4) The common density function, g, of C_i is uniformly bounded and goes to zero in the tail faster than $\frac{1}{t^{1+\alpha_0}}$.

It should be noted here that $R_i(u,\beta)$ is continuous and differentiable with respect to u as well as to β with derivatives

$$\int_0^{\psi^{-1}(u,\bar{Z}_i,\beta)} Z_i(v)[Z_i(u,\beta) - Z_i(v)] \exp\{\beta Z_i(v)\}\, \mathrm{d}v, \quad \text{and} \quad Z_i(u,\beta),$$

respectively. From (C1), $Z_i(t)$ is uniformly bounded, such that the growth rate of $R_i(u,\beta)$ is of the order $O(u)$. These four conditions are close to those used in Lin and Ying (1995).

Proof of Theorem 1 With the definition of $Q_i(\beta)$, we have $V_i(\beta) = U_i(\beta) \wedge Q_i(\beta)$, and $U_i(\beta_0)$ and $Q_i(\beta_0)$ are independent under the censoring assumption given $\bar{Z}_i(t)$ for any t. According to argument used in Ritov and Wellner (1988), we begin by calculating scores for β and f, the density corresponding to the unknown baseline hazard rate, λ, in the model (3). By straightforward calculation, the density of $U_i(\beta)$ derived from $\lambda_{i,\beta}(u)$ is

$$f_{i,\beta}(u) = [1 - F_{i,\beta}(u)]^{-1} \lambda \left[\int_0^{\psi^{-1}(u,\bar{Z}_i,\beta)} \exp\{\beta_0 Z_i(v)\}\, \mathrm{d}v \right]$$
$$\times \exp\{(\beta_0 - \beta) Z_i(\psi^{-1}(u, \bar{Z}_i, \beta))\},$$

and the score function for β is $G^{uc}_{OP,i}(u,\beta) = R_i(u,\beta)\frac{\lambda'(u)}{\lambda(u)} + Z_i(u,\beta)$.

To calculate the score for f, let $\{f_\eta : \eta \in R\}$ be a regular parametric family, and set $a = \frac{\partial}{\partial \eta} \log f_\eta$. Then the score for f is $a(U_i(\beta_0))$.

Under model (3) with right-censored data, the two scores for β and f can be further calculated by the following conditional expectation as

$$s_\beta(u) = \mathcal{E}[G^{uc}_{OP,i}(u,\beta) | \bar{Z}_i(\psi^{-1}(u, \bar{Z}_i, \beta)), \mathcal{F}_u],$$

and

$$s_{a,f}(u) = \mathcal{E}[a(U_i(\beta_0)) | \bar{Z}_i(\psi^{-1}(u, \bar{Z}_i, \beta)), \mathcal{F}_u]$$

evaluated at $u = Q_i(\beta)$ where $\mathcal{F}_u \equiv \sigma\{U_i(\beta_0) \wedge u, I\{U_i(\beta_0) \le u\}\}$.

Further, by applying Proposition 3.1 in Ritov and Wellner (1988) and simple calculations, we have

$$s_\beta(u) = \int_0^u \mathcal{R}[G^{uc}_{OP,i}](v,\beta)\, \mathrm{d}M_i(v), = \int_0^\infty 1_{\{Q_i \ge u\}} [\mathcal{R}[G^{uc}_{OP,i}](v,\beta)\, \mathrm{d}M(v)$$
$$= \int_0^\infty \left\{ R_i(u,\beta_0) \frac{\lambda'(u)}{\lambda(u)} + Z_i(u,\beta_0) \right\} \mathrm{d}M_{i,uc}(v) = \int_0^\infty G_{OP,i}(v,\beta)\, \mathrm{d}M_{i,uc}(v),$$
$$s_{a,f}(u) = \int_0^u \mathcal{R}[a](v)\, \mathrm{d}M_i(v) = \int_0^\infty \mathcal{R}[a](v)\, \mathrm{d}M_{i,uc}(v),$$

where $\mathcal{R}$ is the R operator defined in Ritov and Wellner (1988).

To find the efficient score for β, we need to solve for an a^* with $\int a^* \mathrm{d}F = 0$ in the following equation, for all functions $a \in L^0_2(F)$,

$$
\begin{aligned}
0 &= \mathcal{E}\{[s_\beta - s_{a^*,f}]s_{a,f}\} = \mathcal{E}\mathcal{E}\{[s_{u;\beta} - s_{u;a^*,f}]s_{u;a,f}|\bar{Z}_i(\infty), Q_i\} \\
&= \mathcal{E}\int_0^\infty I_{\{Q_i \ge u\}}[G_{OP,i}(v,\beta) - \mathcal{R}[a^*](v)]\mathcal{R}[a](v)\,\mathrm{d}F(v) \\
&= \int_0^\infty \{\mathcal{E}[I_{\{Q_i\ge u\}}G_{OP,i}(v,\beta)] - \mathcal{E}[I_{Q_i\ge u}]\mathcal{R}[a^*](v)\}\mathcal{R}[a](v)\,\mathrm{d}F(v) \\
&= \int_0^\infty \mathcal{E}[1_{\{Q_i\ge u\}}]\{\mathcal{E}[G_{OP,i}(v,\beta)|Q_i \ge u] - \mathcal{R}[a^*](v)\}\mathcal{R}[a](v)\,\mathrm{d}F(v).
\end{aligned}
$$

Therefore it is easy to make the right choice of a^* such that $\mathcal{R}[a^*](u) = \mathcal{E}[G_{OP,i}(u,\beta)$ $|Q_i(\beta) \ge u]$, and the efficient score is

$$
\begin{aligned}
s_\beta(u) - s_{a^*,f}(u) &= \int_0^u I_{Q_i\ge u}[G_{OP,i}(v,\beta) - \mathcal{E}[G_{OP,i}(v,\beta)|Q_i \ge u]]\,\mathrm{d}M_i(v), \\
&= \int_0^\infty [G_{OP,i}(v,\beta) - \mathcal{E}[G_{OP,i}(v,\beta)|Q_i \ge u]]\,\mathrm{d}M_{i,uc}(v), \\
&= \int_0^\infty [G_{OP,i}(v,\beta) - \mathcal{E}[G_{OP,i}(v,\beta)|V_i(\beta) \ge u]]\,\mathrm{d}M_{i,uc}(v),
\end{aligned}
$$

where the last equation follows from the independence of (Z_i, Q_i) and U_i. □

Proof of Theorem 2 It is noted that for any finite u, $M_i(u;\beta) = 0$, $i = 1, 2$, is a truncated (at u) log-rank estimating equation. And, in this theorem, we are only concerned with the two asymptotic properties on the sequence of intervals $[0, u_{k(n)}]$. Furthermore, both asymptotic properties of $M_2(u;\beta)$ have been well treated in Theorem 1 of Lin and Ying (1995). So that we only need to prove these two properties for $M_1(u;\beta)$.

The argument below condition C(4) assures that $R_i(u,\beta)$ is continuous and differentiable with bounded derivatives, with respect to β, of the order $O(\{\log n\}^{k_0})$ on $[0, u_{k(n)}]$. Therefore $\{R_i(u,\beta) : 0 < u \le t_{k(n)}, |\beta| < B\}$ as a family of weights for the martingale process $M_1(t;\beta)$ has a sequence of packing numbers of a polynomial order of n and $\log n$. Hence Bennett's inequality (Shorack and Wellner 1986) can be applied in the same way as in Lemma 1 of Lin and Ying (1995) to establish the following approximations.

Let $J(u,\beta)$ be any of the two processes $\{N^R(u,\beta), Y^R(u,\beta)\}$ satisfying the following:

(i) For any $B > 0$ and small positive number $\varepsilon > 0$

$$
\sup_{|\beta-\beta_0|\le B} |J(u;\beta) - \mathcal{E}J(u;\beta)| = o(n^{1/2+\varepsilon}) \text{ almost surely;}
$$

(ii) for every $\gamma > 0$, there exists $\eta > 0$ such that

$$
\begin{aligned}
&\sup_{|\beta-\beta_0|\le n^{-\gamma}} |J(u;\beta) - \mathcal{E}J(u;\beta) - J(u;\beta_0) + \mathcal{E}J(u;\beta_0)| \\
&\quad = o(n^{1/2+\varepsilon})\text{almost surely.}
\end{aligned}
$$

With these approximations together with integration by part, we then have

$$
\begin{aligned}
&\Delta_j M_1(u;\beta) - \Delta_j \bar{M}_1(u;\beta) \\
&= \sum_1^n \int_{u_{j-1}}^{u_j} R_i(v,\beta)\mathrm{d}[N_i(u;\beta) - \mathcal{E}N_i(u;\beta)] \\
&\quad + \sum_1^n \int_{u_{j-1}}^{u_j} \frac{\mathcal{E}Y^R(u,\beta)}{\mathcal{E}Y(u,\beta)}\mathrm{d}[N_i(u;\beta) - \mathcal{E}N_i(u;\beta)] \\
&\quad - \sum_1^n \int_{u_{j-1}}^{u_j} \left\{\frac{Y^R(u,\beta)}{Y(u,\beta)} - \frac{\mathcal{E}Y^R(u,\beta)}{\mathcal{E}Y(u,\beta)}\right\}\mathrm{d}N_i(v,\beta) \\
&= o(n^{1/2+\varepsilon}).
\end{aligned}
$$

Hence result (i) of this theorem can be established in the above manner. Result (ii) of this theorem can be proven by using the above results and an argument similar with that of the Lemma 5 of Ying (1993) for $\Delta_j \bar{M}_i(u;\beta) - \Delta_j \bar{M}_i(u;\beta_0) = n\{\Delta_j A_{i,n}(u)(\beta - \beta_0) + r_{j,n}(\beta)\}$ with $r_{j,n}(\beta)$ satisfying $\sup_{|\beta-\beta_0|\le dn} \frac{r_{i,n}(u)(\beta)}{n^{-1/2}+|\beta-\beta_0|)} = o(n^{\varepsilon})$. □

Proof of Theorem 3 The consistency and asymptotic normality of $\hat{\beta}_{k(n)}^{\mathrm{OEE}}$ are immediate consequences of results (i) and (ii) of Theorem 2. An approximated log-likelihood function of β based on $\{\Delta_j M_1(u;\beta), \Delta_j M_2(u;\beta_0) : j = 1,\ldots,k(n)\}$, denoted by $l_{k(n)}(\beta)$, can be derived as

$$
l_{k(n)}(\beta) = \sum_1^{k(n)} \{\Delta_j M_1(u;\beta), \Delta_j M_2(u;\beta)\}\{\Delta_j \hat{H}\}^{-1}\{\Delta_j M_1(u;\beta), \Delta_j M_2(u;\beta)\}^T.
$$

Further, via the asymptotical linearity in Theorem 2 and the derivative of $l_{k(n)}(\beta)$ with respect to β, we arrive at $\hat{\beta}_{k(n)}^{\mathrm{OEE}}$ estimate by solving the following equation:

$$
0 = \sum_{i=1}^{k(n)} \{\Delta_j A_{1,n}(u), \Delta_j A_{2,n}(u)\}\{\Delta_j \hat{H}\}^{-1}\{\Delta_j M_1(u;\beta), \Delta_j M_2(u;\beta)\}^T.
$$

We then have approximated equation for the estimate $\hat{\beta}_{k(n)}^{\mathrm{OEE}}$

$$
\begin{aligned}
&n^{1/2}(\hat{\beta}_k{}^{\mathrm{OEE}} - \beta_0) \\
&= \sum_1^{k(n)} \left\{(\Delta_j A_{1,n}(u), \Delta_j A_{2,n}(u))\{\Delta_j \hat{H}\}^{-1}(\Delta_j A_{1,n}(u), \Delta_j A_{2,n}(u))^T\right\}^{-1} \\
&\quad \times \{\Delta_j A_{1,n}(u), \Delta_j A_{2,n}(u)\}\{\Delta_j \hat{H}\}^{-1}\{\Delta_j M_1(u;\beta), \Delta_j M_2(u;\beta)\}^T.
\end{aligned}
$$

Hence the asymptotical variance of the normalized error $n^{1/2}(\hat{\beta}_{k(n)}^{\mathrm{OEE}} - \beta_0)$ is then calculated as the inverse of the following limit

$$\sigma_\infty^{-2} = \lim_{n\to\infty} \sum_1^{k(n)} (\Delta_j A_{1,n}(u), \Delta_j A_{2,n}(u)) \{\Delta_j \hat{H}\}^{-1} (\Delta_j A_{1,n}(u), \Delta_j A_{2,n}(u))^T,$$

$$= \lim_{n\to\infty} \sum_1^n \frac{1}{n} \int_0^\infty \left\{ \left[Z_i(v,\beta) - \frac{\mathcal{E}Y^R(v,\beta)}{\mathcal{E}Y(v,\beta)} \right] + \frac{\lambda'(v)}{\lambda(v)} \left[R_i(v,\beta) - \frac{\mathcal{E}Y^R(v,\beta)}{\mathcal{E}Y(v,\beta)} \right] \right\}^2 L_i(v,\beta)\,\mathrm{d}F(v).$$

□

Acknowledgments The author would like to thank the editor, associate editor and referees for their helpful comments and suggestions. This research was also supported in part by the National Science Council (NSC) and the Academia Sinica of Taiwan.

References

Aalen, O. O., Gjessing, H. K. (2001). Understanding the shape of the hazard rate: A process point of view. *Statistial Science, 16*, 1–22.

Andersen, P. K., Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Annals of Statistics, 10*, 1100–1120.

Andersen, P. K., Borgan, Ø., Gill, R. D., Keiding, N. (1993). *Statistical models based on counting processes.* New York: Springer.

Bickel, P. J., Doksum, K. A. (1981). An analysis of transformation revisited. *Journal of American Statistical Association, 76*, 296–311.

Box, G. E. P., Cox, D. R. (1964). An analysis of transformation. *Journal of Royal Statistical Society Series B, 26*, 211–243.

Carey, J. R., Liedo, P., Orozco, D., Vaupel, J. W. (1992). Slowing of mortality rates at older ages in large medfly cohort. *Science, 258*, 457–461.

Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of Royal Statistical Society Series B, 34*, 187–220.

Cox, D. R., Oakes, D. (1984). *Analysis of survival data.* London: Chapman & Hall.

Dabrowska, D. M., Doksum, K. A. (1988). Partial likelihood in transformation models with censored data. *Scandinavia Journal of Statistics, 15*, 1–24.

Escobar, L. A., Meeker, W. Q. (1992). Assessing influence in regression analysis with censored data. *Biometrics, 48*, 507–528.

Fygenson, M., Ritov, Y. (1994). Monotone estimating equations for censored data. *Annals of Statistics, 22,* 732–746.

Griffiths, W. S. (1988). Shock models. In *Encyclopedia of statistical sciences* (Vol. 8, pp. 442–446). New York: Wiley.

Heckman, J. J., Singer, B. (1984). A model for minimizing the impact of distribution assumptions in econometric models for duration data. *Econometrica, 52*, 271–320.

Horowitz, J. L. (1996). Semiparametric estimation of a regression model with an unknown transformation of the dependent variable. *Econometrica, 64*, 103–137.

Hougaard, P. (1986). Survival models for heterogeneous populations derived from stable distributions. *Biometrika, 73*, 387–396.

Hsieh, F. (1995). The empirical process approach for semiparametric two-sample models with heterogeneous treatment effect. *Journal of Royal Statistical Society Series B, 57*, 735–748.

Hsieh, F. (1996a). The empirical process approach for a generalized hazards model. *Biometrika, 83*, 519–528.

Hsieh, F. (1996b). Analysis of accelerated failure time model with heterogenous treatment effects. *Annals of Statistics, 24*, 2705–2719.

Hsieh, F. (1997). Estimations in homoscedastic linear regression models with censored data: An empirical process approach. *Annals of Statistics, 25,* 2665–2681.

Hsieh, F. (2001). On heteroscedastic hazards regression models: Theory and application. *Journal of Royal Statistical Society Series B, 63*, 63–79.

Hsieh, F. (2003). Patterns of age-specific mortality and molding of senescence by natural selection derived via the life-history regression model. *Evolutionary Ecology, 17*, 457–475.

Keiding, N., Andersen, P. K., Klein, J. P. (1998). The role of frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates. *Statistics in Medicine, 16*, 215–224.

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

Lin, D. Y., Ying, Z. (1995). Semiparametric inference for the accelerated life model with time-dependent covariates. *Journal of Statistical Planning and Inference, 44*, 47–63.

Murhpy, S. A., Rossini, A. J., van der Vaart, A. W. (1997). Maximum likelihood estimation in the proportional odds model. *Journal of American Statistical Association, 90*, 1399–1405.

Nielsen, G. G., Gill, R. D., Andersen, P. K., Sørensen, T. I. A. (1992). A counting process approach to maximum likelihood estimation in Frailty models. *Scandinavia Journal of Statistics, 19*, 25–43.

Ramlau-Hansen, H. (1983). Smoothing counting process intensities by means of kernel functions. *Annals of Statistics, 11,* 453–466.

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

Ritov, Y., Wellner, J. A. (1988). Censoring, martingales and the Cox model. In N. U. Prabhu (Ed.), *Contemporary mathematics. Statistical inference from stochastic processes* (Vol. 80, pp. 191–220).

Robins, J., Tsiatis, A. A. (1992). Semiparametric estimation of an accelerated failure time model with time-dependent covariates. *Biometrika, 79*, 311–319.

Shorack, G. R., Wellner, J. A. (1986). *Empirical processes with applications in statistics*. Wiley, New York.

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

Vaupel, J. W., Carey, J. R. (1993). Compositional interpretation of medfly mortality. *Science, 260*, 1666–1667.

Vaupel, J. W., Manton, K. G., Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography, 16*, 439–454.

Wei, L. J., Ying, Z., Lin, D. Y. (1990). Linear regression analysis of censored survival data based on rank test. *Biometrika, 77*, 845–852.

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