Biometrika (2017), **104**, 4, *pp.* 971–986
Printed in Great Britain
doi: 10.1093/biomet/asx055
Advance Access publication 3 October 2017

Robust rank estimation for transformation models with random effects

By YUANYUAN LIN
Department of Statistics, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong, China
ylin@sta.cuhk.edu.hk

YANG LUO
Department of Mathematics, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China
yluoac@ust.hk

SHANGYU XIE
School of Banking and Finance, University of International Business and Economics, Beijing 100029, China
xshyu0102@gmail.com

AND KANI CHEN
Department of Mathematics, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China
makchen@ust.hk

Summary

Semiparametric transformation models with random effects are useful in analysing recurrent and clustered data. With specified error and random effect distributions, Zeng & Lin (2007a) proved that nonparametric maximum likelihood estimators are semiparametric efficient. In this paper we consider a more general class of transformation models with random effects, under which an unknown monotonic transformation of the response is linearly related to the covariates and the random effects with unspecified error and random effect distributions. This includes many popular models. We propose an estimator based on the maximum rank correlation, which relies on symmetry of the random effect distribution, and establish its consistency and asymptotic normality. A random weighting resampling scheme is employed for inference. The proposed method can be extended to censored and clustered data. Numerical studies demonstrate that the proposed method performs well in practical situations. Application of the method is illustrated with the Framingham cholesterol data.

Some key words: Censored data; Clustered data; Maximum rank correlation; Random effect; Transformation model.

1. Introduction

Semiparametric models with random effects are useful in analysing dependent survival data. The dependence of correlated failure times is represented in the model through a random effect or frailty term, which complicates inference (Clayton & Cuzick, 1985; Oakes, 1989; Hougaard, 2000). The proportional hazards model with gamma frailties has been extensively studied in the literature; see Nielsen et al. (1992), Klein (1992), Murphy (1994, 1995), Andersen et al. (1997) and Parner (1998). Mathematical tractability makes the gamma frailty popular, but it implies a restrictive form of dependence among failure times. Cai et al. (2002) proposed estimating equation-based inference for linear transformation models with random effects for clustered failure time data. For correlated failure time data, Zeng et al. (2005) considered a proportional odds model with Gaussian random effects and developed a nonparametric maximum likelihood estimator which was found to be more efficient than that of Cai et al. (2002). A class of semiparametric transformation models with random effects was introduced in Zeng & Lin (2007a). Zeng & Lin (2007b) studied a class of semiparametric transformation models with random effects for the intensity function of the counting process for recurrent data. Zeng et al. (2009) generalized the usual gamma frailty model and proposed a class of transformation models for multivariate failure times. With the error and random effect/frailty distributions specified, they proposed a maximum likelihood estimator that is semiparametric efficient, together with a relatively simple and efficient EM algorithm. For multivariate failure time data, a number of works consider Bayesian inference with frailty cure rate models or a general class of transformation models with gamma frailty; see Yin (2008), Nieto-Barajas & Yin (2008), Wienke (2011) and Castro et al. (2014).

In this paper, we consider a more general class of transformation models with random effects, under which an unknown monotonic transformation of the response is linearly related to the covariates and the random effects with unspecified error and random effect distributions. This class of models allows various frailty distributions and includes many popular models, such as the linear model, the proportional hazards model or the proportional odds model with random effects. For such a general class of models, the nonparametric maximum likelihood estimate is unavailable. An easy-to-implement estimation method that is insensitive to the choice of unknown monotonic function, the error distribution and the random effect distribution is much desired.

The maximum rank correlation based on Kendall's τ (Han, 1987) provides a nonparametric and distribution-free estimator for transformation models or generalized accelerated failure time models. Khan & Tamer (2007) extended this approach to accommodate censoring and proposed the partial rank estimator. A smoothed version was considered by Song et al. (2007) and a similar monotone rank estimator was introduced in Cavanagh & Sherman (1998). In this paper, for the general transformation models with random effects, we propose an estimator based on the maximum rank correlation which relies only on the symmetry of the random effect distribution. Motivated by the partial rank estimator (Khan & Tamer, 2007), the method can be easily extended to handle censored data. With or without censoring, estimation of the transformation function, similar to that of Zeng & Lin (2007b), is not required. With the help of a direct search algorithm, computation of the proposed estimator is feasible and straightforward. A random weighting resampling scheme is used for variance estimation.

2. Model and inference

We consider the linear transformation model

$$H(T) = \tilde{\beta}_0^{\mathrm{T}}\tilde{X} + b^{\mathrm{T}}Z + \epsilon, \tag{1}$$

where $H(\cdot)$ is an unknown increasing function, T is the time to event, $\tilde{\beta}_0$ is a (d_1+1)-vector of unknown regression coefficients, b is a set of d_2-dimensional subject-specific unobservable random effects with an unknown but symmetric distribution about μ_0, $\tilde{X}$ and Z are the covariates associated with the fixed and random effects, and ϵ is an error term with unspecified distribution, independent of $\tilde{X}$, Z and b. In the absence of the random effects b, model (1) reduces to the general transformation model or the generalized accelerated failure time model. If the transformation $H(\cdot)$ is specified, model (1) is the semiparametric linear random effects model. If $\tilde{X}$ and Z are assumed to be independent, model (1) could be regarded as a heteroscedastic transformation model. Without assuming a known error distribution, a known random effect distribution, or the independence of $\tilde{X}$ and Z, model (1) is general enough to include many popular models and allows various frailty distributions, so it avoids possible model misspecification.

The objects of interest are $\tilde{\beta}_0$ and μ_0. For identifiability, we restrict the first component of $\tilde{\beta}_0$ to be 1, that is, $\tilde{\beta}_0 = (1, \beta_0^{\mathrm{T}})^{\mathrm{T}}$, where β_0 denotes the other components. Write $\theta_0 = (\beta_0^{\mathrm{T}}, \mu_0^{\mathrm{T}})^{\mathrm{T}}$ and $\theta = (\beta^{\mathrm{T}}, \mu^{\mathrm{T}})^{\mathrm{T}}$. Accordingly, we can decompose $\tilde{X} = (X^0, X)$, where X^0 is the predictor corresponding to the fixed regression coefficient and X is the other d_1-dimensional covariate. Let $(T_i, \tilde{X}_i, Z_i, \epsilon_i)$ $(i = 1, \ldots, n)$ be independent copies of $(T, \tilde{X}, Z, \epsilon)$. For complete data, the observations are $(T_i, \tilde{X}_i, Z_i)$ $(i = 1, \ldots, n)$. Motivated by the maximum rank correlation estimator in Han (1987), our proposed estimator of θ_0 is defined as the maximizer of

$$U_n(\beta, \mu) = \sum_{i \neq j} I\left(T_i < T_j\right) I\left(X_i^0 + \beta^{\mathrm{T}} X_i + \mu^{\mathrm{T}} Z_i < X_j^0 + \beta^{\mathrm{T}} X_j + \mu^{\mathrm{T}} Z_j\right), \qquad (2)$$

where $I(\cdot)$ is the indicator function. Let $\hat{\theta}_n \equiv \arg\max_{\beta,\mu} U_n(\beta, \mu)$. Some regularity conditions are assumed. Let $d = d_1 + d_2$.

Condition 1. The unknown parameter (β_0, μ_0) lies in a bounded set $\mathbb{B} \subset \mathbb{R}^d$.

Condition 2. The covariates $(\tilde{X}, Z)$ are of full rank, and X^0 has an everywhere-positive Lebesgue density conditional on (X, Z).

Condition 3. The density function of the random effect b is symmetric about μ_0, independent of $\tilde{X}$, Z and ϵ.

Let $W = (X^{\mathrm{T}}, Z^{\mathrm{T}})^{\mathrm{T}}$ and $\theta = (\beta^{\mathrm{T}}, \mu^{\mathrm{T}})^{\mathrm{T}}$. Let $\Omega = (Y, X^0, W)$ denote an observation from the distribution P on the set $S \subseteq \mathbb{R} \otimes \mathbb{R} \otimes \mathbb{R}^d$. For each $\omega = (y, x, w)$ in S and each θ in $\mathbb{B}$, define

$$\tau(\omega, \theta) = E\left\{I(Y < y) I\left(X^0 + \theta^{\mathrm{T}} W < x + \theta^{\mathrm{T}} w\right)\right\} + E\left\{I(Y > y) I\left(X^0 + \theta^{\mathrm{T}} W > x + \theta^{\mathrm{T}} w\right)\right\}.$$

Write ∇_m for the mth partial derivative operator of a function σ with respect to $\theta = (\theta_1, \ldots, \theta_d) \in \mathbb{R}^d$, and let

$$|\nabla_m| \sigma(\theta) = \sum_{i_1, \ldots, i_m} \left| \frac{\partial^m}{\partial \theta_{i_1} \cdots \partial \theta_{i_m}} \sigma(\theta) \right|.$$

The next two conditions guarantee a Taylor expansion of $\tau(\omega, \cdot)$ about θ_0:

Condition 4. $E\left\{|\nabla_1 \tau(\cdot, \theta_0)|^2\right\} < \infty$;

Condition 5. $E\left\{|\nabla_2| \tau(\cdot, \theta_0)\right\} < \infty$.

Condition 6. The matrix $E\{\nabla_2\tau(\cdot,\theta_0)\}$ is negative definite.

THEOREM 1. *Assume Conditions* 1–6 *hold. Then, $\hat\theta_n$ is consistent for θ_0, and*

$$n^{1/2}(\hat\theta_n-\theta_0)\to N\left\{0,A^{-1}B(A^{-1})^{\mathrm{T}}\right\}$$

in distribution as $n\to\infty$, where $A=-E\{\ddot h(\theta_0,T_1,T_2)I(T_1<T_2)\}$, $B=\mathrm{var}\{\dot b_1(\theta_0)\}$ and

$$\begin{aligned}
b_i(\theta)&=E[a_{ij}(\theta)+a_{ji}(\theta)-2E\{a_{ij}(\theta)\}\mid \tilde X_i,Z_i,T_i],\\
a_{ij}(\theta)&=\{I(X_i^0+\beta^{\mathrm{T}}X_i+\mu^{\mathrm{T}}Z_i<X_j^0+\beta^{\mathrm{T}}X_j+\mu^{\mathrm{T}}Z_j)\\
&\qquad-I(X_i^0+\beta_0^{\mathrm{T}}X_i+\mu_0^{\mathrm{T}}Z_i<X_j^0+\beta_0^{\mathrm{T}}X_j+\mu_0^{\mathrm{T}}Z_j)\}I(T_i<T_j),\\
h(\theta,t,s)&=\mathrm{pr}\{X_1^0+\beta'X_1+\mu'Z_1<X_2^0+\beta'X_2+\mu'Z_2\mid T_1=t,T_2=s\},
\end{aligned}$$

with a superscript dot denoting a derivative.

Remark 1. Like the maximum rank correlation objective function, $U_n(\theta)$ depends on the responses only through their ranks, which are unchanged by the unknown increasing transformation $H(\cdot)$. Hence, $\hat\theta_n$ is invariant under $H(\cdot)$, whose estimation is avoided.

Remark 2. Unlike the nonparametric maximum likelihood estimator (Zeng & Lin, 2007a), which requires parametric specification of the error and random effect distributions, the proposed estimator is distribution-free, only assuming symmetry of the random effect distribution. This makes the proposal appealing and easy to implement in practice.

Remark 3. Minimization of (2) can be carried out by the Nelder–Mead simplex search (Nelder & Mead, 1965), which does not require convexity or continuity. In simulation studies, we use this algorithm to search with different starting values.

The limiting covariance matrix of $\hat\theta_n$ in Theorem 1 is complicated and could be difficult to approximate. To circumvent this, we propose a random weighting approximation by generating independent and identically distributed nonnegative random variables $e_1,\ldots,e_n$ with mean 1 and variance 1. Define

$$\tilde U_n(\beta,\mu)=\sum_{i\neq j}e_ie_jI\left(T_i<T_j\right)I\left(X_i^0+\beta^{\mathrm{T}}X_i+\mu^{\mathrm{T}}Z_i<X_j^0+\beta^{\mathrm{T}}X_j+\mu^{\mathrm{T}}Z_j\right)\tag{3}$$

and $\tilde\theta_n=\arg\max_{\beta,\mu}\tilde U_n(\beta,\mu)$. The distribution of $n^{1/2}(\hat\theta_n-\theta_0)$ can be approximated by the resampling distribution of $n^{1/2}(\tilde\theta_n-\hat\theta_n)$ given the data $(T_i,\tilde X_i,Z_i)$ $(i=1,\ldots,n)$.

PROPOSITION 1. *Given $\{(T_i,\tilde X_i,Z_i),i=1,\ldots,n\}$, under Conditions* 1–6, *$n^{1/2}(\tilde\theta_n-\hat\theta_n)\to N\left\{0,A^{-1}B(A^{-1})^{\mathrm{T}}\right\}$ in distribution as $n\to\infty$, which is the limiting distribution of $n^{1/2}(\hat\theta_n-\theta_0)$.*

The random weighting resampling for a U-statistic objective function in (3) is given by Jin et al. (2001). The proof of Proposition 1 is omitted.

3. Random effect models for censored data

Time-to-event data are often subject to right censoring. For model (1) under random censoring, the observations are $(Y_i, \tilde{X}_i, Z_i, \delta_i)$ $(i = 1, \ldots, n)$, where $Y_i = \min(T_i, C_i), \delta_i = I(T_i \leqslant C_i)$ and C_i is the censoring variable. We assume that C is independent of T conditional on $(\tilde{X}, Z)$.

Analogous to the partial rank estimator in Khan & Tamer (2007), the proposed estimator of θ_0 is the maximizer of

$$U_n(\beta, \mu) = \sum_{i \neq j} \delta_i I\left(Y_i < Y_j\right) I\left(X_i^0 + \beta^{\mathrm{T}} X_i + \mu^{\mathrm{T}} Z_i < X_j^0 + \beta^{\mathrm{T}} X_j + \mu^{\mathrm{T}} Z_j\right). \tag{4}$$

Let $\hat{\theta}_n \equiv \arg\max_{\beta,\mu} U_n(\beta, \mu)$. Aside from Conditions 1–6, we also assume the following.

Condition 7. Let $\mathcal{S}$ denote the support of $W = (\tilde{X}^{\mathrm{T}}, Z^{\mathrm{T}})^{\mathrm{T}}$, and let $\mathcal{X}_{\mathrm{uc}}$ denote the set $\mathcal{X}_{\mathrm{uc}} = \{w \in \mathcal{S} : \mathrm{pr}(\delta_i = 1 \mid W_i = w) > 0\}$. Then, $\mathrm{pr}(\mathcal{X}_{\mathrm{uc}}) > 0$.

Condition 8. The set $\mathcal{X}_{\mathrm{uc}}$ is not contained in any proper linear subspace of $\mathbb{R}^{d+1}$.

Condition 9. Let $\eta_i = X_i^0 + \beta_0^{\mathrm{T}} X_i + b_i^{\mathrm{T}} Z_i$ and $\eta_j = X_j^0 + \beta_0^{\mathrm{T}} X_j + b_j^{\mathrm{T}} Z_j$. If $X_i^0 + \beta_0^{\mathrm{T}} X_i + \mu_0^{\mathrm{T}} Z_i < X_j^0 + \beta_0^{\mathrm{T}} X_j + \mu_0^{\mathrm{T}} Z_j$, then for any $t \in \mathbb{R}$, $\mathrm{pr}(\eta_i < t) > \mathrm{pr}(\eta_j < t)$.

Condition 7 requires that the probability of censoring does not equal 1 for all $w \in \mathcal{S}$, while Condition 8 is stronger than a full-rank condition, but is usually required for the censored case. In the presence of censoring and random effects, Condition 9 is imposed to prove the consistency of $\hat{\theta}_n$.

Theorem 2. *Under Conditions* 1–9, $\hat{\theta}_n$ *is consistent for* θ_0, *and* $n^{1/2}(\hat{\theta}_n - \theta_0) \to N\left\{0, A^{-1}B(A^{-1})^{\mathrm{T}}\right\}$ *in distribution as* $n \to \infty$, *where* $B = \mathrm{var}\{\dot{b}_1(\theta_0)\}$, $A = -E\{\ddot{h}(\theta_0) I(Y_1 < Y_2)\}$ *and*

$$\begin{aligned}
b_i(\theta) &= E\left[a_{ij}(\theta) + a_{ji}(\theta) - 2E\left\{a_{ij}(\theta)\right\} \Big| Y_i, \tilde{X}_i, Z_i, \delta_i\right], \\
a_{ij}(\theta) &= \Big\{\delta_i I\left(X_i^0 + \beta^{\mathrm{T}} X_i + \mu^{\mathrm{T}} Z_i < X_j^0 + \beta^{\mathrm{T}} X_j + \mu^{\mathrm{T}} Z_j\right) - \\
&\qquad \delta_i I\left(X_i^0 + \beta_0^{\mathrm{T}} X_i + \mu_0^{\mathrm{T}} Z_i < X_j^0 + \beta_0^{\mathrm{T}} X_j + \mu_0^{\mathrm{T}} Z_j\right)\Big\} I\left(Y_i < Y_j\right), \\
h(\theta) &= E\left\{\delta_1 I\left(X_1^0 + \beta^{\mathrm{T}} X_1 + \mu^{\mathrm{T}} Z_1 < X_2^0 + \beta^{\mathrm{T}} X_2 + \mu^{\mathrm{T}} Z_2\right) \Big| H(Y_1) < H(Y_2)\right\}.
\end{aligned}$$

Even though the objective function (4) takes a form similar to that for the partial rank estimator, it is more challenging in the presence of random effects. The same random weighting scheme in (3) can be employed for variance estimation. Though the partial rank correlation in (4) is defined for right censoring, it can be easily modified for left censoring.

4. Random effect models for clustered data

The failure times of the subjects within the same cluster are often correlated. In this section, we extend model (1) to allow correlated failure time data. Suppose there is a random sample of n clusters from an underlying population and there are K_i members in the ith cluster. In the presence of censoring, let T_{ik} be the failure time, C_{ik} be the censoring time, $Y_{ik} = \min(T_{ik}, C_{ik})$ be the observed time and $\delta_{ik} = I(Y_{ik} \leqslant C_{ik})$ be the failure or censoring indicator $(i = 1, \ldots, n; k =$

$1,\ldots,K_i)$, where i pertains to a cluster and k to the individual within a cluster. We consider the transformation model

$$H(T_{ik}) = \tilde{\beta}_0^{\mathrm{T}}\tilde{X}_{ik} + b_i^{\mathrm{T}}Z_{ik} + \epsilon_{ik} \quad (i = 1,\ldots,n;\ k = 1,\ldots,K_i), \tag{5}$$

where b_i are the unobservable random effects for the ith cluster, with an unspecified symmetric distribution about μ_0, and ϵ_{ik} $(i = 1,\ldots,n; k = 1,\ldots,K_i)$ are independent and identically distributed with unspecified distribution, independent of the random effect b_i. The observations are $\{Y_{ik}, \delta_{ik}, \tilde{X}_{ik}, Z_{ik}, i = 1,\ldots,n, k = 1,\ldots,K_i\}$. The dependence of failure times among individuals of the same cluster is characterized through the random effect b_i. We consider the full rank correlation

$$\begin{aligned} &U_n^F(\beta,\mu) \\ &= \sum_{1\leqslant k\leqslant K_i, 1\leqslant l\leqslant K_j}\ \sum_{1\leqslant i,j\leqslant n} \delta_{ik} I\left(Y_{ik} < Y_{jl}\right) I\left(X_{ik}^0 + \beta^{\mathrm{T}}X_{ik} + \mu^{\mathrm{T}}Z_{ik} < X_{jl}^0 + \beta^{\mathrm{T}}X_{jl} + \mu^{\mathrm{T}}Z_{jl}\right) \\ &= \sum_{1\leqslant k\leqslant K_i, 1\leqslant l\leqslant K_j}\ \sum_{i\neq j} \delta_{ik} I\left(Y_{ik} < Y_{jl}\right) I\left(X_{ik}^0 + \beta^{\mathrm{T}}X_{ik} + \mu^{\mathrm{T}}Z_{ik} < X_{jl}^0 + \beta^{\mathrm{T}}X_{jl} + \mu^{\mathrm{T}}Z_{jl}\right) \\ &\quad + \sum_{i=1}^{n}\sum_{k\neq l}^{K_i} \delta_{ik} I(Y_{ik} < Y_{il}) I\left(X_{ik}^0 + \beta^{\mathrm{T}}X_{ik} + \mu^{\mathrm{T}}Z_{ik} < X_{il}^0 + \beta^{\mathrm{T}}X_{il} + \mu^{\mathrm{T}}Z_{il}\right). \end{aligned} \tag{6}$$

The robust rank estimator $(\hat{\beta}_n, \hat{\mu}_n)$ is the maximizer of $U_n^F(\beta,\mu)$.

Remark 4. The first term in (6) pertains to the ranks across different clusters, while the second term pertains to the ranks within the same cluster. Under the independent and identically distributed assumption of ϵ_{ik} $(i = 1,\ldots,n; k = 1,\ldots,K_i)$ in model (5), a theorem analogous to Theorem 2 can be established for $(\hat{\beta}_n, \hat{\mu}_n)$, as the second term in (6) is asymptotically negligible. Details are omitted.

5. SIMULATION STUDIES

In the first set of studies, we generate data from the model

$$\log T = X_0 + \beta_1 X_1 + \beta_2 X_2 + bZ + \epsilon, \tag{7}$$

where (X_0, X_1, X_2, Z) follows a multivariate normal distribution with mean $(1, 0, 0, 0)$, unit variance and equal covariance 0·2 and $(\beta_1, \beta_2) = (-1, 1)$. Two error distributions are used: the standard normal distribution and the extreme value distribution. The simulations are based on 1000 replications. The external random weights are standard exponential variables. The resampling size is $N = 500$. For the censored case, the censoring times are Un$[0, \tau]$, where τ is chosen to produce an approximately 40% censoring rate. In the numerical computation, we treat the coefficient of X_0 as a parameter β_0 in the optimization. After obtaining the estimate of $(\beta_0, \beta_1, \beta_2)$, we rescale it by its norm so that the first component is 1.

First, we consider symmetric normal and logistic random effect distributions. We also consider gamma and lognormal random effect distributions. We set the mean and variance of the random effect distributions to be 1 for simplicity. Thus, the true parameter of model (7) is $\theta_0 = (-1, 1, 1)$

Table 1. *Parameter estimation results* ($\times 10^{-2}$) *for simulation studies with different random effect distributions in the uncensored case*

	θ	BIAS	SE ($\epsilon \sim$ normal)	SEE	CP	BIAS	SE ($\epsilon \sim$ extreme value)	SEE	CP
					$b \sim$ normal				
$n = 200$	β_1	$-0{\cdot}14$	14·17	14·36	96·2	$-1{\cdot}79$	16·27	16·37	96·7
	β_2	0·84	15·84	15·29	95·7	1·58	17·33	17·37	96·3
	μ	0·41	18·99	17·11	94·4	1·74	21·33	19·14	96·0
$n = 400$	β_1	$-0{\cdot}77$	10·58	11·27	96·8	$-1{\cdot}22$	11·60	12·41	96·2
	β_2	0·95	12·00	12·38	96·7	1·49	13·20	13·53	96·2
	μ	0·59	13·95	13·78	96·2	1·12	15·23	15·01	94·8
					$b \sim$ logistic				
$n = 200$	β_1	$-0{\cdot}74$	14·51	14·23	95·2	$-0{\cdot}84$	15·34	16·51	96·0
	β_2	0·53	15·20	15·32	96·5	0·60	15·97	17·64	96·4
	μ	1·09	17·81	16·67	96·3	0·54	17·61	18·84	96·5
$n = 400$	β_1	$-0{\cdot}04$	10·54	11·05	96·9	$-0{\cdot}32$	11·57	12·05	96·5
	β_2	0·82	12·09	12·14	96·5	0·28	12·95	13·11	96·2
	μ	$-0{\cdot}43$	12·54	13·39	96·2	0·29	14·09	14·43	97·8
					$b \sim$ gamma				
$n = 200$	β_1	$-1{\cdot}62$	13·65	13·75	96·6	$-1{\cdot}86$	14·44	15·78	96·6
	β_2	1·07	14·75	14·60	96·2	1·66	17·16	17·00	97·0
$n = 400$	β_1	$-0{\cdot}91$	10·23	10·67	94·8	$-0{\cdot}81$	11·41	11·78	96·7
	β_2	0·86	11·48	11·77	95·4	0·78	12·29	12·81	96·0
					$b \sim$ lognormal				
$n = 200$	β_1	$-0{\cdot}52$	13·38	13·28	95·7	$-1{\cdot}66$	14·74	15·19	97·0
	β_2	0·55	14·39	14·10	96·6	1·80	16·39	15·95	96·1
$n = 400$	β_1	$-0{\cdot}88$	9·85	10·37	96·2	$-1{\cdot}29$	10·98	11·52	96·7
	β_2	1·10	10·83	11·37	97·8	1·36	12·07	12·53	96·3

For the asymmetric gamma and lognormal random effects, we only report the estimation results for β_1 and β_2. BIAS, the bias for the parameter estimate; SE, the empirical standard error; SEE, the estimated standard error, CP, the 95% coverage probability.

with normal and logistic random effects and with $n = 200, 400$. The results are summarized in Tables 1 and 2. For both uncensored and censored cases, the proposed estimators are virtually unbiased and stable in all the cases, whether or not the random effect distribution is symmetric. In Table 1, the standard errors based on random weighting resampling are generally close to the empirical standard errors.

Second, we also compute the nonparametric maximum likelihood estimate of Zeng & Lin (2007a) by the EM algorithm when the error distribution is both correctly specified and misspecified with normal random effects with mean 1 and variance 1. Table 3 shows that, for both censored and uncensored cases, when the error and the random effect distributions are correctly specified, the proposed method is competitive with the nonparametric maximum likelihood estimator. When the error distribution is misspecified, the proposed method is superior in terms of accuracy and stability.

Thirdly, we evaluate the performance of the proposed method with strong correlation between X and Z. The data are generated from model (7) but (X_0, X_1, X_2, Z) has a multivariate normal distribution with mean $(1, 0, 0, 0)$, unit variance and equal covariance 0·5. Table 4 suggests that the proposed method works well for both censored and uncensored cases.

Lastly, we generate clustered data with $K = 2$ from the model

$$\log T_k = X_{0k} + \beta_1 X_{1k} + \beta_2 X_{2k} + bZ_k + \epsilon_k \quad (k = 1, 2),$$

Table 2. *Parameter estimation results* ($\times 10^{-2}$) *for simulation studies with different random effect distributions in the censored case with* $n = 400$

	$\epsilon \sim$ normal				$\epsilon \sim$ extreme value			
θ	BIAS	SE	SEE	CP	BIAS	SE	SEE	CP
				$b \sim$ normal				
β_1	−0·68	12·90	13·04	96·4	−1·19	14·43	15·17	97·3
β_2	0·79	14·15	14·25	95·8	1·46	16·02	16·61	97·2
μ	0·41	16·68	15·79	95·4	1·44	18·03	18·02	96·6
				$b \sim$ logistic				
β_1	−1·28	12·44	12·98	96·1	−0·70	13·60	14·61	97·3
β_2	0·93	13·76	14·12	96·3	1·22	15·60	16·07	97·2
μ	0·96	15·56	15·65	96·7	0·80	16·80	17·31	98·0
				$b \sim$ gamma				
β_1	−0·34	11·95	12·53	96·3	−0·91	14·45	14·25	96·5
β_2	0·85	13·67	13·60	95·7	0·93	15·57	15·44	97·5
				$b \sim$ lognormal				
β_1	−1·15	11·97	12·04	96·6	−1·21	13·23	13·77	97·0
β_2	1·12	13·00	13·06	96·4	1·19	14·89	14·95	97·0

Table 3. *Parameter estimation results* ($\times 10^{-2}$)*: the proposed estimator and the nonparametric maximum likelihood estimation with normal random effect and* $n = 200$

				Uncensored				
	Proposed				NPMLE			
θ	BIAS	SE	SEE	CP	BIAS	SE	BIAS	SE
					Correctly specified		Misspecified as	
	$\epsilon \sim$ normal						$\epsilon \sim$ extreme value	
β_1	−0·14	14·17	14·36	96·2	0·50	9·61	−3·15	11·38
β_2	0·84	15·84	15·29	95·7	−0·53	9·59	3·43	11·27
μ	0·41	18·99	17·11	94·4	−1·49	11·93	4·47	13·32
	$\epsilon \sim$ extreme value						$\epsilon \sim$ logistic	
β_1	−1·79	16·27	16·37	96·7	0·54	9·95	−11·86	13·58
β_2	1·58	17·33	17·37	96·3	−1·80	10·61	16·91	13·22
μ	1·74	21·33	19·14	96·0	−1·20	12·57	10·97	13·92
				Censored				
	Proposed				NPMLE			
θ	BIAS	SE	SEE	CP	BIAS	SE	BIAS	SE
					Correctly specified		Misspecified as	
	$\epsilon \sim$ normal						$\epsilon \sim$ extreme value	
β_1	−1·14	18·13	18·77	95·8	1·99	10·35	−5·81	13·01
β_2	1·41	18·80	21·55	96·7	−3·02	10·91	8·41	13·49
μ	2·02	21·73	22·56	96·0	−4·07	12·59	5·84	14·21
	$\epsilon \sim$ extreme value						$\epsilon \sim$ logistic	
β_1	−2·57	20·15	23·86	97·2	−3·42	11·61	−12·96	16·37
β_2	2·73	22·44	28·09	97·3	2·73	12·30	16·39	15·67
μ	2·84	24·32	29·10	96·6	1·80	12·89	9·25	15·68

NPMLE, the nonparametric maximum likelihood estimation proposed by Zeng & Lin (2007a).

Table 4. *Parameter estimation results* ($\times 10^{-2}$) *for simulation studies with correlation* $\rho = 0{\cdot}5$ *among covariates and normal random effect*

θ	BIAS	SE	SEE	CP	BIAS	SE	SEE	CP
	Uncensored				Censored			
	$n = 200$				$n = 400$			
				$\epsilon \sim$ normal				
β_1	−0·00	15·44	16·25	96·3	−0·97	13·71	13·95	96·4
β_2	1·06	18·15	18·91	96·2	1·65	16·85	16·45	97·1
μ	0·64	20·14	20·23	95·8	0·85	17·85	17·53	95·7
				$\epsilon \sim$ extreme value				
β_1	−1·21	17·08	17·70	96·2	−0·67	15·60	16·18	97·3
β_2	2·55	20·85	20·13	95·7	1·38	18·87	19·16	97·2
μ	1·07	21·52	21·22	96·8	1·37	19·95	20·16	97·6

Table 5. *Parameter estimation results* ($\times 10^{-2}$) *for simulation studies with longitudinal data, normal random effect and* $n = 200$

	Uncensored						Censored					
	Proposed				NPMLE		Proposed				NPMLE	
θ	BIAS	SE	SEE	CP	BIAS	SE	BIAS	SE	SEE	CP	BIAS	SE
			$\epsilon \sim$ normal						$\epsilon \sim$ normal			
β_1	−0·96	10·57	11·28	96·9	1·70	8·55	−0·44	12·50	13·19	97·3	3·19	11·33
β_2	1·40	11·98	12·30	97·5	−2·47	10·18	0·72	14·05	14·40	97·0	−3·33	11·99
μ	0·25	13·90	13·96	96·9	−0·69	13·13	0·83	16·36	16·03	96·2	−1·69	15·67
			$\epsilon \sim$ extreme value						$\epsilon \sim$ extreme value			
β_1	−0·43	11·39	12·21	96·4	1·85	9·70	−0·25	14·39	14·72	95·9	−1·56	11·79
β_2	0·76	13·33	13·42	96·5	−1·46	10·88	0·39	15·41	15·97	96·9	2·48	12·97
μ	0·84	15·57	14·97	96·0	−1·57	13·63	0·95	18·63	17·59	96·5	−1·98	14·00

where $(X_{0k}, X_{1k}, X_{2k}, Z_k)$, b and ϵ_k are the same as in model (7). The results in Table 5 agree with the theory that the proposed method works well for dependent data.

6. Applications

We apply the proposed methods to the Framingham cholesterol data (Zhang & Davidian, 2001), which consist of 200 randomly selected participants' age at baseline, gender, and cholesterol level measured at the beginning of the study and then every two years for 10 years. We apply the proposed method to characterize change in cholesterol level over time, the effect of baseline covariates, and between-subject variation in cholesterol levels.

Zhang & Davidian (2001) fitted a semiparametric linear mixed model

$$T_{ik} = b_{i0} + \beta_1 \mathrm{age}_i + \beta_2 \mathrm{sex}_i + b_{i1} t_{ik} + \epsilon_{ik}.$$

Here, T_{ik} is cholesterol level divided by 100 at the kth time for subject i and t_{ik} is $(\mathrm{time} - 5)/10$, with time measured in years from baseline, where the transformations of level and time were done for reasons of numerical stability; age_i is age at baseline; sex_i is a gender indicator. In Zhang & Davidian (2001), the errors ϵ_{ik} are assumed to be independent $N(0, \sigma^2)$ and $b_i = (b_{0i}, b_{1i})^{\mathrm{T}} = \mu + RZ_i$, where Z_i are approximated by the seminonparametric representation of

Table 6. *Regression analysis results for the Framingham cholesterol data*

	Proposed		SNP	
Parameter	Estimate	SE	Estimate	SE
β (sex)	$-0{\cdot}0199$	0·0312	$-0{\cdot}0626$	0·0455
μ (time)	0·2425	0·0691	0·2817	0·0242

SNP, the method of Zhang & Davidian (2001).

Gallant & Nychka (1987) with an order parameter K. To avoid possible misspecification, we relax these parametric assumptions on the errors, the random effect b_i and the linear model structure. Based on preliminary analysis similar to that of Zhang & Davidian (2001), we consider fitting our proposed model (5), that is,

$$H(T_{ik}) = \beta_1 \text{age}_i + \beta_2 \text{sex}_i + b_{i1} t_{ik} + \epsilon_{ik},$$

where b_i and ϵ_{ik} are the same as in model (5). We standardize the coefficient of age to be unity. We present the estimates of β_2 and the mean μ of the random effect b_1 in Table 6, along with the results of Zhang & Davidian (2001) with tuning parameter $K = 1$. Our estimates have the same significances as those of Zhang & Davidian (2001); in particular, there is no significant difference in cholesterol level between women and men. Our rank method gives a smaller point estimate of the mean of the random effect, though it has larger standard error. The significantly positive effect of time suggests that the possibility of a subpopulation having higher baseline cholesterol is high, even after adjusting for age and gender.

7. REMARKS

Unlike maximum likelihood estimation of the commonly used transformation models with random effects and known error distribution in survival analysis, our method cannot estimate the transformation function, so it cannot be used for prediction. It is also very difficult to recover the entire distribution of b under our model assumptions. However, the proposed estimator can serve as a check on the validity of estimates obtained under parametric assumptions, or as an initial/consistent estimate for other semiparametric approaches.

ACKNOWLEDGEMENT

The authors are indebted to the editor, the associate editor and two reviewers for comments that led to substantial improvements in the paper. The work was supported by the Research Grant Council of Hong Kong, The Chinese University of Hong Kong, the Natural Science Foundation of China and the Fundamental Research Funds for the Central Universities

APPENDIX

Proof of Theorem 1

Define

$$g(\theta) = E\left\{I\left(T_1 < T_2\right) I\left(X_1^0 + \beta^{\mathrm{T}} X_1 + \mu^{\mathrm{T}} Z_1 < X_2^0 + \beta^{\mathrm{T}} X_2 + \mu^{\mathrm{T}} Z_2\right)\right\},$$

$$g_n(\theta) = \frac{1}{n^2 - n} \sum_{i \neq j} I\left(T_i < T_j\right) I\left(X_i^0 + \beta^{\mathrm{T}} X_i + \mu^{\mathrm{T}} Z_i < X_j^0 + \beta^{\mathrm{T}} X_j + \mu^{\mathrm{T}} Z_j\right).$$

To establish consistency, we first show that $E\{g_n(\theta)\}$ has its maximum at $\theta = \theta_0$. Write $W = (X^{\mathrm{T}}, Z^{\mathrm{T}})^{\mathrm{T}}$ and $\theta = (\beta^{\mathrm{T}}, \mu^{\mathrm{T}})^{\mathrm{T}}$. For $i \neq j$, we have that

$$\begin{aligned}
&E\left\{I\left(T_i < T_j\right) I\left(X_i^0 + \beta^{\mathrm{T}} X_i + \mu^{\mathrm{T}} Z_i < X_j^0 + \beta^{\mathrm{T}} X_j + \mu^{\mathrm{T}} Z_j\right) \mid X_i^0, X_j^0, W_i, W_j\right\} \\
&\quad = E\left\{I\left(T_i < T_j\right) \mid X_i^0, X_j^0, W_i, W_j\right\} I\left(X_i^0 + \beta^{\mathrm{T}} X_i + \mu^{\mathrm{T}} Z_i < X_j^0 + \beta^{\mathrm{T}} X_j + \mu^{\mathrm{T}} Z_j\right) \\
&\quad = \mathrm{pr}\left(X_i^0 + \beta_0^{\mathrm{T}} X_i + b_i^{\mathrm{T}} Z_i + \epsilon_i < X_i^0 + \beta_0^{\mathrm{T}} X_j + b_j^{\mathrm{T}} Z_j + \epsilon_j \mid X_i^0, X_j^0, W_i, W_j\right) \\
&\qquad \times I(X_i^0 + \theta^{\mathrm{T}} W_i < X_j^0 + \theta^{\mathrm{T}} W_j) \\
&\quad \equiv \xi_{ij} I\left(X_i^0 + \theta^{\mathrm{T}} W_i < X_j^0 + \theta^{\mathrm{T}} W_j\right),
\end{aligned}$$

where $\xi_{ij} = \mathrm{pr}(X_i^0 + \beta_0^{\mathrm{T}} X_i + b_i^{\mathrm{T}} Z_i + \epsilon_i < X_j^0 + \beta_0^{\mathrm{T}} X_j + b_j^{\mathrm{T}} Z_j + \epsilon_j \mid X_i^0, X_j^0, W_i, W_j)$. A direct calculation yields that

$$\xi_{ij} = \mathrm{pr}\{\epsilon_i - \epsilon_j + (b_i - \mu_0)^{\mathrm{T}} Z_i - (b_j - \mu_0)^{\mathrm{T}} Z_j < X_j^0 + \theta_0^{\mathrm{T}} W_j - X_i^0 - \theta_0^{\mathrm{T}} W_i \mid X_i^0, X_j^0, W_i, W_j\}.$$

Under our model assumptions, $\epsilon_i - \epsilon_j + (b_i - \mu_0)^{\mathrm{T}} Z_i - (b_j - \mu_0)^{\mathrm{T}} Z_j$ is symmetric about zero, so ξ_{ij} is greater than 0·5 if $X_i^0 + \theta_0^{\mathrm{T}} W_i < X_j^0 + \theta_0^{\mathrm{T}} W_j$; otherwise, ξ_{ij} is less than 0·5 if $X_i^0 + \theta_0^{\mathrm{T}} W_i > X_j^0 + \theta_0^{\mathrm{T}} W_j$. Hence, the conditional expectation of each pair in the summation of $g_n(\theta)$ given X_i^0, X_j^0, W_i, W_j is

$$\begin{aligned}
&E\{I(T_i < T_j) \mid X_i^0, X_j^0, W_i, W_j\} I(X_i^0 + \theta^{\mathrm{T}} W_i < X_j^0 + \theta^{\mathrm{T}} W_j) \\
&\quad + E\{I(T_j < T_i) \mid X_i^0, X_j^0, W_i, W_j\} I(X_j^0 + \theta^{\mathrm{T}} W_j < X_i^0 + \theta^{\mathrm{T}} W_i) \\
&= \begin{cases} \xi_{ij}, & X_i^0 + \theta^{\mathrm{T}} W_i < X_j^0 + \theta^{\mathrm{T}} W_j, \\ \xi_{ji}, & X_j^0 + \theta^{\mathrm{T}} W_j < X_i^0 + \theta^{\mathrm{T}} W_i, \end{cases}
\end{aligned} \tag{A1}$$

whose value may be greater than 0·5 or less than 0·5, depending on θ. However, when $\theta = \theta_0$, the quantity in (A1) would always take the value greater than 0·5 and hence $g(\theta)$ is maximized by taking the expectation over X_i^0, X_j^0, W_i, W_j. As a result, $E\{g_n(\theta)\}$ is maximized at $\theta = \theta_0$.

A key step for consistency is the uniform convergence $\sup_\theta |g_n(\theta) - g(\theta)| = o_{\mathrm{p}}(1)$. Similar to Khan & Tamer (2007, p. 272), this can be shown by the uniform laws of large numbers for U-statistics with bounded kernel functions satisfying a Euclidean property. The class of functions inside the summation of the maximum rank correlation objective function (2) can be proved to be Euclidean with constant envelope 1 from the identical subgraph set and Vapnik–Chervonenkis class set arguments of Sherman (1993, § 5). Once the Euclidean property is shown, we apply Corollary 7 in Sherman (1994, p. 450) and hence prove uniform convergence.

What remains for proving consistency is to establish compactness and continuity. Compactness holds under Condition 1. By Condition 2, the distribution of the random variable $X^0 + \theta_0^{\mathrm{T}} W$ is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}$, so $g(\theta)$ is continuous at $\theta = \theta_0$. Hence, consistency is proved.

To establish asymptotic normality, let $\epsilon_n(\theta) = g_n(\theta) - g(\theta)$. A standard decomposition of U-statistics gives

$$\epsilon_n(\theta) - \epsilon_n(\theta_0) = \frac{1}{n}\sum_{i=1}^{n} b_i(\theta) + \frac{1}{n^2 - n}\sum_{i<j} d_{ij}(\theta),$$

where

$$\begin{aligned}
b_i(\theta) &= E[a_{ij}(\theta) + a_{ji}(\theta) - 2E\{a_{ij}(\theta)\} \mid \tilde{X}_i, Z_i, T_i], \\
d_{ij}(\theta) &= a_{ij}(\theta) + a_{ji}(\theta) - 2E\{a_{ij}(\theta)\} - b_i(\theta) - b_j(\theta), \\
a_{ij}(\theta) &= \{I(X_i^0 + \theta^{\mathrm{T}} W_i < X_j^0 + \theta^{\mathrm{T}} W_j) - I(X_i^0 + \theta_0^{\mathrm{T}} W_i < X_j^0 + \theta_0^{\mathrm{T}} W_j)\} I(T_i < T_j).
\end{aligned}$$

Note that $E\{b_i(\theta)\} = 0$. Under Condition 4, Taylor expansion gives

$$\frac{1}{n}\sum_{i=1}^{n} b_i(\theta) = (\theta-\theta_0)^{\mathrm{T}}\frac{1}{n}\sum_{i=1}^{n}\dot{b}_i(\theta_0) + o_{\mathrm{p}}(\|\theta-\theta_0\|^2).$$

Similar to the second step of the consistency proof, we need to show that, for any sequence r_n of order $o(1)$,

$$\sup_{\|\theta-\theta_0\|\leqslant r_n}\left|\frac{1}{n^2-n}\sum_{i<j} d_{ij}(\theta)\right| = o_{\mathrm{p}}(n^{-1}). \tag{A2}$$

Again by applying the identical subgraph set and Vapnik–Chervonenkis class set arguments of Sherman (1993, § 5), together with Corollary 17 and Corollary 21 in Nolan & Pollard (1987), we can show that the class of function $d_{ij}(\theta)$ is Euclidean. The Euclidean property together with Corollary 7 of Sherman (1994) guarantee that (A2) holds.

For θ in a neighbourhood of θ_0, we have

$$\begin{aligned}
g(\theta) &= \int_{t_1<t_2} h(\theta,t_1,t_2)f_T(t_1)f_T(t_2)\,\mathrm{d}t_1\,\mathrm{d}t_2 \\
&= \int_{t_1<t_2} h(\theta_0,t_1,t_2)f_T(t_1)f_T(t_2)\,\mathrm{d}t_1\,\mathrm{d}t_2 \\
&\quad + (\theta-\theta_0)^{\mathrm{T}}\int_{t_1<t_2} \dot{h}(\theta_0,t_1,t_2)f_T(t_1)f_T(t_2)\,\mathrm{d}t_1\,\mathrm{d}t_2 \\
&\quad + \frac{1}{2}(\theta-\theta_0)^{\mathrm{T}}\left\{\int_{t_1<t_2} \ddot{h}(\theta_0,t_1,t_2)f_T(t_1)f_T(t_2)\,\mathrm{d}t_1\,\mathrm{d}t_2\right\}(\theta-\theta_0) + o_{\mathrm{p}}\left(\|\theta-\theta_0\|^2\right) \\
&= a - \frac{1}{2}(\theta-\theta_0)^{\mathrm{T}}A(\theta-\theta_0) + o_{\mathrm{p}}(\|\theta-\theta_0\|^2),
\end{aligned}$$

where $a = E\{h(\theta_0,T_1,T_2)I(T_1<T_2)\}$, $A = -E\{\ddot{h}(\theta_0,T_1,T_2)I(T_1<T_2)\}$ and $h(\theta,t_1,t_2) = \mathrm{pr}(X_1^0 + \theta^{\mathrm{T}}W_1 < X_2^0 + \theta^{\mathrm{T}}W_2 \mid T_1 = t_1, T_2 = t_2)$. Note that the matrix A is invertible under Condition 6. It then follows that

$$\begin{aligned}
g_n(\theta) &= g(\theta) + \epsilon_n(\theta) \\
&= -\frac{1}{2}(\theta-\theta_0)^{\mathrm{T}}A(\theta-\theta_0) + (\theta-\theta_0)^{\mathrm{T}}\frac{1}{n}\sum_{i=1}^{n}\dot{b}_i(\theta_0) \\
&\quad + a + \epsilon_n(\theta_0) + o_{\mathrm{p}}(\|\theta-\theta_0\|^2) + o_{\mathrm{p}}(n^{-1}) \\
&= f_n(\theta) + a + \epsilon_n(\theta_0) + o_{\mathrm{p}}(n^{-1}),
\end{aligned}$$

where

$$\begin{aligned}
f_n(\theta) &= -\frac{1}{2}(\theta-\theta_0)^{\mathrm{T}}A(\theta-\theta_0) + (\theta-\theta_0)^{\mathrm{T}}\frac{1}{n}\sum_{i=1}^{n}\dot{b}_i(\theta_0) + o_{\mathrm{p}}\left(\|\theta-\theta_0\|^2\right) \\
&= -\frac{1}{2}(\theta-\theta_0)^{\mathrm{T}}A_n(\theta-\theta_0) + (\theta-\theta_0)^{\mathrm{T}}\frac{1}{n}\sum_{i=1}^{n}\dot{b}_i(\theta_0) \\
&= -\frac{1}{2}\left[A_n^{1/2}\left\{\theta-\theta_0-A_n^{-1}\frac{1}{n}\sum_{i=1}^{n}\dot{b}_i(\theta_0)\right\}\right]^{\mathrm{T}}\left[A_n^{1/2}\left\{\theta-\theta_0-A_n^{-1}\frac{1}{n}\sum_{i=1}^{n}\dot{b}_i(\theta_0)\right\}\right] \\
&\quad + \frac{1}{2}\left\{\frac{1}{n}\sum_{i=1}^{n}\dot{b}_i(\theta_0)\right\}^{\mathrm{T}}A_n^{-1}\left\{\frac{1}{n}\sum_{i=1}^{n}\dot{b}_i(\theta_0)\right\},
\end{aligned}$$

in which $o_p(\|\theta-\theta_0\|^2)=c_n\|\theta-\theta_0\|^2$, $c_n=o_p(1)$ and $A_n=A-2c_nI$. Hence, the maximizer of $f_n(\theta)$ is $\hat{\gamma}_n=\theta_0+A_n^{-1}n^{-1}\sum_{i=1}^n \dot{b}_i(\theta_0)$.

Since $\hat{\theta}_n$ is the maximizer of g_n,

$$\begin{aligned}
0 &\leqslant f_n(\hat{\gamma}_n)-f_n(\hat{\theta}_n)\\
&=\left\{f_n\left(\hat{\gamma}_n\right)+a+\epsilon_n(\theta_0)-g_n\left(\hat{\gamma}_n\right)\right\}-\left\{f_n\left(\hat{\theta}_n\right)+a+\epsilon_n(\theta_0)-g_n\left(\hat{\theta}_n\right)\right\}\\
&\quad-\left\{g_n\left(\hat{\theta}_n\right)-g_n\left(\hat{\gamma}_n\right)\right\}\\
&\leqslant \left\{f_n\left(\hat{\gamma}_n\right)+a+\epsilon_n(\theta_0)-g_n\left(\hat{\gamma}_n\right)\right\}-\left\{f_n\left(\hat{\theta}_n\right)+a+\epsilon_n(\theta_0)-g_n\left(\hat{\theta}_n\right)\right\}\\
&=o_p\left(n^{-1}\right)+o_p\left(n^{-1}\right)=o_p\left(n^{-1}\right).
\end{aligned}\tag{A3}$$

On the other hand, in view of the expression for f_n,

$$f_n(\hat{\gamma}_n)-f_n(\hat{\theta}_n)=\frac{1}{2}\left[A_n^{1/2}\left\{\hat{\theta}_n-\theta_0-A_n^{-1}\frac{1}{n}\sum_{i=1}^n\dot{b}_i(\theta_0)\right\}\right]^{\mathrm{T}}\left[A_n^{1/2}\left\{\hat{\theta}_n-\theta_0-A_n^{-1}\frac{1}{n}\sum_{i=1}^n\dot{b}_i(\theta_0)\right\}\right].\tag{A4}$$

Combining (A3) and (A4), we obtain

$$\begin{aligned}
\hat{\theta}_n&=\theta_0+A_n^{-1}\frac{1}{n}\sum_{i=1}^n\dot{b}_i(\theta_0)+o_p\left(n^{-1/2}\right)\\
&=\theta_0+A^{-1}\frac{1}{n}\sum_{i=1}^n\dot{b}_i(\theta_0)+\left(A_n^{-1}-A^{-1}\right)\frac{1}{n}\sum_{i=1}^n\dot{b}_i(\theta_0)+o_p\left(n^{-1/2}\right)\\
&=\theta_0+A^{-1}\frac{1}{n}\sum_{i=1}^n\dot{b}_i(\theta_0)+o_p\left(n^{-1/2}\right).
\end{aligned}$$

The last equation holds because $A_n^{-1}-A^{-1}=o_p(1)$ and $n^{-1}\sum_{i=1}^n\dot{b}_i(\theta_0)=O_p(n^{-1/2})$. Thus, as $n\to\infty$,

$$n^{1/2}(\hat{\theta}_n-\theta_0)=A^{-1}n^{-1/2}\sum_{i=1}^n\dot{b}_i(\theta_0)+o_p(1)\to N(0,\Sigma)$$

in distribution, where $\Sigma=A^{-1}\mathrm{var}\{\dot{b}_1(\theta_0)\}(A^{-1})^{\mathrm{T}}$. The proof of Theorem 1 is complete

Proof of Theorem 2

Define $g(\theta)=E\{\delta_1 I(Y_1<Y_2)I(X_1^0+\beta^{\mathrm{T}}X_1+\mu^{\mathrm{T}}Z_1<X_2^0+\beta^{\mathrm{T}}X_2+\mu^{\mathrm{T}}Z_2)\}$ and $g_n(\theta)=1/\{n(n-1)\}\times\sum_{i\neq j}\delta_i I(Y_i<Y_j)I(X_i^0+\beta^{\mathrm{T}}X_i+\mu^{\mathrm{T}}Z_i<X_j^0+\beta^{\mathrm{T}}X_j+\mu^{\mathrm{T}}Z_j)$. Similar to the proofs of consistency in Theorem 1, we first show that $E\{g_n(\theta)\}$ has its maximizer at $\theta=\theta_0$. Write $W=(X^{\mathrm{T}},Z^{\mathrm{T}})^{\mathrm{T}}$ and $\theta=(\beta^{\mathrm{T}},\mu^{\mathrm{T}})^{\mathrm{T}}$. For $i\neq j$,

$$\begin{aligned}
&E\{\delta_i I(Y_i<Y_j)I(X_i^0+\beta^{\mathrm{T}}X_i+\mu^{\mathrm{T}}Z_i<X_j^0+\beta^{\mathrm{T}}X_j+\mu^{\mathrm{T}}Z_j)\mid X_i^0,X_j^0,W_i,W_j\}\\
&=E\{\delta_i I(Y_i<Y_j)\mid X_i^0,X_j^0,W_i,W_j\}I(X_i^0+\theta^{\mathrm{T}}W_i<X_j^0+\theta^{\mathrm{T}}W_j)\\
&=E[I\{X_i^0+\theta_0^{\mathrm{T}}W_i+(b_i-\mu_0)^{\mathrm{T}}Z_i+\epsilon_i\leqslant H(C_i)\}I\{H(Y_i)<H(Y_j)\}\mid X_i^0,X_j^0,W_i,W_j]\\
&\quad\times I(X_i^0+\theta^{\mathrm{T}}W_i<X_j^0+\theta^{\mathrm{T}}W_j)
\end{aligned}$$

$$
\begin{aligned}
&= \mathrm{pr}[X_i^0 + \theta_0^{\mathrm{T}} W_i + (b_i - \mu_0)^{\mathrm{T}} Z_i + \epsilon_i \leqslant \min\{X_j^0 + \theta_0^{\mathrm{T}} W_j + (b_j - \mu_0)^{\mathrm{T}} Z_j + \epsilon_j, \\
&\qquad H(C_i) \wedge H(C_j)\} \mid X_i^0, X_j^0, W_i, W_j] I(X_i^0 + \theta' W_i < X_j^0 + \theta' W_j) \\
&\equiv \xi_{ij} I(X_i^0 + \theta^{\mathrm{T}} W_i < X_j^0 + \theta^{\mathrm{T}} W_j),
\end{aligned}
$$

where

$$
\begin{aligned}
\xi_{ij} &= \mathrm{pr}[X_i^0 + \theta_0^{\mathrm{T}} W_i + (b_i - \mu_0)^{\mathrm{T}} Z_i + \epsilon_i \leqslant \min\{X_j^0 + \theta_0' W_j + (b_j - \mu_0)^{\mathrm{T}} Z_j + \epsilon_j, \\
&\qquad H(C_i) \wedge H(C_j)\} \mid X_i^0, X_j^0, W_i, W_j] \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{pr}\{X_i^0 + \theta_0^{\mathrm{T}} W_i + (b_i - \mu_0)^{\mathrm{T}} Z_i + t_1 \leqslant X_j^0 + \theta_0^{\mathrm{T}} W_j + (b_j - \mu_0)^{\mathrm{T}} Z_j + t_2\} \\
&\quad \times \mathrm{pr}\{H(C_i) \wedge H(C_j) \geqslant X_i^0 + \theta_0^{\mathrm{T}} W_i + (b_i - \mu_0)^{\mathrm{T}} Z_i + t_1\} f_\epsilon(t_1) f_\epsilon(t_2)\, \mathrm{d}t_1\, \mathrm{d}t_2.
\end{aligned}
$$

Analogously, we can derive an expression for ξ_{ji}, that is,

$$
\begin{aligned}
\xi_{ji} &= \mathrm{pr}[X_j^0 + \theta_0^{\mathrm{T}} W_j + (b_j - \mu_0)^{\mathrm{T}} Z_j + \epsilon_j \leqslant \min\{X_i^0 + \theta_0^{\mathrm{T}} W_i + (b_i - \mu_0)^{\mathrm{T}} Z_i + \epsilon_i, \\
&\qquad H(C_i) \wedge H(C_j)\} \mid X_i^0, X_j^0, W_i, W_j] \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{pr}\{X_j^0 + \theta_0^{\mathrm{T}} W_j + (b_j - \mu_0)^{\mathrm{T}} Z_j + t_1 \leqslant X_i^0 + \theta_0^{\mathrm{T}} W_i + (b_i - \mu_0)^{\mathrm{T}} Z_i + t_2\} \\
&\quad \times \mathrm{pr}\{H(C_i) \wedge H(C_j) \geqslant X_j^0 + \theta_0^{\mathrm{T}} W_j + (b_j - \mu_0)^{\mathrm{T}} Z_j + t_1\} f_\epsilon(t_1) f_\epsilon(t_2)\, \mathrm{d}t_1\, \mathrm{d}t_2.
\end{aligned}
$$

When $X_i^0 + \theta_0^{\mathrm{T}} W_i < X_j^0 + \theta_0^{\mathrm{T}} W_j$, under Conditions 3 and 9, we can show that

$$
\begin{aligned}
&\mathrm{pr}\{X_i^0 + \theta_0^{\mathrm{T}} W_i + (b_i - \mu_0)^{\mathrm{T}} Z_i + t_1 \leqslant X_j^0 + \theta_0^{\mathrm{T}} W_j + (b_j - \mu_0)^{\mathrm{T}} Z_j + t_2\} \\
&\quad > \mathrm{pr}\{X_j^0 + \theta_0^{\mathrm{T}} W_j + (b_j - \mu_0)^{\mathrm{T}} Z_j + t_1 \leqslant X_i^0 + \theta_0^{\mathrm{T}} W_i + (b_i - \mu_0)^{\mathrm{T}} Z_i + t_2\}.
\end{aligned} \tag{A5}
$$

This is because, under Condition 9,

$$
\begin{aligned}
\mathrm{pr}(\eta_i \leqslant \eta_j + t_2 - t_1) &= \int_{-\infty}^{\infty} \mathrm{pr}(\eta_i \leqslant x + t_2 - t_1)\, \mathrm{d}F_{\eta_j}(x) \\
&> \int_{-\infty}^{\infty} \mathrm{pr}(\eta_j \leqslant x + t_2 - t_1)\, \mathrm{d}F_{\eta_j}(x) \\
&= \mathrm{pr}(\eta_j \leqslant x + t_2 - t_1)\, \mathrm{d}F_{\eta_j}(x)\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} F_{\eta_j}(x)\, \mathrm{d}F_{\eta_j}(x + t_2 - t_1) \\
&= 1 - \int_{-\infty}^{\infty} F_{\eta_j}(x - t_2 + t_1)\, \mathrm{d}F_{\eta_j}(x) \\
&\geqslant 1 - \int_{-\infty}^{\infty} F_{\eta_i}(x - t_2 + t_1)\, \mathrm{d}F_{\eta_j}(x) \\
&= 1 - \left\{1 - \int_{-\infty}^{\infty} F_{\eta_j}(x)\, \mathrm{d}F_{\eta_i}(x - t_2 + t_1)\right\} \\
&= \int_{-\infty}^{\infty} F_{\eta_j}(x + t_2 - t_1)\, \mathrm{d}F_{\eta_i}(x) \\
&= \mathrm{pr}(\eta_j \leqslant \eta_i + t_2 - t_1).
\end{aligned}
$$

On the other hand, if $X_i^0 + \theta_0^{\mathrm{T}} W_i < X_j^0 + \theta_0^{\mathrm{T}} W_j$, it is not hard to check that under Conditions 3 and 9,

$$\begin{aligned} &\mathrm{pr}\{H(C_i) \wedge H(C_j) \geqslant X_i^0 + \theta_0^{\mathrm{T}} W_i + (b_i - \mu_0)^{\mathrm{T}} Z_i + t_1\} \\ &\quad > \mathrm{pr}\{H(C_i) \wedge H(C_j) \geqslant X_j^0 + \theta_0^{\mathrm{T}} W_j + (b_j - \mu_0)^{\mathrm{T}} Z_j + t_1\}. \end{aligned} \tag{A6}$$

Combining (A5) and (A6), we have shown that $\xi_{ij} > \xi_{ji}$ if $X_i^0 + \theta_0^{\mathrm{T}} W_i < X_j^0 + \theta_0^{\mathrm{T}} W_j$. Similarly, one can show that $\xi_{ji} > \xi_{ij}$ if $X_j^0 + \theta_0^{\mathrm{T}} W_j < X_i^0 + \theta_0^{\mathrm{T}} W_i$.

Hence, the conditional expectation of each pair in the summation of $g_n(\theta)$ given X_i^0, X_j^0, W_i, W_j is

$$\begin{aligned} &E\{\delta_i I(Y_i < Y_j)\} I(X_i^0 + \theta^{\mathrm{T}} W_i < X_j^0 + \theta^{\mathrm{T}} W_j) + E\{\delta_j I(Y_j < Y_i)\} \\ &\quad \times I(X_j^0 + \theta^{\mathrm{T}} W_j < X_i^0 + \theta^{\mathrm{T}} W_i) \\ &= \begin{cases} \xi_{ij}, & X_i^0 + \theta^{\mathrm{T}} W_i < X_j^0 + \theta^{\mathrm{T}} W_j, \\ \xi_{ji}, & X_j^0 + \theta^{\mathrm{T}} W_j < X_i^0 + \theta^{\mathrm{T}} W_i, \end{cases} \end{aligned} \tag{A7}$$

whose value depends on θ. However, when $\theta = \theta_0$, the quantity in (A7) would always take the larger value of ξ_{ij} and ξ_{ji}, and hence $g(\theta)$ is maximized by taking the expectation over W_i and W_j. As a result, $E\{g_n(\theta)\}$ is maximized at $\theta = \theta_0$.

What remains for proving consistency is to show $\sup_\theta |g_n(\theta) - g(\theta)| = o_{\mathrm{p}}(1)$ and that $\hat{\theta}_n = (\hat{\beta}_n, \hat{\mu}_n)$ converges to $\theta_0 = (\beta_0, \mu_0)$ in probability, which can be done along the lines of the proofs for Theorem 1. Asymptotic normality can be proved in the same fashion as in the uncensored case in the proof of Theorem 1; the proof is omitted.

References

Andersen, P. K., Klein, J. P., Knudsen, K. M. & Tabanera y Placios, R. (1997). Estimation of Cox's regression model with shared gamma frailties. *Biometrics* **53**, 1475–84.

Cai, T. Cheng, S. C. & Wei, L. J. (2002). Semiparametric mixed-effects models for clustered failure time data. *J. Am. Statist. Assoc.* **97**, 514–22.

Castro, M. Chen, M. H. Ibrahim, J. G. & Klein, J. P. (2014). Bayesian transformation models for multivariate survival data. *Scand. J. Statist.* **41**, 187–99.

Cavanagh, C. & Sherman, R. P. (1998). Rank estimators for monotonic index models. *J. Economet.* **84**, 351–81.

Clayton, D. & Cuzick, J. (1985). Multivariate generalizations of the proportional hazards model (with Discussion). *J. R. Statist. Soc.* A **148**, 82–117.

Gallant, A. R. & Nychka, D. W. (1987). Semi-nonparametric maximum likelihood estimation. *Econometrica* **55**, 363–90.

Han, A. K. (1987). Non-parametric analysis of a generalized regression model: The maximum rank correlation estimator. *J. Economet.* **35**, 303–16.

Hougaard, P. (2000). *Analysis of Multivariate Survival Data.* New York: Springer.

Jin, Z. Ying, Z. & Wei, L. J. (2001). A simple resampling method by perturbing the minimand. *Biometrika* **88**, 381–90.

Khan, S. & Tamer, E. (2007). Partial rank estimation of duration models with general forms of censoring. *J. Economet.* **136**, 251–80.

Klein, J. (1992). Semiparametric estimation of random effects using the Cox model based on the EM algorithm. *Biometrics* **48**, 795–806.

Murphy, S. A. (1994). Consistency in a proportional hazards model incorporating a random effect. *Ann. Statist.* **22**, 712–31.

Murphy, S. A. (1995). Asymptotic theory for the frailty model. *Ann. Statist.* **23**, 182–98.

Nelder, J. A. & Mead, R. (1965). A simplex algorithm for function minimization. *Comp. J.* **7**, 308–13.

Nielsen, G. G. Gill, R. D. Andersen, P. K. & Sorensen, T. I. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scand. J. Statist.* **19**, 25–43.

Nieto-Barajas, L. E. & Yin, G. (2008). Bayesian semiparametric cure rate model with an unknown threshold. *Scand. J. Statist.* **35**, 540–56.

Nolan, D. & Pollard, D. (1987). U-processes: Rates of convergence. *Ann. Statist.* **15**, 780–99.

Oakes, D. (1989). Bivariate survival models induced by frailties. *J. Am. Statist. Assoc.* **84**, 487–93.

Parner, E. (1998). Asymptotic theory for the correlated gamma-frailty model. *Ann. Statist.* **26**, 183–214.

Sherman, R. (1993). The limiting distribution of the maximum rank correlation estimator. *Econometrica* **61**, 123–37.

Sherman, R. (1994). Maximal inequalities for degenerate U-processes with applications to optimization estimators. *Ann. Statist.* **22**, 439–59.

Song, X. Ma, S. Huang, J. & Zhou, X. H. (2007). A semiparametric approach for the nonparametric transformation survival model with multiple covariates. *Biostatistics* **8**, 197–211.

Wienke, A. (2011). *Frailty Models in Survival Analysis*. Boca Raton, Florida: CRC Press.

Yin, G. (2008). Bayesian transformation cure frailty models with multivariate failure time data. *Statist. Med.* **27**, 5929–40.

Zeng, D. Chen, Q. & Ibrahim, J. G. (2009). Gamma frailty transformation models for multivariate survival times. *Biometrika* **96**, 277–91.

Zeng, D. & Lin, D. Y. (2007a). Maximum likelihood estimation in semiparametric regression models with censored data (with Discussion). *J. R. Statist. Soc.* B **69**, 507–64.

Zeng, D. & Lin, D. Y. (2007b). Semiparametric transformation models with random effects for recurrent events. *J. Am. Statist. Assoc.* **102**, 167–80.

Zeng, D. Lin, D. Y. & Yin, G. (2005). Maximum likelihood estimation for the proportional odds model with random effects. *J. Am. Statist. Assoc.* **100**, 470–83.

Zhang, D. & Davidian, M. (2001). Linear mixed models with flexible distributions of random effects for longitudinal data. *Biometrics* **57**, 795–802.

[*Received on* 4 *September* 2014. *Editorial decision on* 20 *August* 2017]