Ann Inst Stat Math (2016) 68:1095–1109
DOI 10.1007/s10463-015-0528-7

Jackknife empirical likelihood for linear transformation models with right censoring

Hanfang Yang[1] · **Shen Liu**[2] · **Yichuan Zhao**[3]

Received: 17 December 2013 / Revised: 15 March 2015 / Published online: 14 July 2015

Abstract A class of linear transformation models with censored data was proposed as a generalization of Cox models in survival analysis. This paper develops inference procedure for regression parameters based on jackknife empirical likelihood approach. We can show that the limiting variance is not necessary to estimate and the Wilk's theorem can be obtained. The jackknife empirical likelihood method benefits from the simpleness in optimization using jackknife pseudo-value. In our simulation studies, the proposed method is compared with the traditional empirical likelihood and normal approximation methods in terms of coverage probability and computational cost.

Keywords Linear transformation model · Empirical likelihood · Jackknife · Coverage probability

1 Introduction

When the association between the dependent and independent variables is not appropriately represented by conditional mean $m(x) = E(Y|X = x)$, transformations are recommended to apply on either dependent or independent variables or both. Subjectively, transformation models can afford many flexible choices for practical users; objectively, better statistical performance is achieved in choosing appropriate transformation functions, for example, Box–Cox transformation. As pioneers of trans-

✉ Yichuan Zhao
matyiz@langate.gsu.edu; yichuan@gsu.edu

1 School of Statistics, Renmin University of China, Beijing 100872, China

2 Guangxi Institute for Educational Research, Nanning 530021, Guangxi, China

3 Department of Mathematics and Statistics, Georgia State University, Atlanta, GA 30303, USA

formation models, Box and Cox (1964) introduced the power family transformation on the dependent variable. A large amount of literature discussed transformation models, such as Bickel and Doksum (1981), Carroll and Ruppert (1984), Breiman and Friedman (1985).

In survival analysis, covariate effects are considered to be associated with failure time, odds rate and hazard rate, etc. Calling for the better interpretation, researchers proposed the well-known proportional hazards model with right censoring (Cox 1972; Andesen and Gill 1982, etc.), proportional odds model (Bennett 1983), and accelerated failure time models (Pettitt 1982). Generally, those models with right censoring have linear covariate effect and can be obtained by taking transformations of survival function, i.e., $g\{S_Z(t)\} = h(t) + Z^T\beta$, where $S_Z(t)$ is the survival function of the failure time T conditioned on p-dimensional covariate Z, $h(\cdot)$ is a strictly increasing unspecified function and $g(\cdot)$ is a given decreasing function. Cheng et al. (1995) proposed a linear transformation model with censoring which generalizes most popular survival models, including models mentioned above, $h(T) = -Z^T\beta + \varepsilon$, where ε is a random variable independent of covariate Z with the distribution function $F(x) = 1 - g^{-1}(x)$. We can specify $g(x)$ to obtain the Cox regression model and the proportional odds model, respectively. Chen et al. (2002) and Kong et al. (2006) made great contributions to the linear transformation model with right censoring.

To estimate the asymptotic variance in the linear transformation model consistently is complicated and less computational efficient. In small samples, the performances of normal approximation approach often are limited by the accuracy of variance estimation [see Fine et al. (1998)]. Empirical likelihood (EL) method proposed by Owen (1988), Owen (1990) can tackle with this problem. Under mild conditions, the Wilk's theorem holds $-2\ln R(\theta_0) \xrightarrow{\mathfrak{D}} \chi^2_1$, where $R(\theta_0) = \sup_{\Phi_{\theta_0}} L(F)/\sup_{\Phi} L(F)$ and Φ_{θ_0}, a subset of empirical distribution space Φ, is controlled by the targeting parameter θ_0. Zhao (2010) proposed the EL inference method for linear transformation models with right censoring (see Cheng et al. 1995). Using EL approach, Yu et al. (2011) and Yang and Zhao (2012) adjusted the estimation equation of the transformation model to avoid estimating the covariance matrix. Moreover, the optimization problem in EL approach is a critical step to find the precise and reliable solution. Researchers in statistics field are hindered to improve EL method for U-statistics type estimating questions frequently encountered until Jing et al. (2009) introduced the jackknife pseudo- sample into the EL procedure. In this paper, we develop the jackknife empirical likelihood (JEL) method for linear transformation models.

The rest of the paper is organized as follows. In Sect. 2, we develop the jackknife empirical likelihood method for the linear transformation model. Then, we compare the JEL method and the existing methods in Sect. 3. In Sect. 4, we discuss the current method and some possible future work. The proofs are provided in the Appendix.

2 Inference procedure

Consider censored data (X_i, δ_i), $i = 1, \ldots, n$, where $X_i = \min(T_i, C_i)$ and $\delta_i = I(T_i \le C_i)$. C_i with distribution function $G(t)$ and failure time T_i are independent.

In this paper, we adopt the same notations as Yang and Zhao (2012) did. Covariate vector $\{Z_i\}_{i=1}^n \in \mathrm{R}^p$ is the corresponding vectors and $Z_{ij} = Z_i - Z_j$.

Fine et al. (1998) proposed the constrained variable t_0 satisfies $\Pr\{\min(T, C) > t_0\} > 0$, $\alpha = h(t_0)$ and $p+1$ dimensional parameter $\theta = (\alpha, \beta^T)^T$. Denote true $\theta_0 = (\alpha_0, \beta_0^T)^T$. The estimating equation of a transformation model is

$$\Omega_w(\theta) = \sum_{i=1}^{n} \sum_{j=1, i\neq j}^{n} w_{ij}(\theta)\dot{\eta}_{ij}(\theta) \left\{ \frac{\delta_j I\{\min(X_i, t_0) \geq X_j\}}{\hat{G}^2(X_j)} - \eta_{ij}(\theta) \right\}, \quad (1)$$

where $\hat{G}(\cdot)$ is the Kaplan–Meier estimator of G, the positive weighted function $w_{ij}(\cdot)$, $\eta_{ij}(\theta) = \eta(Z_{ij}^T\beta) - \Pr(T_i \geq T_j \geq t_0 | Z_i, Z_j)$ and its derivative vector $\dot{\eta}_{ij}(\theta)$ is defined in Yang and Zhao (2012).

Define $U_i = (Z_i^T, X_i, C_i)$. We have the same notations as Zhao (2010) and Yang and Zhao (2012) did

$$e_{ij}(\theta) = w_{ij}(\theta)\dot{\eta}_{ij}(\theta) \left\{ \frac{\delta_j I(\min(X_i, t_0) \geq X_j)}{G^2(X_j)} - \eta_{ij}(\theta) \right\},$$

and

$$d_i(\theta) = 2 \int_0^{t_0} \frac{q(\theta, t)}{\pi(t)} dM_i(t),$$

where

$$q(\theta, t) = \lim_{n\to\infty} \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} w_{ij}(\theta)\dot{\eta}_{ij}(\theta) \frac{\delta_j I\{\min(X_i, t_0) \geq X_j\}}{G^2(X_j)} I(X_j \geq t),$$

$$\pi(t) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} I(X_i \geq t),$$

and

$$dM_i(t) = I(X_i \leq t, \delta_i = 0) - \int_0^t I(X_i \geq u) d\Lambda_G(u).$$

The term $\Lambda_G(t)$ in the above equation is a cumulative hazard function. Denote $\hat{d}_i(\theta)$ the estimator of $d_i(\theta)$ when $q(\theta, t)$, $\pi(t)$ and $\Lambda_G(t)$ are estimated by finite sample. Denote $\hat{e}_{ij}(\theta)$ the estimator of $e_{ij}(\theta)$ when $G(t)$ is estimated by $\hat{G}(t)$ with finite sample. Furthermore, by Zhao (2010) and Yang and Zhao (2012), we have that

$$b(U_i, U_j; \theta) = \{e_{ij}(\theta) + d_i(\theta) + e_{ji}(\theta) + d_j(\theta)\},$$

$$V(\theta) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \left\{ b(U_i, U_j; \theta) \right\},$$

Zhao (2010) interpreted equation (1) as one U-statistics

$$\hat{V}(\theta) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \{\hat{b}(U_i, U_j; \theta)\},$$

with the symmetric kernel of 2 degrees $\hat{b}(U_i, U_j; \theta) = \{\hat{e}_{ij}(\theta) + \hat{d}_i(\theta) + \hat{e}_{ji}(\theta) + \hat{d}_j(\theta)\}$. In fact, $\hat{V}(\theta_0) = 2/\{n(n-1)\}\Omega_w(\theta_0)$ from Lemma A.1 of Yang and Zhao (2012). Yang and Zhao (2012) developed empirical likelihood method based on dependent vectors $\{\hat{W}_i(\theta), i = 1, \ldots, n\}$, where $\hat{W}_i(\theta) = 1/(n-1)\sum_{j=1, j\neq i}^{n}\{\hat{b}(U_i, U_j; \theta)\}$. Motivated by Jing et al. (2009), we construct the jackknife pseudo-sample for U-statistics $\hat{V}(\theta)$. The kernel $\{\hat{b}(U_i, U_j; \theta)\}$ is not fully determined by independent samples U_i and U_j, but incorporates with entire samples due to $\hat{d}_j(\theta)$ and $\hat{G}(\cdot)$ in $\hat{e}_{ji}(\theta)$. The arguments about jackknife empirical likelihood for U-statistics in Jing et al. (2009) could not guarantee that the Wilk's theorem is valid in this case. However, the consistency of $\hat{G}(\cdot)$ and the property that $\sum_{i=1}^{n} \hat{d}_i(\theta) = 0$ are essential for us to believe that $\hat{b}(U_i, U_j; \theta)$ approaches the kernel of U-statistics from U_i and U_j after using the jackknife procedure. Denote the pseudo-value $\hat{Q}_l(\theta) = n\hat{V}(\theta) - (n-1)\hat{V}_l(\theta)$, where

$$\hat{V}_l(\theta) = \frac{1}{(n-1)(n-2)} \sum_{i=1, i\neq l}^{n} \sum_{j=1, j\neq i,l}^{n} \left\{\hat{b}(U_i, U_j; \theta)\right\}, \quad l = 1, \ldots, n.$$

Hence, the jackknife empirical likelihood ratio at θ based on $\hat{Q}_l$ is given by,

$$R(\theta) = \sup\left\{\prod_{i=1}^{n} np_i : \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i \hat{Q}_i(\theta) = 0, p_i \geq 0\right\}.$$

Using Lagrange multiplier method, the jackknife empirical log-likelihood ratio is as follows,

$$l(\theta) = -2\log\{R(\theta)\} = 2\sum_{i=1}^{n} \log\{1 + \lambda(\theta)^T \hat{Q}_i(\theta)\}, \tag{2}$$

where $p+1$-dimensional Lagrange multiplier $\lambda(\theta)$ satisfies

$$g(\lambda(\theta)) \equiv \frac{1}{n}\sum_{i=1}^{n} \frac{\hat{Q}_i(\theta)}{1 + \lambda(\theta)^T \hat{Q}_i(\theta)} = 0. \tag{3}$$

Assume the standard regularity conditions, which are the same as those in Yang and Zhao (2012). We establish the Wilk's theorem as follows.

Theorem 1 *Under the regularity conditions in Yang and Zhao (2012), as $n \longrightarrow \infty$, we have*

$$l(\theta_0) \xrightarrow{\mathfrak{D}} \chi^2_{p+1}, \tag{4}$$

where χ^2_{p+1} is a standard Chi-squared random variable with $p+1$ degrees of freedom.

The $100(1-\alpha)$ % jackknife empirical likelihood confidence region for θ can be established as $R_\alpha = \{\theta : l(\theta) \leq \chi^2_{p+1}(\alpha)\}$, where $\chi^2_{p+1}(\alpha)$ is the upper α-quantile of distribution of χ^2_{p+1}.

When one is interested in θ_{10}, a sub-vector of $\theta_0 = (\theta_{10}^T, \theta_{20}^T)^T$, the hypothesis is shown as $H_0 : \theta_1 = \theta_{10}$, where $\theta_1 \in R^q$ and $\theta_2 \in R^{p+1-q}$. The Wilk's theorem for the profiled jackknife empirical likelihood ratio $l^*(\theta_1) = \inf_{\theta_2} l(\theta_1, \theta_2)$ (see Qin and Lawless 1994) is given as follows.

Theorem 2 *Under the regularity conditions in Yang and Zhao (2012), as $n \longrightarrow \infty$,*

$$l^*(\theta_{10}) \xrightarrow{\mathfrak{D}} \chi^2_q. \tag{5}$$

Thus, we can construct the jackknife empirical likelihood confidence region for θ_{10} with $100(1-\alpha)$ % level $R^*_\alpha = \left\{\theta_1 : l^*(\theta_1) \leq \chi^2_{\mathrm{q}}(\alpha)\right\}$, where $\chi^2_q(\alpha)$ is the upper α-quantile of distribution of χ^2_q. Unlike the existing methods, the unscaled limiting distribution ensures that inference is driven by data automatically, without estimating covariance matrix.

3 Numerical studies

We carry out simulation studies in terms of coverage probability of confidence regions, comparing jackknife empirical likelihood with the normal approximation and EL methods. Based on (1), Fine et al. (1998) constructed normal approximation (NA)-based confidence regions, which need to estimate a complicated covariance matrix for the inference. The scenarios of simulation are considered as those in Yang and Zhao (2012). We choose the logarithm function as a link function. The ϵ is a standard extreme value function and the transformation model becomes the Cox regression model. We denote $w(\cdot) = 1$. Note that t_0 is corresponding to 20% upper quantile of the censoring data. The censoring data are from $U[0, c]$. The censoring rates are selected as 0.1, 0.2, 0.3 and 0.4, respectively. As Yang and Zhao (2012) did for the EL method, we have setting (**A**), $\beta = (-0.5, 0.5)$. Z_1 follows $U[0, 1]$, and Z_2 is from Bernoulli distribution with success probability 0.2. In setting (**B**), $\beta_1 = 1$ and $\beta_2 = 0$. Z_1 is $U[0, 1]$, and Z_2 is from Bernoulli distribution with 0.2. The sample sizes are 60 and 100, and the data are repeated 1000 times.

In Table 1, we report coverage probabilities of 95 % JEL, EL and normal approximation (NA) confidence regions. For the sample size 60, the NA, JEL and EL methods have good performances. When the sample size increases to 100, the estimated coverage probabilities for NA, EL and JEL methods are close to 95 % nominal level. The

Table 1 Coverage probability of 95 % confidence region for β

n	Censoring rate	JEL (A)	EL (A)	NA (A)	JEL (B)	EL (B)	NA (B)
60	0.1	0.928	0.933	0.929	0.920	0.929	0.936
60	0.2	0.943	0.944	0.938	0.949	0.950	0.932
60	0.3	0.949	0.951	0.913	0.961	0.965	0.931
60	0.4	0.927	0.952	0.912	0.934	0.956	0.906
100	0.1	0.960	0.949	0.941	0.954	0.951	0.950
100	0.2	0.958	0.953	0.929	0.949	0.956	0.948
100	0.3	0.957	0.945	0.933	0.956	0.968	0.952
100	0.4	0.931	0.940	0.915	0.936	0.954	0.935

Table 2 Average time (second) per repetition for computing 95 % coverage probabilities for β

n	Censoring rate	JEL (A)	EL (A)	NA (A)	JEL (B)	EL (B)	NA (B)
60	0.1	0.354	0.353	0.998	0.355	0.351	1.055
60	0.2	0.354	0.352	1.023	0.354	0.351	1.052
60	0.3	0.355	0.353	1.057	0.355	0.352	1.051
60	0.4	0.356	0.354	1.055	0.355	0.352	1.055
100	0.1	0.974	0.961	3.143	0.992	0.977	3.007
100	0.2	0.987	0.973	3.135	1.008	0.993	3.013
100	0.3	0.992	0.978	3.136	1.022	1.008	2.927
100	0.4	0.993	0.980	3.146	1.023	1.009	2.944

average computation times per repetition (with the second as the unit) of the three methods for calculating coverage probability are shown in Table 2 with 1000 repetitions. Even though the EL method spends slightly shorter time than JEL does, we would conclude that JEL and EL methods take similar running time because of very small difference of computational burden. When calculating the estimation equations, the JEL is comparable with EL method in speed because jackknifing technique does not leverage computational complexity. The jackknife procedure in a linear transformation model can be simplified because the estimating equations are calculated by the summation. Moreover, both finding β and the variance estimation for the normal approximation method need more computational cost than JEL and EL methods.

4 Discussion

In this paper, we apply jackknife empirical likelihood to the U-statistics type estimation equation of linear transformation model. The Wilk's theorem for any fixed component has been established, where $\hat{b}(U_i, U_j; \theta)$ is close to the kernel function of U-statistics. Thus, the JEL method can be applied to more general case than U-statistics. Recently, the linear transformation model with the diverging number of dimensions p was extensively discussed. Penalized empirical likelihood methods for large p were investigated by Tang and Leng (2010) and Lahiri and Mukhopadhyay (2012), among others. In the future, we will develop the jackknife EL procedure for linear transformation models to solve the high-dimensional optimization issue and computational cost problem.

Proofs of Theorems

Denote $\Gamma(\theta_0) = \Gamma_1(\theta_0) - \Gamma_2(\theta_0)$, which is a limiting covariance matrix for $n^{(-3/2)}\Omega_w(\theta_0)$ defined by Fine et al. (1998), where

$$\Gamma_1(\theta) = \lim_{n\to\infty} \frac{1}{n^3} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \sum_{k=1, k\neq j}^{n} \{e_{ij}(\theta) + e_{ji}(\theta)\}\{e_{ik}(\theta) + e_{ki}(\theta)\}^T,$$

$$\Gamma_2(\theta) = 4\int_0^{t_0} \frac{q(\theta,t)q^T(\theta,t)}{\pi(t)}\, d\Lambda_G(t).$$

Lemma 1 *Under the conditions of Theorem 1,*

$$\sqrt{n}\frac{1}{n}\sum_{l=1}^{n} \hat{Q}_l(\theta_0) = \sqrt{n}\hat{V}(\theta_0) \xrightarrow{\mathfrak{D}} N(0, 4\Gamma(\theta_0)),$$

as $n \to \infty$.

Proof Noticing that $\hat{G}(\cdot)$ and $\hat{\Lambda}_G(t)$ in $\hat{b}(U_i, U_j; \theta_0)$ are estimated by the full sample, we re-express $\sum_{l=1}^{n} \hat{Q}_l$ as follows,

$$\begin{aligned}
\sum_{l=1}^{n} \hat{Q}_l(\theta_0) &= \sum_{l=1}^{n} n\hat{V}(\theta_0) - \sum_{l=1}^{n}(n-1)\hat{V}_l(\theta_0) \\
&= n^2\hat{V}(\theta_0) - \frac{1}{n-2}\sum_{l=1}^{n}\sum_{i=1, i\neq l}^{n}\sum_{j=1, j\neq i,l}^{n} \left\{\hat{b}(U_i, U_j; \theta_0)\right\} \\
&= n^2\hat{V}(\theta_0) - \frac{1}{n-2}\sum_{\{i,j,l|1\leq i,j,l\leq n, i\neq j, j\neq l, l\neq i\}}^{n} \left\{\hat{b}(U_i, U_j; \theta_0)\right\} \\
&= n^2\hat{V}(\theta_0) - \frac{n-2}{n-2}\sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} \left\{\hat{b}(U_i, U_j; \theta_0)\right\} \\
&= n^2\hat{V}(\theta_0) - n(n-1)\cdot\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} \left\{\hat{b}(U_i, U_j; \theta_0)\right\} \\
&= n^2\hat{V}(\theta_0) - n(n-1)\hat{V}(\theta_0) \\
&= n\hat{V}(\theta_0).
\end{aligned} \tag{6}$$

Thus, the following result holds by Lemma A.1 of Yang and Zhao (2012),

$$\sqrt{n}\frac{1}{n}\sum_{l=1}^{n}\hat{Q}_l(\theta_0) = \sqrt{n}\hat{V}(\theta_0) \xrightarrow{\mathfrak{D}} N(0, 4\Gamma(\theta_0)). \tag{7}$$

□

Lemma 2 *Denote*

$$V_l(\theta) = \frac{1}{(n-1)(n-2)}\sum_{i=1,i\neq l}^{n}\sum_{j=1,j\neq i,l}^{n}\left\{b(U_i, U_j;\theta)\right\},$$

and

$$Q_l(\theta) = nV(\theta) - (n-1)V_l(\theta).$$

Under the conditions of Theorem 1, let

$$\Pi_n(\theta_0) = \frac{1}{n}\sum_{l=1}^{n} Q_l(\theta_0)Q_l^T(\theta_0),$$

and

$$\hat{\Pi}_n(\theta_0) = \frac{1}{n}\sum_{l=1}^{n}\hat{Q}_l(\theta_0)\hat{Q}_l^T(\theta_0).$$

Then, we have (a) $\Pi_n(\theta_0) \xrightarrow{P} 4\Gamma(\theta_0)$*; ii)* $\hat{\Pi}_n(\theta_0) \xrightarrow{P} 4\Gamma(\theta_0)$.

Proof This lemma is to prove that variance of $\hat{Q}_l(\theta_0)$, denoted as $\hat{\Pi}_n(\theta_0)$, whose components are estimated by random samples, converges to $4\Gamma(\theta_0)$ in probability. It is not a strictly full version of traditional jackknife sample in $\hat{Q}_l(\theta_0)$, which means that we cannot apply the properties of jackknife EL to our method directly. Note that $b(U_i, U_j;\theta_0)$ depends on both i-th and the j-th samples, since the terms $q(\theta, t)$, $\pi(t)$, $G(\cdot)$, $\Lambda_G(t)$ are determined, instead of random. Therefore, what we are required to do is to: (a) prove that $\Pi_n(\theta_0)$ converges to $4\Gamma(\theta_0)$ in probability; (b) verify that the gap between $\Pi_n(\theta_0)$ and $\hat{\Pi}_n(\theta_0)$ is close to zero in probability. The details are given as follows.

For (a), similar to Lemma A.3 of Jing et al. (2009), one has

$$\begin{aligned}\Pi_n(\theta_0) &= \frac{1}{n}\sum_{l=1}^{n} Q_l(\theta_0)Q_l^T(\theta_0)\\ &= \frac{1}{n}\sum_{l=1}^{n}\Big\{Q_l(\theta_0)Q_l^T(\theta_0) - 2Q_l(\theta_0)V^T(\theta_0) + V(\theta_0)V^T(\theta_0)\Big\} + V(\theta_0)V^T(\theta_0)\\ &= \frac{1}{n}\sum_{l=1}^{n}(Q_l(\theta_0) - V(\theta_0))(Q_l(\theta_0) - V(\theta_0))^T + V(\theta_0)V(\theta_0)^T.\end{aligned} \tag{8}$$

Like equation (A.4) in Yang and Zhao (2012), it is clear that

$$\mathrm{Var}(V(\theta_0)) = \frac{4\Gamma(\theta_0)}{n} + o\left(n^{-1}\right), \text{ a.s.} \tag{9}$$

Denote the jackknife estimator of $\mathrm{Var}(V(\theta_0))$ by $\hat{\mathrm{Var}}(jack) \equiv \frac{1}{n(n-1)}\sum_{l=1}^{n}(Q_l(\theta_0) - V(\theta_0))(Q_l(\theta_0) - V(\theta_0))^T$, and it is consistent with $\mathrm{Var}(V(\theta_0))$ from Lee (1990), i.e.,

$$n\left[\hat{\mathrm{Var}}(jack) - \mathrm{Var}(V(\theta_0))\right] \to 0, \text{ a.s.} \tag{10}$$

Thus, combining (9) and (10), the first part of (8) equals to

$$\begin{aligned}\frac{1}{n}\sum_{l=1}^{n}(Q_l(\theta_0) - V(\theta_0))(Q_l(\theta_0) - V(\theta_0))^T &= (n-1)\hat{\mathrm{Var}}(jack)\\ &= 4\Gamma(\theta_0) + o(1).\end{aligned}$$

Note that $1/n\sum_{i=1}^{n} Q_l(\theta_0) = V(\theta_0)$, which is similar to (6). Combining (9) and the Law of Large Numbers for U-statistics, we obtain $V(\theta_0) = O(n^{-1/2})$. Therefore, (8) equals to

$$\Pi_n(\theta_0) = 4\Gamma(\theta_0) + o(1), \text{ a.s.},$$

i.e.,

$$\Pi_n(\theta_0) \xrightarrow{P} 4\Gamma(\theta_0). \tag{11}$$

For (b), to prove the difference between $\Pi_n(\theta_0)$ and $\hat{\Pi}_n(\theta_0)$ is close to zero in probability, we need to check that $|\hat{Q}_l(\theta_0) - Q_l(\theta_0)| = o_p(1)$ first. Like Yang and Zhao (2012), it is clear that

$$\sup_{i,j=1,\ldots,n}\left|\hat{b}(U_i, U_j; \theta_0) - b(U_i, U_j; \theta_0)\right| = o_p(1).$$

Then, according to the definition of $Q_l(\theta_0)$, it can be re-expressed as follows.

$$\begin{aligned}Q_l(\theta_0) &= nV(\theta_0) - (n-1)V_l(\theta_0)\\ &= \frac{n}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\left\{b(U_i, U_j; \theta_0)\right\}\\ &\quad - \frac{n-1}{(n-1)(n-2)}\sum_{i=1,i\neq l}^{n}\sum_{j=1,j\neq i,l}^{n}\left\{b(U_i, U_j; \theta_0)\right\}\\ &= \frac{1}{n-1}\sum_{j=1,j\neq l}^{n}\left\{b(U_l, U_j; \theta_0)\right\} + \frac{1}{n-1}\sum_{i=1,i\neq l}^{n}\left\{b(U_i, U_l; \theta_0)\right\}\end{aligned}$$

$$-\frac{1}{(n-1)(n-2)}\sum_{i=1,i\neq l}^{n}\sum_{j=1,j\neq i,l}^{n}\left\{b(U_i,U_j;\theta_0)\right\}$$
$$=2W_l(\theta_0)-V_l(\theta_0),$$

as $b(U_i, U_j; \theta_0)$ is symmetric for any i and j. Similarly,

$$\begin{aligned}\hat{Q}_l(\theta_0) &= n\hat{V}(\theta_0)-(n-1)\hat{V}_l(\theta_0)\\ &= \frac{n}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\left\{\hat{b}(U_i,U_j;\theta_0)\right\}\\ &\quad -\frac{n-1}{(n-1)(n-2)}\sum_{i=1,i\neq l}^{n}\sum_{j=1,j\neq i,l}^{n}\left\{\hat{b}(U_i,U_j;\theta_0)\right\}\\ &= \frac{1}{n-1}\sum_{j=1,j\neq l}^{n}\left\{\hat{b}(U_l,U_j;\theta_0)\right\}+\frac{1}{n-1}\sum_{i=1,i\neq l}^{n}\left\{\hat{b}(U_i,U_l;\theta_0)\right\}\\ &\quad -\frac{1}{(n-1)(n-2)}\sum_{i=1,i\neq l}^{n}\sum_{j=1,j\neq i,l}^{n}\left\{\hat{b}(U_i,U_j;\theta_0)\right\}\\ &= 2\hat{W}_l(\theta_0)-\hat{V}_l(\theta_0).\end{aligned}$$

Therefore, we have

$$\left|\hat{Q}_l(\theta_0)-Q_l(\theta_0)\right| = \left|2\left[\hat{W}_l(\theta_0)-W_l(\theta_0)\right]+\left[V_l(\theta_0)-\hat{V}_l(\theta_0)\right]\right|. \tag{12}$$

For the first part of (12), by Yang and Zhao (2012), we obtain that

$$\begin{aligned}&2|\hat{W}_l(\theta_0)-W_l(\theta_0)|\\ &\quad=\frac{2}{n-1}\left|\sum_{j=1,j\neq l}^{n}\{b(U_l,U_j;\theta_0)-\hat{b}(U_l,U_j;\theta_0)\}\right|\\ &\quad\leq\frac{2}{n-1}\sum_{j=1,j\neq l}^{n}\sup_{l,j=1,\ldots,n}\left|b(U_l,U_j;\theta_0)-\hat{b}(U_l,U_j;\theta_0)\right|\\ &\quad=o_p(1).\end{aligned} \tag{13}$$

Similarly, for the second part of (12), one has that

$$\begin{aligned}&|V_l(\theta_0)-\hat{V}_l(\theta_0)|\\ &\quad=\frac{1}{(n-1)(n-2)}\left|\sum_{i=1,i\neq l}^{n}\sum_{j=1,j\neq i,l}^{n}\{b(U_i,U_j;\theta_0)-\hat{b}(U_i,U_j;\theta_0)\}\right|\end{aligned}$$

$$\leq \frac{1}{(n-1)(n-2)} \sum_{i=1, i\neq l}^{n} \sum_{j=1, j\neq i,l}^{n} \sup_{i,j=1,\ldots,n} \left| b(U_i, U_j; \theta_0) - \hat{b}(U_i, U_j; \theta_0) \right|$$
$$= o_p(1). \tag{14}$$

Combining (12), (13) and (14), it leads to

$$|\hat{Q}_l(\theta_0) - Q_l(\theta_0)| = o_p(1). \tag{15}$$

For any $\alpha \in R^{p+1}$, by (15)

$$\begin{aligned}
&\alpha^T \left\{ \Pi_n(\theta_0) - \hat{\Pi}_n(\theta_0) \right\} \alpha \\
&= \frac{1}{n}\alpha^T \left\{ \sum_{l=1}^{n} Q_l(\theta_0) Q_l^T(\theta_0) - \sum_{l=1}^{n} \hat{Q}_l(\theta_0) \hat{Q}_l^T(\theta_0) \right\} \alpha \\
&= \frac{1}{n}\alpha^T \left\{ \sum_{l=1}^{n} Q_l(\theta_0) Q_l^T(\theta_0) - 2\sum_{l=1}^{n} \hat{Q}_l(\theta_0) Q_l^T(\theta_0) + \sum_{l=1}^{n} \hat{Q}_l(\theta_0) \hat{Q}_l^T(\theta_0) \right. \\
&\quad \left. +2\sum_{l=1}^{n} \hat{Q}_l(\theta_0) Q_l^T(\theta_0) - 2\sum_{l=1}^{n} \hat{Q}_l(\theta_0) \hat{Q}_l^T(\theta_0) \right\} \alpha \\
&= \frac{1}{n}\sum_{l=1}^{n} \left[\alpha^T (Q_l(\theta_0) - \hat{Q}_l(\theta_0)) \right]^2 + \frac{2}{n}\sum_{l=1}^{n} \alpha^T \hat{Q}_l(\theta_0)(Q_l(\theta_0) - \hat{Q}_l(\theta_0))^T \alpha \\
&= o_p(1),
\end{aligned}$$

as both $|Q_l(\theta_0)|$ and $|\hat{Q}_l(\theta_0)|$ are uniformly bounded, due to the boundness of $|b(U_i, U_j; \theta_0)|$ and (15). Thus,

$$\hat{\Pi}_n(\theta_0) = \hat{\Pi}_n(\theta_0) - \Pi_n(\theta_0) + \Pi_n(\theta_0) \xrightarrow{P} 4\Gamma(\theta_0), \tag{16}$$

and we finish the proof. □

Lemma 3 *Under the conditions of Theorem 1,* $\|\lambda(\theta_0)\| = O_p(n^{-1/2})$, *where* $\|\cdot\|$ *denotes the Euclidean norm.*

Proof As in (Owen (1990), p. 101), we let $\lambda(\theta_0) = \rho\nu$ where $\rho \geq 0$ and $\|\nu\| = 1$, and have

$$\begin{aligned}
0 &= \|g(\rho\nu)\| \\
&\geq |\nu^T g(\rho\nu)| \\
&= \frac{1}{n}\left| \nu^T \left[\sum_{l=1}^{n} \frac{\hat{Q}_l(\theta_0)}{1+\rho\nu^T \hat{Q}_l(\theta_0)} \right] \right| = \frac{1}{n}\left| \nu^T \left[\sum_{l=1}^{n} \hat{Q}_l(\theta_0) - \rho \sum_{l=1}^{n} \frac{\hat{Q}_l(\theta_0)\hat{Q}_l^T(\theta_0)\nu}{1+\rho\nu^T \hat{Q}_l(\theta_0)} \right] \right| \\
&\geq \frac{\rho}{n}\left| \nu^T \sum_{l=1}^{n} \frac{\hat{Q}_l(\theta_0)\hat{Q}_l^T(\theta_0)}{1+\rho\nu^T \hat{Q}_l(\theta_0)}\nu \right| - \frac{1}{n}\left| \nu^T \sum_{l=1}^{n} \hat{Q}_l(\theta_0) \right|,
\end{aligned}$$

where $g(\cdot)$ has been defined by (3). It holds that

$$\frac{\rho}{n}\left|\nu^T\sum_{l=1}^{n}\frac{\hat{Q}_l(\theta_0)\hat{Q}_l^T(\theta_0)}{1+\rho\nu^T\hat{Q}_l(\theta_0)}\nu\right| \le \frac{1}{n}\left|\nu^T\sum_{l=1}^{n}\hat{Q}_l(\theta_0)\right|.$$

It is clear that $n^{-1}\sum_{l=1}^{n}\hat{Q}_l(\theta_0) = \hat{V}(\theta_0) = O_p(n^{-1/2})$ by (7), and $\nu^T\hat{\Pi}_n(\theta_0)\nu = \nu^T(\hat{\Pi}_n(\theta_0)-\Gamma(\theta_0))\nu + \nu^T\Gamma(\theta_0)\nu = O_p(1)$ by Lemma 2. As shown in Lemma 2, $\|\hat{Q}_l(\theta_0)\|$ is uniformly bounded by M for all i, thus $1+\rho\nu^T\hat{Q}_l(\theta_0) \le 1+\rho M$. We can obtain that

$$\frac{\rho\nu\hat{\Pi}_n\nu}{1+\rho M} \le \|\nu\|\cdot\left\|\frac{1}{n}\sum_{l=1}^{n}\hat{Q}_l(\theta_0)\right\| = O_p(n^{-1/2}).$$

Hence,

$$\|\lambda(\theta_0)\| = \rho = O_p(n^{-1/2}). \tag{17}$$

□

Proof of Theorem 1 We will derive an asymptotic expression for $\lambda(\theta)$ as the root of $g(\lambda(\theta))$ when θ is replaced by its true value θ_0 as follows

$$\begin{aligned}0 = g(\lambda(\theta_0)) &= \frac{1}{n}\sum_{l=1}^{n}\frac{\hat{Q}_l(\theta_0)}{1+\lambda(\theta_0)^T\hat{Q}_l(\theta_0)}\\ &= \frac{1}{n}\sum_{l=1}^{n}\hat{Q}_l(\theta_0)\left[1-\lambda(\theta_0)^T\hat{Q}_l(\theta_0)+\frac{\lambda(\theta_0)^T\hat{Q}_l(\theta_0)\hat{Q}_l^T(\theta_0)\lambda(\theta_0)}{1+\lambda(\theta_0)^T\hat{Q}_l(\theta_0)}\right]\\ &= \hat{V}(\theta_0)-\lambda(\theta_0)^T\hat{\Pi}_n(\theta_0)+\frac{1}{n}\sum_{l=1}^{n}\frac{\hat{Q}_l(\theta_0)\lambda(\theta_0)^T\hat{Q}_l(\theta_0)\hat{Q}_l^T(\theta_0)\lambda(\theta_0)}{1+\lambda(\theta_0)^T\hat{Q}_l(\theta_0)}.\end{aligned}$$

Since $\|\hat{Q}_l(\theta_0)\|$ is uniformly bounded by M, and $\|\hat{\Pi}_n(\theta_0)\|$ is $O_p(1)$. From Lemma 3, $1+\lambda(\theta_0)^T\hat{Q}_l(\theta_0) = O_p(1)$. Thus, the last term of the above equation can be written as

$$\left|\frac{1}{n}\sum_{l=1}^{n}\frac{\hat{Q}_l(\theta_0)\lambda(\theta_0)^T\hat{Q}_l(\theta_0)\hat{Q}_l^T(\theta_0)\lambda(\theta_0)}{1+\lambda(\theta_0)^T\hat{Q}_l(\theta_0)}\right| = O_p(n^{-1}).$$

Hence, we write

$$\lambda(\theta_0) = \hat{\Pi}_n^{-1}(\theta_0)\hat{V}(\theta_0) + O_p(n^{-1}). \tag{18}$$

By (18), $l(\theta_0)$ can be expanded as

$$
\begin{aligned}
l(\theta_0) &= 2\sum_{l=1}^{n}\log\{1+\lambda(\theta_0)^T\hat{Q}_l(\theta_0)\} \\
&= 2\sum_{l=1}^{n}\left[\lambda(\theta_0)^T\hat{Q}_l(\theta_0) - \frac{1}{2}\lambda(\theta_0)^T\hat{Q}_l(\theta_0)\hat{Q}_l^T(\theta_0)\lambda(\theta_0) + O((\lambda(\theta_0)^T\hat{Q}_l(\theta_0))^3)\right] \\
&= 2n\hat{V}^T(\theta_0)\hat{\Pi}_n^{-1}(\theta_0)\hat{V}(\theta_0) - n\hat{V}^T(\theta_0)\hat{\Pi}_n^{-1}(\theta_0)\hat{\Pi}_n(\theta_0)\hat{\Pi}_n^{-1}(\theta_0)\hat{V}(\theta_0) \\
&\quad + O_p(n^{-1/2}) \\
&= \sqrt{n}\hat{V}^T(\theta_0)\cdot\hat{\Pi}_n^{-1}(\theta_0)\cdot\sqrt{n}\hat{V}(\theta_0) + o_p(1),
\end{aligned}
$$

since $\|\hat{Q}_l(\theta_0)\|$ is uniformly bounded, $\hat{V}(\theta_0) = O_p(n^{-1/2})$, and $\|\lambda(\theta_0)\| = O_p(n^{-1/2})$. Thus, combining the results of Lemmas 1 and 2, we obtain that

$$
l(\theta_0) \xrightarrow{\mathfrak{D}} \chi^2_{p+1}. \tag{19}
$$

□

Proof of Theorem 2 The proof is along the lines of Zhang and Zhao (2013). Let $\tilde{\theta}_2 = \arg\inf_{\theta_2} l(\theta_{10}, \theta_2)$, $\tilde{D}(\theta_0) = \lim_{n\to\infty} n^{-2}\sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} w_{ij}(\theta_0)\tilde{\eta}_{ij}(\theta_0)\tilde{\eta}_{ij}^T(\theta_0)$, where $\tilde{\eta}_{ij}^T(\theta_0)$ is the partial derivative of $\eta_{ij}^T(\theta_0)$ with respect to θ_2, and $\tilde{\Phi}(\theta_0) = \tilde{D}^T(\theta_0)\Pi^{-1}(\theta_0)\tilde{D}(\theta_0)$. Then, following the similar arguments in Qin and Lawless (1994) and Fine et al. (1998), we have

$$
\sqrt{n}(\tilde{\theta}_2 - \theta_{20}) = -\tilde{\Phi}(\theta_0)^{-1}\tilde{D}(\theta_0)^T\Pi(\theta_0)^{-1}\frac{1}{\sqrt{n}}\sum_{l=1}^{n}\hat{Q}_l(\theta_0) + o_p(1),
$$

and the Lagrange multiplier λ_2 satisfying

$$
\sqrt{n}\lambda_2 = S_1\frac{1}{\sqrt{n}}\sum_{l=1}^{n}\hat{Q}_l(\theta_0) + o_p(1),
$$

where $S_1 = \Pi^{-1}(\theta_0) - \Pi^{-1}(\theta_0)\tilde{D}(\theta_0)\tilde{\Phi}^{-1}(\theta_0)\tilde{D}(\theta_0)^T\Pi^{-1}(\theta_0)$. Although $\hat{Q}_l$'s are not independent of each other, which is different from situation in Qin and Lawless (1994), Lemmas 1 and 2 guarantee the convergence rate of $\sum_{l=1}^{n}\hat{Q}_l(\theta_0)$ and boundness of $\hat{Q}_l(\theta_0)$. Then, similar to Zhang and Zhao (2013) and Theorem 1, one has that

$$
l^*(\theta_{10}) = \left\{\frac{1}{\sqrt{n}}\sum_{l=1}^{n}\hat{Q}_l(\theta_0)\right\}^T S_1\left\{\frac{1}{\sqrt{n}}\sum_{l=1}^{n}\hat{Q}_l(\theta_0)\right\} + o_p(1)
$$

$$= \left\{ \Pi^{-1/2}(\theta_0) \frac{1}{\sqrt{n}} \sum_{l=1}^{n} \hat{Q}_l(\theta_0) \right\}^T \left\{ I - \Pi^{-1/2}(\theta_0) \tilde{D}(\theta_0) \tilde{\Phi}^{-1}(\theta_0) \tilde{D}(\theta_0)^T \Pi^{-1/2}(\theta_0) \right\}$$

$$\times \left\{ \Pi^{-1/2}(\theta_0) \frac{1}{\sqrt{n}} \sum_{l=1}^{n} \hat{Q}_l(\theta_0) \right\} + o_p(1)$$

$$\equiv \left\{ \Pi^{-1/2}(\theta_0) \frac{1}{\sqrt{n}} \sum_{l=1}^{n} \hat{Q}_l(\theta_0) \right\}^T S_2 \left\{ \Pi^{-1/2}(\theta_0) \frac{1}{\sqrt{n}} \sum_{l=1}^{n} \hat{Q}_l(\theta_0) \right\} + o_p(1).$$

Since $S_2 = I - \Pi^{-1/2}(\theta_0) \tilde{D}(\theta_0) \tilde{\Phi}^{-1}(\theta_0) \tilde{D}(\theta_0)^T \Pi^{-1/2}(\theta_0)$ is a symmetric and idempotent matric with trace q, we have

$$l^*(\theta_{10}) \xrightarrow{\mathfrak{D}} \chi_q^2.$$

□

Acknowledgments The authors would like to thank the two referees and AE for their useful suggestions which lead to a significant improvement. Hanfang Yang's research is supported by the Fundamental Research Funds for the Central Universities, and the Research Funds of Renmin University of China (13XNF059). Yichuan Zhao was partially supported by the NSF grant DMS-1406163, NSA grant H98230-12-1-0209, and the RIG grant, Georgia State University. The authors acknowledge the help in language by Jenny Zhao.

References

Aalen, O., Borgan, O., Gjessing, H. (2008). *Survival and event history analysis: a process point of view*. New York: Springer.

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

Bennett, S. (1983). Analysis of survival data by the proportional odds model. *Statistics in Medicine*, *2*, 273–277.

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

Box, G. E. P., Cox, D. R. (1964). An analysis of transformations. *Journal of The Royal Statistical Society Series*, *26*, 211–252.

Breiman, L., Friedman, J. H. (1985). Estimating optimal transformations for multiple regression and correlation. *Journal of the American Statistical Association*, *80*, 580–619.

Cai, J., Fan, J., Li, R., Zhou, H. (2005). Variable selection for multivariate failure time data. *Biometrika*, *92*, 303–316.

Cai, T., Wei, L. J., Wilcox, M. (2000). Semiparametric regression analysis for clustered failure time data. *Biometrika*, *87*, 867–878.

Carroll, R. J., Ruppert, D. (1984). Power transformation when fitting theoretical models to data. *Journal of the American Statistical Association*, *79*, 321–328.

Chen, K., Jin, Z., Ying, Z. (2002). Semiparametric analysis of transformation models with censored data. *Biometrika*, *89*, 659–668.

Chen, S., Peng, L., Qin, Y. (2009). Effects of data dimension on empirical likelihood. *Biometrika*, *96*, 711–722.

Cheng, S. C., Wei, L. J., Ying, Z. (1995). Analysis of transformation models with censored data. *Biometrika*, *82*, 835–845.

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

Fine, J. P., Ying, Z., Wei, L. J. (1998). On the linear transformation model with censored data. *Biometrika*, *85*, 980–986.

Jing, B. Y., Yuan, J. Q., Zhou, W. (2009). Jackknife empirical likelihood. *Journal of the American Statistical Association*, *104*, 1224–1232.

Kong, L., Cai, J., Sen, P. (2006). Asymptotic results for fitting semiparametric transformation models to failure time data from case-cohort studies. *Statistica Sinica*, *16*, 135–151.

Lahiri, S., Mukhopadhyay, S. (2012). A penalized empirical likelihood method in high dimensions. *Annals of Statistics*, *40*, 2511–2540.

Lee, A. J. (1990). *U-Statistics*. Theory and Practice. New York: CRC Press.

Owen, A. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, *75*, 237–249.

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

Owen, A. (2001). *Empirical Likelihood*. Florida: Chapman and Hall/CRC.

Pettitt, A. N. (1982). Inference for the linear model using a likelihood based on ranks. *Journal of the Royal Statistical Society: Series B*, *44*, 234–243.

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

Tang, C. Y., Leng, C. (2010). Penalized high-dimensional empirical likelihood. *Biometrika*, *4*, 905–920.

Yu, W., Sun, Y., Zheng, M. (2011). Empirical likelihood method for linear transformation models. *Annals of the Institute of Statistical Mathematics*, *63*, 331–346.

Yang, H., Zhao, Y. (2012). New empirical likelihood inference for linear transformation models. *Journal of Statistical Planning and Inference*, *142*, 1659–1668.

Zhang, Z., Zhao, Y. (2013). Empirical likelihood for linear transformation models with interval-censored failure time data. *Journal of Multivariate Analysis*, *116*, 398–409.

Zhao, Y. (2010). Semiparametric inference for transformation models via empirical likelihood. *Journal of Multivariate Analysis*, *101*, 1846–1858.