Empirical likelihood for linear transformation models with interval-censored failure time data

Zhigang Zhang\textsuperscript{a}, Yichuan Zhao\textsuperscript{b,}\textsuperscript{*}

\textsuperscript{a} Department of Epidemiology and Biostatistics, Memorial Sloan–Kettering Cancer Center, New York, NY 10065, United States
\textsuperscript{b} Department of Mathematics and Statistics, Georgia State University, Atlanta, GA 30303, United States

\textbf{A R T I C L E  I N F O}

\textbf{Article history:}
Received 27 July 2012
Available online 23 January 2013

\textbf{AMS subject classifications:}
62G05
62G10
62F12

\textbf{Keywords:}
Confidence intervals/regions
Coverage probability
Interval-censored failure time data
Jackknife empirical likelihood
Linear transformation models
Estimating equations

\textbf{A B S T R A C T}

For regression analysis of interval-censored failure time data, Zhang et al. (2005) [40] proposed an estimating equation approach to fit linear transformation models. In this paper, we develop two empirical likelihood (EL) inference approaches for the regression parameters based on the generalized estimating equations. The limiting distributions of log-empirical likelihood ratios are derived and empirical likelihood confidence intervals for any specified component of regression parameters are obtained. We carry out extensive simulation studies to compare the proposed methods with the method discussed by Zhang et al. (2005) [40]. The simulation results demonstrate that the EL and jackknife EL methods for linear transformation models have better performance than the existing normal approximation method based on coverage probability of confidence intervals in most cases, and they enable us to overcome an under-coverage problem for the confidence intervals of the regression parameters using a normal approximation when sample sizes are small and right censoring is heavy. Two real data examples are provided to illustrate our procedures.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

Let T denote the time of occurrence of an event of interest. We say that T is interval-censored if its true value is not observed but only known to lie in an interval, say, (L, R). “Such data are frequently observed in clinical trials or longitudinal studies that entail periodic follow-ups” (see [40]). For example, after radiotherapy, a cancer patient is often required to visit the physician regularly to examine whether or not there is disease progression. If progression was not observed at, say, the 1-year follow-up but was at the 1.5-year follow-up, then the disease progression time would be known to be in (1, 1.5], contributing an interval-censored failure time.

In this paper we focus on regression analysis of interval-censored failure time data. To this end, several methods have been proposed in the literature. For a comprehensive review, see [29,39]. In particular, [40] considered a class of linear transformation models which contain the proportional hazards model and the proportional odds model as special cases. In that paper they proposed an estimating equation approach to estimate the regression parameters and showed that the estimators always exist, are unique and consistent. To draw inference, they also established a normal approximation for the asymptotic distribution of the estimators using a heuristic argument, on which approximate confidence intervals are based. However, when sample sizes are small or the right censoring rate is heavy, the empirical coverage probabilities of the normality based confidence intervals may be seriously below the nominal levels, suggesting poor approximation. In this paper, we try to solve this problem by applying empirical likelihood (EL) approaches.
The EL is an appealing nonparametric approach for constructing confidence regions for parameter of interest. Owen [22,23] systematically developed novel EL methods for the mean of a random vector and other parameters of interest in the complete data setting. Since [26], the EL method has been commonly applied to make inferences based on all kinds of estimating equations in different contexts. New advances in the empirical likelihood research include the copula data analysis [5,24,25], the jackknife EL method [15,9,34], the high dimensional EL method [13,5], the penalized high dimensional EL method [31,18,17], the ROC curve analysis using EL procedures [8,11,1,20,35], general Cox models with EL [30], etc.

Based on the pseudo-observation idea of [27], [16] proposed an innovative EL procedure for general regression parameters. For special cases, for example, when \( F \) is the logistic distribution, \( F \) is the extreme value distribution, i.e.,

\[
F(z) = \begin{cases} 
1 - \exp(1/z) & \text{for } z \geq 0, \\
\exp(-1/z) & \text{for } z < 0.
\end{cases}
\]

Throughout the paper, we adopt the same notations as those in [40] for simplicity. Let \( T \) be the continuous failure time of interest as in Section 1 and \( Z \) denote a \( p \times 1 \) vector of covariates. One considers the following linear transformation model like [40],

\[
u(T) = Z^T \beta + \epsilon,
\]

where \( u \) is an unknown strictly increasing function, \( \beta \) is a \( p \times 1 \) vector of regression parameters and \( \epsilon \) has a pre-specified distribution function \( F \). It is well known that the linear transformation model includes some commonly used models as special cases. For example, when \( F(t) \) is the extreme value distribution, i.e., \( F(t) = 1 - \exp(-\exp(t)) \), (2.1) is the proportional hazards model. When \( F \) is the standard logistic distribution, (2.1) is exactly the proportional odds model. Equivalently, when \( F \) is strictly increasing, (2.1) can be written as

\[
g[S_2(t)] = u(t) - z^T \beta,
\]

where \( g^{-1}(s) = 1 - F(s) \) and \( S_2 \) denotes the survival function of \( T \) given \( Z = z \).

Now suppose that we have \( n \) i.i.d. replicates \( \{(L_i, R_i, T_i, Z_i); i = 1, \ldots, n\} \) of \( (L, R, T, Z) \) and the fact that \( P(L < T \leq R) = 1 \). Due to the interval censoring, we can only observe \( \{(L_i, R_i, Z_i); i = 1, \ldots, n\} \). Let \( H_r = 1 - S_r \) denote the true distribution function of \( T \) given \( Z = z \) and \( H_r \) a consistent estimator of \( H_r \) based on the observed data \( (L_i, R_i) \) on subjects with \( Z_i = z \) (for instance, the [32] estimator). Zhang et al. [40] discussed the fitting of model (2.1) to interval-censored failure time data when \( Z \) is a categorical variable that takes finitely many values. Under the conditional non-informative censoring assumption [21], Eq. (2) of [40], [38,39], they showed that

\[
E \left\{ (a_j)_{i=1}^{n} \int_{L_i}^{R_i} I(t_i \geq t_j) \, dH_{Z_i}(t_i) \, dH_{Z_i}(t_i) \bigg| Z_i, Z_j \right\} = E(I(T_i \geq T_j)|Z_i, Z_j),
\]

where \( a_i = \int_{L_i}^{R_i} dH_{Z_i}(t), i = 1, \ldots, n \). Because

\[
E(I(T_i \geq T_j)|Z_i, Z_j) = E(I(\epsilon_i - \epsilon_j \geq Z_i^T \beta)|Z_i, Z_j) = \tau(Z_i^T \beta),
\]

(2.3)

where \( \epsilon_i = u(T_i) - Z_i^T \beta, \beta_j = Z_j - Z_i, i, j = 1, \ldots, n \) and \( \tau(t) = \int_{t}^{\infty} (1 - F(s + t)) \, dF(s) \), [40] proposed the following estimating equation to estimate \( \beta \)

\[
U(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{n} \tau'(Z_i^T \beta)Z_i \left\{ \int_{L_i}^{R_i} I(t_i \geq t_j) \, d\hat{H}_{Z_i}(t_j) \, d\hat{H}_{Z_i}(t_i) \middle/ \hat{a}_i \hat{a}_j \right\} - \tau(Z_i^T \beta) = 0,
\]

(2.4)

where \( \hat{a}_i = \int_{L_i}^{R_i} d\hat{H}_{Z_i}(t) \) and \( \tau'(t) \) is the first derivative of \( \tau(t) \) (see condition C.2 in Section 2.2) and has the form

\[
\tau'(t) = - \int_{-\infty}^{t} f(s + t) \, dF(s),
\]

where \( f \) is the density function of \( \epsilon \) and is assumed to be bounded.

2. Main results

2.1. Preliminaries

Throughout the paper, we adopt the same notations as those in [40] for simplicity. Let \( T \) be the continuous failure time of interest as in Section 1 and \( Z \) denote a \( p \times 1 \) vector of covariates. One considers the following linear transformation model like [40],

\[
u(T) = Z^T \beta + \epsilon,
\]

where \( u \) is an unknown strictly increasing function, \( \beta \) is a \( p \times 1 \) vector of regression parameters and \( \epsilon \) has a pre-specified distribution function \( F \). It is well known that the linear transformation model includes some commonly used models as special cases. For example, when \( F(t) \) is the extreme value distribution, i.e., \( F(t) = 1 - \exp(-\exp(t)) \), (2.1) is the proportional hazards model. When \( F \) is the standard logistic distribution, (2.1) is exactly the proportional odds model. Equivalently, when \( F \) is strictly increasing, (2.1) can be written as

\[
g[S_2(t)] = u(t) - z^T \beta,
\]

where \( g^{-1}(s) = 1 - F(s) \) and \( S_2 \) denotes the survival function of \( T \) given \( Z = z \).

Now suppose that we have \( n \) i.i.d. replicates \( \{(L_i, R_i, T_i, Z_i); i = 1, \ldots, n\} \) of \( (L, R, T, Z) \) and the fact that \( P(L < T \leq R) = 1 \). Due to the interval censoring, we can only observe \( \{(L_i, R_i, Z_i); i = 1, \ldots, n\} \). Let \( H_r = 1 - S_r \) denote the true distribution function of \( T \) given \( Z = z \) and \( H_r \) a consistent estimator of \( H_r \) based on the observed data \( (L_i, R_i) \) on subjects with \( Z_i = z \) (for instance, the [32] estimator). Zhang et al. [40] discussed the fitting of model (2.1) to interval-censored failure time data when \( Z \) is a categorical variable that takes finitely many values. Under the conditional non-informative censoring assumption [21], Eq. (2) of [40], [38,39], they showed that

\[
E \left\{ (a_j)_{i=1}^{n} \int_{L_i}^{R_i} I(t_i \geq t_j) \, dH_{Z_i}(t_i) \, dH_{Z_i}(t_i) \bigg| Z_i, Z_j \right\} = E(I(T_i \geq T_j)|Z_i, Z_j),
\]

where \( a_i = \int_{L_i}^{R_i} dH_{Z_i}(t), i = 1, \ldots, n \). Because

\[
E(I(T_i \geq T_j)|Z_i, Z_j) = E(I(\epsilon_i - \epsilon_j \geq Z_i^T \beta)|Z_i, Z_j) = \tau(Z_i^T \beta),
\]

where \( \epsilon_i = u(T_i) - Z_i^T \beta, Z_i = Z_j - Z_i, i, j = 1, \ldots, n \) and \( \tau(t) = \int_{t}^{\infty} (1 - F(s + t)) \, dF(s) \), [40] proposed the following estimating equation to estimate \( \beta \)

\[
U(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{n} \tau'(Z_i^T \beta)Z_i \left\{ \int_{L_i}^{R_i} I(t_i \geq t_j) \, d\hat{H}_{Z_i}(t_j) \, d\hat{H}_{Z_i}(t_i) \middle/ \hat{a}_i \hat{a}_j \right\} - \tau(Z_i^T \beta) = 0,
\]

where \( \hat{a}_i = \int_{L_i}^{R_i} d\hat{H}_{Z_i}(t) \) and \( \tau'(t) \) is the first derivative of \( \tau(t) \) (see condition C.2 in Section 2.2) and has the form

\[
\tau'(t) = - \int_{-\infty}^{t} f(s + t) \, dF(s),
\]

where \( f \) is the density function of \( \epsilon \) and is assumed to be bounded.
Assume that \( \hat{\beta} \) is the solution to \( U(\beta) = 0 \) and \( \beta_0 \) is the true value of \( \beta \), it was shown in [40] that \( \hat{\beta} \) is unique for large \( n \) and consistent, and heuristically the distribution of \( n^{1/2}(\hat{\beta} - \beta_0) \) can be approximated by a normal distribution with mean zero and covariance matrix \( \Lambda^{-1} \Sigma \Lambda^{-1} \) (see Appendix C of [40]), where

\[
\Lambda = \lim_{n \to \infty} \left\{ \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left( \tau' (Z_i^T \beta_0) Z_j^T \right)^2 Z_i Z_j^T \right\},
\]

\[
\Sigma = \lim_{n \to \infty} \left\{ \frac{1}{n^3} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k \neq j} \left( d_{ij}(\beta_0) + d_{ik}(\beta_0) \right) \left( d_{kj}(\beta_0) + d_{kl}(\beta_0) \right) \right\},
\]

\[
d_{ij}(\beta) = \tau'(Z_i^T \beta) Z_{ij} \left\{ (a_i a_j)^{-1} \int_{t_i}^{R_i} \int_{t_j}^{R_j} \chi(t_i, t_j) dH_{Z_i}(t_i) dH_{Z_j}(t_j) \right\}.
\]

Let

\[
\hat{\Lambda} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left( \tau' (Z_i^T \hat{\beta}) Z_j^T \right)^2 Z_i Z_j^T,
\]

and

\[
\hat{\Sigma} = \frac{1}{n^3} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k \neq j} \left( \hat{d}_{ij}(\hat{\beta}) + \hat{d}_{ik}(\hat{\beta}) \right) \left( \hat{d}_{kj}(\hat{\beta}) + \hat{d}_{kl}(\hat{\beta}) \right),
\]

where \( \hat{d}_{ij}(\beta) \) is equal to \( d_{ij}(\beta) \) with replacing \( H_Z, \), \( \hat{a}_{ij} \), and \( \hat{a}_i \), respectively. In the Appendix we show that \( \Sigma \) and \( \Lambda \) are consistently estimated by \( \hat{\Sigma} \) and \( \hat{\Lambda} \), respectively (see Lemma A.2). Thus an asymptotic 100(1 - \( \alpha \))% confidence region for \( \beta \) based on the above normal approximation is given by

\[
\mathcal{R}_1 = \{ \beta : n(\hat{\beta} - \beta)^T \hat{\Lambda}^{-1} \hat{\Lambda} (\hat{\beta} - \beta) \leq \chi^2_p(\alpha) \},
\]

where \( \chi^2_p(\alpha) \) is the upper \( \alpha \)-quantile of the chi-square distribution with degrees of freedom \( p \). Although the overall performance of this normal approximation is satisfactory as shown in the simulation results of [40], there exists an under-coverage problem for small sample sizes and heavy right censoring scenarios.

2.2. EL-based inference procedures

To overcome the under-coverage problem for the normal approximation method discussed above, we adopt a pseudo EL approach proposed by jing et al. [16]. We define \( X_i = (L_i, R_i, Z_i) \). To symmetrize, let \( b(X_i, X_j; \beta) = d_{ij}(\beta) + d_{ji}(\beta) \) and \( \hat{b}(X_i, X_j; \beta) = \hat{d}_{ij}(\beta) + \hat{d}_{ji}(\beta) \). As [16], we define

\[
W_i(\beta) = \frac{1}{n - 1} \sum_{j=1, j \neq i}^{n} b(X_i, X_j; \beta) \quad \text{and} \quad \hat{W}_i(\beta) = \frac{1}{n - 1} \sum_{j=1, j \neq i}^{n} \hat{b}(X_i, X_j; \beta)
\]

for \( i = 1, \ldots, n \). For each fixed \( \beta \), the \( W_i(\beta) \)'s are identically distributed (but not independent) and because of (2.2) and (2.3), we have \( E(W_i(\beta_0)) = 0, i = 1, \ldots, n \). Define

\[
Q(\beta) = \frac{1}{n} \sum_{i=1}^{n} W_i(\beta) \quad \text{and} \quad \hat{Q}(\beta) = \frac{1}{n} \sum_{i=1}^{n} \hat{W}_i(\beta).
\]

Obviously, \( Q(\beta) \) is a \( p \)-dimensional multivariate \( U \)-statistic for fixed \( \beta \) and can be approximated by \( \hat{Q}(\beta) \), which is merely an alternative expression of the estimating function \( U(\beta) \) in (2.4) because \( \binom{n}{2} \hat{Q}(\beta) = U(\beta) \). Then based on the pseudo-observations \( W_i \), the proposed EL at \( \beta \) is given by

\[
L(\beta) = \sup \left\{ \prod_{i=1}^{n} p_i : \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i W_i(\beta) = 0, p_i \geq 0 \right\}.
\]

Since the \( W_i \)'s depend on \( H_Z(\cdot) \) which is unknown, we replace them by the \( \hat{W}_i \)'s. Therefore, by introducing the notation \( \hat{L}(\beta) \), an estimated EL evaluated at \( \beta \) is given by

\[
\hat{L}(\beta) = \sup \left\{ \prod_{i=1}^{n} p_i : \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i \hat{W}_i(\beta) = 0, p_i \geq 0 \right\}.
\]

In the above, a unique maximum exists for a given \( \beta \), provided that 0 is inside the convex hull of the points \((\hat{W}_1(\beta), \ldots, \hat{W}_n(\beta))\) (if not we will set \( \hat{L}(\beta) = 0 \) to complete the definition). Note that \( \prod_{i=1}^{n} p_i \) attains its maximum at
\( p_i = 1/n \) under the restrictions \( p_i \geq 0 \) for \( i = 1, \ldots, n \) and \( \sum_{i=1}^{n} p_i = 1 \). Thus, the empirical likelihood ratio at \( \beta \) may be defined by
\[
\hat{R}(\beta) = \sup \left\{ \prod_{i=1}^{n} n p_i : \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i \hat{W}_i(\beta) = 0, p_i \geq 0 \right\}.
\]

Using the Lagrange multiplier approach as in [22,23,26], we have \( p_i = 1/\{n(1 + \lambda^T \hat{W}_i(\beta))\} \) for \( i = 1, \ldots, n \) and thus
\[
\hat{l}(\beta) := -2 \log \hat{R}(\beta) = 2 \sum_{i=1}^{n} \log(1 + \lambda^T \hat{W}_i(\beta)),
\]
where \( \lambda = (\lambda_1, \ldots, \lambda_p)^T \) satisfies the equation
\[
\hat{g}(\lambda) := \frac{1}{n} \sum_{i=1}^{n} \frac{\hat{W}_i(\beta)}{1 + \lambda^T \hat{W}_i(\beta)} = 0.
\]

Accordingly, we let \( \hat{R}(\beta) = 0 \) and \( \hat{l}(\beta) = \infty \) if \( 0 \) is outside of the convex hull of the points \( (\hat{W}_1(\beta), \ldots, \hat{W}_n(\beta)) \).

To obtain our main results, we require the following regularity conditions that are commonly used in survival analysis (cf. [40]):

C.1. The covariate vector \( Z \) is bounded, i.e., \( ||Z|| \leq D \) for some positive constant \( D \), where \( || \cdot || \) is the Euclidean norm.

C.2. The function \( \tau(\cdot) \) is twice continuously differentiable. Denote its first derivatives by \( \tau'(\cdot) \).

C.3. Both matrices \( \Lambda \) and \( \Sigma \) are finite and positive definite.

C.4. There exists a \( \delta > 0 \) such that \( P[H_z(R) - H_z(L) > \delta] = 1 \) for all values of \( z \).

**Remark 1.** Theorem 2.1 below only requires that \( \tau(\cdot) \) is continuously differentiable. Existence and continuity of the second derivative of \( \tau(\cdot) \), however, are needed by Theorem 2.2 in order to apply the results from [26].

**Remark 2.** In the third regularity condition above, note that \( \Sigma \), defined as a limit in the last subsection, can also be written as \( E(\tau(\beta_0)) \beta_0 \beta_0^T \) by using the strong law of large numbers for \( U \)-statistics. Furthermore, we adopt the notion in [40, Appendix C] that the variance of \( n^{-3/2}U(\beta_0) \) can be approximated by \( \Sigma \). Note that this approximation essentially ignores the fact that \( H_z \) is an estimator of \( H_z \), hence the extra variability brought in. As will be shown in the technical proofs and by the simulation results, this approximation greatly simplifies the technical arguments and appears to work quite well. More discussions will be found in Section 5.

**Remark 3.** The fourth condition essentially says that the width of the interval containing \( T \) is wide enough to be a real interval (non-degenerate) and does not give rise to an exact observation. Similar conditions are imposed, for instance, in [12, p. 82].

Now we establish the main theorem and explain how they can be used to construct confidence regions for the true, unknown regression parameter \( \beta_0 \).

**Theorem 2.1.** Under (C.1)–(C.4), and approximating the variance of \( n^{-3/2}U(\beta_0) \) by \( \Sigma \), the EL statistic \( \hat{l}(\beta_0) \) converges in distribution to \( 4 \chi^2_p \), where \( \chi^2_p \) is a chi-square distribution with \( p \) degrees of freedom.

Using **Theorem 2.1**, an asymptotic 100(1 – \( \alpha \))% EL confidence region for \( \beta_0 \) is given by
\[
\mathcal{R}_2 = \left\{ \beta : \hat{l}(\beta) \leq 4 \chi^2_p(\alpha) \right\},
\]
where \( \chi^2_p(\alpha) \) is the upper \( \alpha \)-quantile of the distribution of \( \chi^2_p \).

In practice, confidence regions for the regression parameter \( \beta \) may not be very useful. One is often more interested in constructing confidence intervals for single components of regression parameters \( \beta \). Therefore, we present a way to construct EL confidence regions for a \( q \)-dimensional \( (q < p) \) subvector \( \beta^{(1)} \) of \( \beta = (\beta^{(1)^T}, \beta^{(2)^T})^T \) (the true value \( \beta_0 \) is accordingly partitioned as \( \beta_0 = (\beta_0^{(1)^T}, \beta_0^{(2)^T})^T \)). Yu et al. [37] proposed a profile EL for linear transformation models in this regard. Motivated by their method, we propose the profile EL for single components \( \beta^{(1)} \) of \( \beta \) by dealing with nuisance parameters \( \beta^{(2)} \) from the full EL.

We define the profile EL ratio at \( \beta^{(1)} \) as
\[
\tilde{l}(\beta^{(1)}) = \inf_{\beta^{(2)} \in \mathcal{R}^{p-q}} \hat{l}(\beta^{(1)^T}, \beta^{(2)^T})^T),
\]
where \( \mathcal{R}^{p-q} \) denotes the \( (p - q) \)-dimensional Euclidean space. We give the theorem as that for the full EL as follows.

**Theorem 2.2.** Under (C.1)–(C.4), and approximating the variance of \( n^{-3/2}U(\beta_0) \) by \( \Sigma \), the profile EL statistic \( \tilde{l}(\beta_0^{(1)}) \) converges in distribution to \( 4 \chi^2_q \).
Using Theorem 2.2, an asymptotic 100(1 − α)% EL confidence region for \( \beta_0^{(1)} \) is given by

\[
\mathcal{R}_3 = \left\{ \beta^{(1)} : \hat{l}(\beta^{(1)}) \leq 4\chi^2_q(\alpha) \right\}.
\] (2.8)

2.3. Jackknife EL-based inference procedures

Recently, [15] proposed a jackknife empirical likelihood (JEL) method that can be applied to inferences involving \( U \)-statistics. This approach can be adapted to our problem as well. Note that

\[
Q(\beta) = \frac{1}{n} \sum_{i=1}^{n} w_i(\beta) = \frac{1}{\binom{n}{2}} \sum_{i,j} b(X_i, X_j; \beta),
\]

where, to simplify the notation, we use \( \sum_{i,j} \) instead of \( \sum_{i=1,j=1,i,j\neq i} \) in this section. To take advantage of the jackknife method, we define

\[
Q^{(-i)}(\beta) = \frac{1}{\binom{n-1}{2}} \sum_{i,j\neq i} b(X_i, X_j; \beta),
\]

where \( \sum_{i,j\neq i} \) denotes \( \sum_{i=1,j=1,i,j\neq i}^{n} \). Let \( V_i(\beta) = nQ(\beta) - (n - 1)Q^{(-i)}(\beta), \ l = 1, \ldots, n \), be the jackknife pseudo-values. It can be easily shown that \( \frac{1}{n} \sum_{i=1}^{n} V_i(\beta) = Q(\beta) \). Jing et al. [15] showed that inferences about \( \beta \) can be made via \( V_i(\beta) \) in the following way. Define the JEL at \( \beta \) to be

\[
L_j(\beta) = \sup \left\{ \frac{n}{\sum_{i=1}^{n} p_i} : \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i V_i(\beta) = 0, p_i \geq 0 \right\}.
\]

Then by similar arguments as in Section 2.2, it can be shown that the JEL ratio at \( \beta \) is

\[
R_j(\beta) = \sup \left\{ \frac{n}{\sum_{i=1}^{n} p_i} : \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i V_i(\beta) = 0, p_i \geq 0 \right\},
\]

and

\[
l_j(\beta) := -2 \log R_j(\beta) = 2 \sum_{i=1}^{n} \log(1 + \lambda^T V_i(\beta)) \]

where \( \lambda \) satisfies the equation

\[
g_j(\lambda) := \frac{1}{n} \sum_{i=1}^{n} \frac{V_i(\beta)}{1 + \lambda^T V_i(\beta)} = 0.
\]

We shall define \( R_j(\beta) = 0 \) and \( l_j(\beta) = \infty \) if 0 is outside of the convex hull of the points \( (V_1(\beta), \ldots, V_n(\beta)) \).

Let \( \hat{Q}^{(-i)}(\beta), \hat{V}_i(\beta), \hat{L}_i(\beta), \hat{R}_i(\beta) \) and \( \hat{g}_j(\lambda) \) be approximations of their corresponding values by using \( \hat{b}(X_i, X_j; \beta) \) instead of \( b(X_i, X_j; \beta) \) (or equivalently, by replacing \( F_0 \) with \( F_0 \)) as in Section 2.2. We have the following results for the JEL procedure. These Wilks’s theorems enable us to construct confidence regions for regression parameters.

**Theorem 2.3.** Under (C.1)–(C.4), and approximating the variance of \( n^{-3/2}U(\beta_0) \) by \( \Sigma \), the JEL statistic \( \hat{l}_j(\beta_0) \) converges in distribution to \( \chi^2_p \).

By Theorem 2.3, an asymptotic 100(1 − α)% JEL confidence region for \( \beta_0 \) is given by

\[
\mathcal{R}_4 = \left\{ \beta : \hat{l}_j(\beta) \leq \chi^2_p(\alpha) \right\}.
\] (2.9)

**Theorem 2.4.** Under (C.1)–(C.4), and approximating the variance of \( n^{-3/2}U(\beta_0) \) by \( \Sigma \), the profile JEL statistic \( \hat{l}_j(\beta_0^{(1)}) \) converges in distribution to \( \chi^2_q \).

Based on Theorem 2.4, an asymptotic 100(1 − α)% JEL confidence region for \( \beta_0^{(1)} \) is given by

\[
\mathcal{R}_5 = \left\{ \beta^{(1)} : \hat{l}_j(\beta^{(1)}) \leq \chi^2_q(\alpha) \right\}.
\] (2.10)

3. Simulation results

To evaluate the performance of the EL and JEL estimators, we conduct several simulation studies. We consider the two special cases of the linear transformation models as in [40], i.e., the proportional hazards model and the proportional odds model. For the former, a continuous survival time \( T \) is generated from a proportional hazards model with a hazard function
Table 1
Comparison of coverage probability and mean width of 95% confidence interval derived from the three methods.

<table>
<thead>
<tr>
<th>True</th>
<th>N</th>
<th>C%</th>
<th>Bias</th>
<th>SSD</th>
<th>MSE</th>
<th>NA</th>
<th>WCI</th>
<th>EL</th>
<th>WCI</th>
<th>JEL</th>
<th>WCI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td></td>
<td>CP</td>
<td></td>
<td>CP</td>
<td></td>
</tr>
<tr>
<td>β = 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>10</td>
<td>−0.005</td>
<td>0.521</td>
<td>0.409</td>
<td>0.884</td>
<td>1.601</td>
<td>0.929</td>
<td>1.843</td>
<td>0.915</td>
<td>1.976</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>−0.009</td>
<td>0.562</td>
<td>0.437</td>
<td>0.881</td>
<td>1.713</td>
<td>0.918</td>
<td>1.974</td>
<td>0.913</td>
<td>2.154</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>−0.005</td>
<td>0.667</td>
<td>0.496</td>
<td>0.876</td>
<td>1.944</td>
<td>0.898</td>
<td>2.233</td>
<td>0.900</td>
<td>2.539</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>−0.006</td>
<td>0.378</td>
<td>0.320</td>
<td>0.911</td>
<td>1.255</td>
<td>0.934</td>
<td>1.359</td>
<td>0.928</td>
<td>1.410</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>−0.007</td>
<td>0.412</td>
<td>0.342</td>
<td>0.903</td>
<td>1.342</td>
<td>0.919</td>
<td>1.450</td>
<td>0.919</td>
<td>1.518</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>−0.002</td>
<td>0.484</td>
<td>0.386</td>
<td>0.893</td>
<td>1.513</td>
<td>0.905</td>
<td>1.613</td>
<td>0.909</td>
<td>1.725</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>0.000</td>
<td>0.251</td>
<td>0.228</td>
<td>0.925</td>
<td>0.894</td>
<td>0.929</td>
<td>0.933</td>
<td>0.945</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.001</td>
<td>0.273</td>
<td>0.243</td>
<td>0.921</td>
<td>0.954</td>
<td>0.926</td>
<td>0.991</td>
<td>0.926</td>
<td>1.011</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.007</td>
<td>0.313</td>
<td>0.274</td>
<td>0.912</td>
<td>1.074</td>
<td>0.915</td>
<td>1.107</td>
<td>0.919</td>
<td>1.141</td>
<td></td>
<td></td>
</tr>
<tr>
<td>β = 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>10</td>
<td>0.077</td>
<td>0.547</td>
<td>0.444</td>
<td>0.908</td>
<td>1.739</td>
<td>0.938</td>
<td>2.012</td>
<td>0.933</td>
<td>2.344</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.092</td>
<td>0.590</td>
<td>0.468</td>
<td>0.896</td>
<td>1.835</td>
<td>0.930</td>
<td>2.133</td>
<td>0.924</td>
<td>2.567</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.102</td>
<td>0.695</td>
<td>0.542</td>
<td>0.891</td>
<td>2.125</td>
<td>0.908</td>
<td>2.432</td>
<td>0.912</td>
<td>2.925</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.005</td>
<td>0.734</td>
<td>0.643</td>
<td>0.925</td>
<td>2.521</td>
<td>0.959</td>
<td>2.923</td>
<td>0.952</td>
<td>3.159</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.006</td>
<td>0.850</td>
<td>0.751</td>
<td>0.926</td>
<td>2.944</td>
<td>0.956</td>
<td>3.414</td>
<td>0.954</td>
<td>3.817</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.006</td>
<td>0.625</td>
<td>0.593</td>
<td>0.941</td>
<td>2.325</td>
<td>0.955</td>
<td>2.490</td>
<td>0.953</td>
<td>2.659</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>0.033</td>
<td>0.274</td>
<td>0.251</td>
<td>0.935</td>
<td>0.982</td>
<td>0.945</td>
<td>1.024</td>
<td>0.942</td>
<td>1.049</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.034</td>
<td>0.291</td>
<td>0.264</td>
<td>0.932</td>
<td>1.036</td>
<td>0.942</td>
<td>1.081</td>
<td>0.938</td>
<td>1.109</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.036</td>
<td>0.335</td>
<td>0.303</td>
<td>0.933</td>
<td>1.188</td>
<td>0.940</td>
<td>1.229</td>
<td>0.936</td>
<td>1.276</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: SSD represents the sample standard deviation based on the 3000 point estimates. MSE represents mean of the 3000 standard errors. CP represents the coverage probability of the 95% confidence interval. WCI represents the width of the 95% confidence interval. C% represents the percentage of right censoring. NA represents the normal approximation. EL represents the empirical likelihood method. JEL represents the jackknife empirical likelihood method.

The survival time is generated by taking the link function $u(t) = \log(t)$ and $F(t)$ is the standard logistic distribution. The censoring intervals and right censoring rate are generated similarly as in [38].

The censoring intervals and right censoring rate are generated similarly as in the proportional hazards model. All numerical results, summarized in Table 1, are based on 3000 simulated replicates.

We compare the three methods, normal approximation, EL and JEL via empirical coverage probability and width of the confidence interval at 95% nominal level. It can be seen from Table 1 that the coverage probabilities of EL and JEL are closer to their nominal level than those of normal approximation. This is possibly due to the widening of the confidence intervals as it is apparent in Table 1 (comparing SSD and MSE) that the standard errors of the regression parameter estimator are underestimated by the normal approximation. However, we caution here that the EL and JEL confidence intervals do not completely eliminate the estimation bias because of the approximation of the variance of $n^{-3/2}U(\beta_0)$ adopted in Theorems 2.1 and 2.2.

$e^{\beta}$ (i.e., $u(t) = \log(t)$ and $F(t)$ is the extreme value distribution), where $Z$ is a binary covariate taking values 0 and 1 with probabilities 0.5. $\beta$ is the regression parameter to be estimated with the true value 0 or 1. The generated survival time is then added by 1 to ensure that it stays away from 0 to facilitate the generation of interval censoring. This modification does not alter estimation of $\beta$ since the model is rank invariant. Non-informative interval censoring is generated by adding or subtracting continuous uniform $[0, 0.5]$ random variables from the survival time and then being modified as in [38]. A right censoring time following a uniform distribution $\cup_{a,b}$ is also implemented, with $a$ and $b$ varying to control the right censoring rate to be 10%, 30% and 50%. Sample sizes examined are 30, 50 and 100. For the proportional odds model, the survival time is generated by taking the link function $u(t) = \log(0.08t)$ and $F(t)$ to be the standard logistic distribution. The censoring intervals and right censoring rate are generated similarly as in the proportional hazards model. All numerical results, summarized in Table 1, are based on 3000 simulated replicates.
4. Applications

To further illustrate our methods we apply the EL and JEL approaches to two interval-censored data sets analyzed in [40].

The first data set was based on a breast cancer study discussed in [10], in which totally 94 patients were included. Among them, 46 patients received radiation therapy only \((Z = 0)\) while the remaining 48 patients received both radiation therapy and chemotherapy \((Z = 1)\). We are interested in investigating the association between the treatment modality and the failure time of interest, which is time to appearance of breast retraction [40]. Applying estimating equation (2.4) with \(F\) being the extreme value distribution (i.e., a proportional hazards model), we obtained the point estimate of \(\beta\) as \(-0.697\) with an estimated standard error of 0.251 based on the normal approximation as [40] did. The corresponding 95% confidence interval, together with the EL and JEL 95% confidence intervals, are shown in Table 2. The results suggest that \(\beta\) is significantly different from 0, indicating that patients who had both radiation therapy and chemotherapy tend to have shorter time to breast retraction. If \(F\) takes the standard logistic distribution, that is, when the linear transformation model reduces to the proportional odds model, the point estimate and its estimated standard error for \(\beta\) are \(-1.041\) and 0.372, which are reported in Table 2. The three 95% confidence intervals for \(\beta\) from the normal approximation, EL and JEL methods, also given in Table 2, suggest the same conclusion.

The second data set was based on an HIV study conducted in the 1980s and discussed in [3]. The purpose of this study is to investigate the HIV-1 infection rate among a group of haemophiliac patients. “These patients were at risk of HIV-1 infection because they received blood products containing factors VIII and IX concentrate made from plasma of many donors” [40]. The clinical endpoint is therefore time to HIV-1 infection, which was interval-censored due to missed follow-ups and long gap time between monitoring visits. Here for illustration purpose we focus on a subset of patients who received no factor VIII concentrate (236 patients with \(Z = 0\)) or low dose (1–20000 U/year) VIII concentrate (132 patients with \(Z = 1\)). Applying the estimating equation (2.4) with \(F\) being the extreme value distribution, that is, the proportional hazards model, we obtained the point estimate of \(\beta\) as \(-1.016\) with an estimated standard error of 0.124 based on the normal approximation (see [40]). If we let \(F\) be the standard logistic distribution and thus (2.1) becomes the proportional odds model, the point estimate and its estimated standard error for \(\beta\) are \(-1.511\) and 0.182, which are reported in Table 3. The three 95% confidence intervals for \(\beta\) from the normal approximation, EL and JEL methods, also shown in Table 2, suggest the same conclusion.

Note: NA represents the normal approximation. EL represents the empirical likelihood method. JEL represents the jackknife empirical likelihood method.

Table 2
Breast cancer study analysis results.

<table>
<thead>
<tr>
<th>Proportional hazards model</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(\hat{\beta})</td>
<td>95% NA CI</td>
<td>95% EL CI</td>
<td>95% JEL CI</td>
</tr>
<tr>
<td>(-0.697)</td>
<td>((-1.181, -0.213))</td>
<td>((-1.160, -0.151))</td>
<td>((-1.214, -0.179))</td>
</tr>
</tbody>
</table>

Proportional odds model

| \(-1.041\) | \((-1.769, -0.310)\) | \((-1.722, -0.227)\) | \((-1.800, -0.269)\) |

Note: NA represents the normal approximation. EL represents the empirical likelihood method. JEL represents the jackknife empirical likelihood method.

Table 3
HIV study analysis results.

<table>
<thead>
<tr>
<th>Proportional hazards model</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(\hat{\beta})</td>
<td>95% NA CI</td>
<td>95% EL CI</td>
<td>95% JEL CI</td>
</tr>
<tr>
<td>(-1.016)</td>
<td>((-1.259, -0.773))</td>
<td>((-1.266, -0.779))</td>
<td>((-1.272, -0.780))</td>
</tr>
</tbody>
</table>

Proportional odds model

| \(-1.511\) | \((-1.868, -1.154)\) | \((-1.876, -1.162)\) | \((-1.884, -1.165)\) |

Note: NA represents the normal approximation. EL represents the empirical likelihood method. JEL represents the jackknife empirical likelihood method.

5. Discussion

There exist several algorithms for deriving the nonparametric maximum likelihood estimator (NPMLE) of the survival function of a random variable when it is interval censored. Though a consistent estimator, this NPMLE has an unusual convergence rate of \(n^{-\frac{1}{3}}\) and the limiting distribution is non-normal and complicated. Therefore, in this study we have adopted an approximation for estimating the asymptotic variance of \(n^{-\frac{1}{2}}U(\beta_0)\). This approximation results in a slight underestimation of the variability of \(n^{-\frac{1}{2}}U(\beta_0)\), which can be seen from the simulation results. Namely, the empirical coverage probability of the 95% confidence intervals is still slightly below the nominal level even for the EL and JEL methods. As a matter of fact, if the real, unknown \(H_Z\) (instead of its estimator \(\hat{H}_Z\)) is used in estimation Eq. (2.4) for simulation, this under coverage problem would be completely gone (results not shown here). We are currently conducting research in hope of precisely estimating the asymptotic variance of \(n^{-\frac{1}{2}}U(\beta_0)\).
In our simulation studies, we find that results from the JEL approach and the EL approach are comparable, with EL perhaps slightly better in terms of both coverage probability and width of the confidence interval. We also tried larger sample sizes such as 300 or 500 and obtained similar results. In [15], they reported simulation results for JEL, NA and [33]'s sequential EL, Scaled-EL methods. They claimed that JEL is the best one among others. However there is not a direct comparison of EL and JEL. We will investigate the comparison of EL and JEL in the transformation model and other statistical models in the future.

Acknowledgments

The authors are grateful to the referee and the AE for their useful suggestions, which improved the paper significantly. Zhiqiang Zhang's research was partially supported by an NCI core grant. Yichuan Zhao's research was partially supported by a grant under the National Security Agency.

Appendix. Proofs of Theorems

To prove Theorems 2.1 and 2.2 we need the following two lemmas. The first one establishes the asymptotic normality of \( Q(\beta_0) \), which is Lemma A.1 of [41].

**Lemma A.1.** Under conditions (C.1)–(C.4), when \( n \to \infty \), we have \( n^{1/2}Q(\beta_0) \overset{D}{\to} N(0, 4\Sigma) \), where \( \overset{D}{\to} \) denotes convergence in distribution. In particular, the asymptotic variance of \( Q(\beta_0) \) has the following order of convergence

\[
\text{Var}(Q(\beta_0)) = \frac{4\Sigma}{n} + O(n^{-2}), \quad \text{a.s.} \quad (A.1)
\]

**Proof.** This is a standard result for multivariate U-statistics of degree 2. Detailed proof is summarized in [14,28,2], etc. \( \square \)

The next lemma establishes several convergence results regarding the variance estimators.

**Lemma A.2.** Let \( \Sigma_n = \sum_{i=1}^{n} W_i(\beta_0)W_i^T(\beta_0)/n \) and \( \Sigma_n = \sum_{i=1}^{n} \hat{W}_i(\beta_0)\hat{W}_i^T(\beta_0)/n \). Then under conditions (C.1)–(C.4), we have (i) \( \Sigma_n \overset{p}{\to} \Sigma \), (ii) \( \hat{\Sigma} \overset{p}{\to} \Sigma \), (iii) \( \hat{\Sigma} \overset{p}{\to} \Sigma \), and (iv) \( \Lambda \overset{p}{\to} \Lambda \), where \( \overset{p}{\to} \) denotes convergence in probability.

**Proof.** To prove (i), we first rewrite

\[
\Sigma_n = \frac{1}{n} \sum_{i=1}^{n} (W_i(\beta_0) - Q(\beta_0))W_i(\beta_0) - Q(\beta_0))T + Q(\beta_0)Q^T(\beta_0). \quad (A.2)
\]

By the strong law of large numbers for U-statistics we have that \( Q(\beta_0) = o(1), \) a.s. (recall that the expectation of \( Q(\beta_0) \) is 0). Thus, the result follows by Lemma A.2 of [41].

Next we prove (ii), we only need to show \( \Sigma_n = \Sigma_n + o_p(1) \). Since

\[
P \left\{ \limsup_{n \to \infty} \sup_{t \in \mathbb{R}} |H(t) - H(t)| = 0 \right\} = 1.
\]

Groeneboom and Wellner [12, p. 85] and Z takes finitely many values, we see that

\[
K_n \equiv \sup_{i=1, \ldots, n} \sup_{\{i, j\} \subseteq \mathbb{R}} \left| \int_{s}^{u} d\hat{H}_z(t) - \int_{s}^{u} dH_z(t) \right| = o_p(1).
\]

Using (C.4), the above result and the fact that all \( a_i \)'s and \( \hat{a}_i \)'s are at most 1, we have

\[
\sup_{i,j=1, \ldots, n} \left| \frac{1}{\hat{a}_i \hat{a}_j} - \frac{1}{a_i a_j} \right| \leq \sup_{i,j=1, \ldots, n} \frac{|a_i (a_j - \hat{a}_j)| + |\hat{a}_i (a_j - \hat{a}_j)|}{\hat{a}_i \hat{a}_i a_i a_j} \leq 2K_n \delta^{-4} + o_p(1) = o_p(1).
\]

Moreover,

\[
\sup_{i,j=1, \ldots, n} \left| \int_{l_i}^{R_i} \int_{l_j}^{R_j} I(t_i \geq t_j) d\hat{H}_z(t_j) d\hat{H}_z(t_i) - \int_{l_i}^{R_i} \int_{l_j}^{R_j} I(t_i \geq t_j) dH_z(t_j) dH_z(t_i) \right|
\]

\[
\leq \sup_{i,j=1, \ldots, n} \left| \int_{l_i}^{R_i} \left( \hat{H}(R_j \wedge t_i) - \hat{H}(L_j \wedge t_i) \right) - \left( H(R_j \wedge t_i) - H(L_j \wedge t_i) \right) \right| d\hat{H}_z(t_i) \]

\[
+ \left| \int_{l_i}^{R_i} \left( H(R_i \wedge t_i) - H(L_j \wedge t_i) \right) \right| d\hat{H}_z(t_i) - \int_{l_i}^{R_i} \left( H(R_i \wedge t_i) - H(L_j \wedge t_i) \right) dH_z(t_i) \right|.
\]
In the above, the first term on the right hand side is bounded by $2K_n$ uniformly. The second term uniformly (i.e., for all $i$ and $j$) converges to 0 because of Lemma 1 in the Appendix of [19]. Therefore, the supremum of the sum of the two terms is $o_P(1)$.

Using the above results, we obtain the uniform (over all $i$ and $j$) convergence of $\hat{d}_i(\beta_0)$ to $d_i(\beta_0)$:

$$
\sup_{i,j=1,\ldots,n} \left\| \hat{d}_i(\beta_0) - d_i(\beta_0) \right\|
\leq \sup_{i,j=1,\ldots,n} \left\| \tau'(Z^{ij}_n \beta) Z_i \right\| \cdot \left\{ (a_i a_j)^{-1} - (\hat{a}_i \hat{a}_j)^{-1} \right\}
\int_{t_i}^{R_i} \int_{t_j}^{R_j} \left( l(t_j \geq t_i) dH(t_j) dH(t_i) \right)
+ \left( \hat{a}_i \hat{a}_j \right)^{-1} \int_{t_i}^{R_i} \int_{t_j}^{R_j} \left( l(t_j \geq t_i) dH(t_j) dH(t_i) \right) - \int_{t_i}^{R_i} \int_{t_j}^{R_j} \left( l(t_j \geq t_i) d\hat{H}(t_j) d\hat{H}(t_i) \right)
= o_P(1),
$$

because both $\tau'(\cdot)$ and $Z$ are bounded. Thus a straightforward algebraic calculation shows that $\sup_{i=1,\ldots,n} \left\| W_i(\beta_0) - \hat{W}_i(\beta_0) \right\|$ is $o_P(1)$.

Note that $\frac{1}{n} \sum_{i=1}^n W_i(\beta_0) = O_P(n^{-1/2}) = o_P(1)$ due to Lemma A.1 and that $\left\| W_i(\beta_0) \right\|$ is uniformly bounded (since $\left\| d_j(\beta_0) \right\|$ is uniformly bounded) by a constant, say, $M_0$. For each fixed $a \in \mathbb{R}^p$, we finally have

$$
a^T(\hat{\Sigma}_n - \Sigma_n)a = \frac{1}{n} \sum_{i=1}^n \left[ a^T \hat{W}_i(\beta_0) - W_i(\beta_0) \right]^2 + \frac{2}{n} \sum_{i=1}^n \left[ a^T W_i(\beta_0) \right] \left[ a^T \hat{W}_i(\beta_0) - W_i(\beta_0) \right].
$$

In the above, the first term on the right hand side is $o_P(1)$ because $\sup_{i=1,\ldots,n} \left\| W_i(\beta_0) - \hat{W}_i(\beta_0) \right\|$ is $o_P(1)$. The second term is also $o_P(1)$ because $\left\| W_i(\beta_0) \right\|$ is uniformly bounded (for all $i$). Hence the claimed result follows.

To prove (iii), we first note that $\sup_{i=1,\ldots,n} \left\| \hat{d}_i(\beta) - d_i(\beta_0) \right\| = o_P(1)$ due to the consistency of $\hat{\beta}$, the continuity of $\tau(\cdot)$ and $\tau'(\cdot)$ and the fact that $Z$ is bounded. Thus we also have $\sup_{i=1,\ldots,n} \left\| \hat{W}_i(\beta) - W_i(\beta_0) \right\| = o_P(1)$. This means that both $\hat{W}_i(\beta_0)$ and $\hat{W}_i(\beta)$ are uniformly bounded. Following the same argument as in the proof of (ii), we obtain the desired results.

The proof of (iv) is exactly the same as above and thus omitted.

**Proof of Theorem 2.1.** We first show that the root of (2.6) satisfies $\|\lambda\| = O_P(n^{-1/2})$ when $\beta$ is replaced by its true value $\beta_0$. Similar to [23, p. 101], we let $\lambda = \rho \theta$ where $\rho \geq 0$ and $\|\theta\| = 1$. Then

$$
0 = g(\rho \theta) = |\theta^T g(\rho \theta)|
= \frac{1}{n} \left\| \theta^T \left\{ \sum_{i=1}^n \hat{W}_i(\beta_0) - \rho \sum_{i=1}^n \hat{W}_i(\beta_0) \theta^T \hat{W}_i(\beta_0) \right\} \right\|
\geq \frac{\rho}{n} \left\| \theta^T \left\{ \sum_{i=1}^n \hat{W}_i(\beta_0) \hat{W}_i(\beta_0)^T \right\} \right\|
- \frac{1}{n} \left\| \theta^T \sum_{i=1}^n \hat{W}_i(\beta_0) \right\|.
$$

Note that the second term on the right hand side above is $O_P(n^{-1/2})$ by Lemma A.1 and the approximation adopted in Appendix C of [40]. For simplifying the first term, we need the fact that $1 + \rho \theta^T \hat{W}_i(\beta_0) > 0$, which is true because $p_i > 0$ for all $i$. Moreover, since $\|W_i(\beta_0)\|$ are uniformly bounded by, say, $M$, for all $i$, we have $1 + \rho \theta^T \hat{W}_i(\beta_0) \leq 1 + \rho M$ for all $i$. Thus it follows that

$$
\frac{\rho \theta^T \hat{\Sigma}_n \theta}{1 + \rho M} = O_P(n^{-1/2}).
$$

Because $\|\theta\| = 1$, $\hat{\Sigma}_n = \Sigma + o_P(1)$ and $\Sigma$ is assumed to be positive definite, we have $\theta^T \hat{\Sigma}_n \theta \geq \sigma_p / 4 + o_P(1)$ where $\sigma_p > 0$ is the smallest eigenvalue of $\Sigma$. Therefore, we have $\|\lambda\| = \rho = O_P(n^{-1/2})$.

Next we derive an asymptotic expression for $\lambda$ as the root of (2.6) when $\beta$ is replaced by its true value $\beta_0$. Reorganizing the summand in (2.6) and we have

$$
0 = g(\lambda) = \frac{1}{n} \sum_{i=1}^n \hat{W}_i(\beta_0) \left\{ 1 - \lambda^T \hat{W}_i(\beta_0) + \frac{\lambda^T \hat{W}_i(\beta_0) \hat{W}_i(\beta_0)^T \lambda}{(1 + \lambda^T \hat{W}_i(\beta_0))} \right\}
= \hat{Q}(\beta_0) - \hat{\Sigma}_n \lambda + \frac{\sum_{i=1}^n \hat{W}_i(\beta_0) \lambda^T \hat{W}_i(\beta_0) \hat{W}_i(\beta_0)^T \lambda}{1 + \lambda^T \hat{W}_i(\beta_0)}.
In the above, the bound of the last term on the right hand side is $O_P(n^{-1/2})$. Since $\|\hat{W}_t(\beta_0)\|$ is uniformly bounded and $\|\lambda\| = O_P(n^{-1/2})$. Therefore, we may write
\[ \lambda = \hat{\Sigma}_n^{-1} \hat{Q}(\beta_0) + O_P(n^{-1}). \] (A.3)

Note that in the above, $\hat{\Sigma}_n$ is invertible because $\hat{\Sigma}_n = \Sigma + o_P(1)$ and $\Sigma$ is assumed to be positive definite.

Finally we expand $I(\beta_0)/4$ as follows:
\[
\frac{1}{4} I(\beta_0) = \frac{1}{2} \sum_{i=1}^{n} \log(1 + \lambda^T \hat{W}_t(\beta_0))
= \frac{1}{2} \sum_{i=1}^{n} \{\lambda^T \hat{W}_t(\beta_0) - \lambda^T \hat{W}_t(\beta_0) \hat{W}_t(\beta_0)^T \lambda/2 + O(\lambda^T \hat{W}_t(\beta_0)^2)\}
= \frac{n}{2} \hat{Q}(\beta_0)^T \hat{\Sigma}_n^{-1} \hat{Q}(\beta_0) - \frac{n}{4} \hat{Q}(\beta_0)^T \hat{\Sigma}_n^{-1} \hat{\Sigma}_n \hat{\Sigma}_n^{-1} \hat{Q}(\beta_0) + O_P(n^{-1/2})
= \left\{ n^{-3/2} U(\beta_0)^T \right\} \hat{\Sigma}_n^{-1} \left\{ n^{-3/2} U(\beta_0) \right\} + o_P(1).
\]

Using again the approximation adopted in Appendix C of [40], i.e., $n^{-3/2} U(\beta_0)$ asymptotically follows a normal distribution with mean 0 and variance $\Sigma$, and the fact that $\hat{\Sigma}_n \to \Sigma$, we see that Theorem 2.1 follows. □

**Proof of Theorem 2.2.** Corresponding to $\beta = (\beta(1)^T, \beta(2)^T)^T$, write $Z = (Z(1)^T, Z(2)^T)^T$. Let $\check{\beta}(2) = \arg \inf_{\beta(2) \in \mathcal{R}^{p-q}} I((\beta_0(1)^T, \beta(2)^T)^T)$. Equivalently, $\check{\beta}(2)$ maximizes $\prod_{i=1}^{n} p_i$, where
\[
p_i \geq 0, \quad \sum_{i=1}^{n} p_i = 1, \quad \sum_{i=1}^{n} p_i \hat{W}_i \left( (\beta_0(1)^T, \beta(2)^T)^T \right) = 0. \quad (A.4)
\]

Note that in (A.4) $\hat{W}_i$’s are vectors of dimension $p$, while only $p - q$ parameters $(\beta(2))$ are unknown. Therefore, using exactly the same arguments as in Lemma 1 and Theorem 1 of [26] (the conditions can be easily verified), we can show that $\check{\beta}(2)$ exists and is unique. Furthermore, we have
\[
\sqrt{n}(\check{\beta}(2) - \beta_0(2)) = -(A_2^T \Sigma^{-1} A_2)^{-1} A_2^T \Sigma^{-1/2} \sum_{i=1}^{n} \hat{W}_i(\beta_0)/2 + o_P(1),
\]
\[
\sqrt{n} \tilde{\lambda}_2 = (I - \Sigma^{-1} A_2(A_2^T \Sigma^{-1} A_2)^{-1} A_2^T) \Sigma^{-1/2} \sum_{i=1}^{n} \hat{W}_i(\beta_0) + o_P(1)
\]
\[
:= S_n^{-1/2} \sum_{i=1}^{n} \hat{W}_i(\beta_0) + o_P(1),
\]

where $A_2 = \lim_{n \to \infty} \sum_{i=1}^{n} \sum_{j=1}^{n} \tau_i (Z_{ij}(\beta_0))^2 Z_{ij}(\beta_2^{(2)})^2 / n^2$ and $\tilde{\lambda}_2$ is the Lagrange multiplier corresponding to $\beta(2) = \check{\beta}(2)$. Note that here we have a slightly different situation than [26] in that $\hat{W}_i$’s are not independent of each other. However, the arguments in Lemmas A.1 and A.2 guarantee the convergence rate of $\sum_{i=1}^{n} \hat{W}_i(\beta_0)$ and boundedness of $\hat{W}_i(\beta_0)$.

Since $A$ is positive definite, $A_2$ is a $p \times (p - q)$ dimensional matrix of full rank. Using the approximation adopted in Appendix C of [40], we see that $\sqrt{n}(\check{\beta}(2) - \beta_0(2))$ has an asymptotic normal distribution with mean 0 and variance $[A_2^T \Sigma^{-1} A_2]^{-1}$, and $\tilde{\lambda}_2 = O_P(n^{-1/2})$. To simplify the notation we now write $\left( (\beta_0(1)^T, \beta(2)^T)^T \right)$ as $\beta_0$. Let $\beta^*$ be on the line segment between $\check{\beta}_0$ and $\beta_0$, we expand
\[
\sum_{i=1}^{n} \hat{W}_i(\beta_0) = n^{-1/2} \left\{ S_n^{-1/2} \sum_{i=1}^{n} \hat{W}_i(\beta_0) + o_P(1) \right\} T \left\{ \sum_{i=1}^{n} \hat{W}_i(\beta_0) + \sum_{i=1}^{n} \frac{\partial \hat{W}_i(\beta^*)}{\partial \beta} (\check{\beta}_2 - \beta_0) \right\}
= \left[ \sum_{i=1}^{n} \hat{W}_i(\beta_0)^T / \sqrt{n} \right] S \left[ \sum_{i=1}^{n} \hat{W}_i(\beta_0) / \sqrt{n} \right]
+ \left[ \sum_{i=1}^{n} \hat{W}_i(\beta_0)^T / \sqrt{n} \right] S (2A_2) \left[ \sqrt{n}(\check{\beta}_2 - \beta_0) \right] + o_P(1)
= \left[ \sum_{i=1}^{n} \hat{W}_i(\beta_0)^T / \sqrt{n} \right] S \left[ \sum_{i=1}^{n} \hat{W}_i(\beta_0) / \sqrt{n} \right] + o_P(1)
\]
because $S A_2 = 0$. Similarly we have

$$
\hat{\lambda}_2^T \sum_{i=1}^n \hat{W}_i(\beta_0) \hat{W}_i(\beta_0)^T \hat{\lambda}_2 = \left\{ \sum_{i=1}^n \hat{W}_i(\beta_0)^T / \sqrt{n} \right\} S^T \Sigma S \left\{ \sum_{i=1}^n \hat{W}_i(\beta_0) / \sqrt{n} \right\} + o_p(1)
$$

$$
= \left\{ \sum_{i=1}^n \hat{W}_i(\beta_0)^T / \sqrt{n} \right\} S \left\{ \sum_{i=1}^n \hat{W}_i(\beta_0) / \sqrt{n} \right\} + o_p(1).
$$

Therefore, using a similar Taylor expansion as in the proof of Theorem 2.1, we have that

$$
\frac{1}{4} \hat{l}(\beta_0^{(1)}) = \frac{1}{2} \sum_{i=1}^n \log \left\{ 1 + \hat{\lambda}_2^T \hat{W}_i(\beta_0) \right\}
$$

$$
= \frac{1}{4} \left\{ \sum_{i=1}^n \hat{W}_i(\beta_0)^T / \sqrt{n} \right\} S \left\{ \sum_{i=1}^n \hat{W}_i(\beta_0) / \sqrt{n} \right\} + o_p(1)
$$

$$
= \left\{ n^{-3/2} U(\beta_0)^T \Sigma^{-1/2} \right\} \left\{ I - \Sigma^{-1/2} A_2 (A_2^T \Sigma^{-1} A_2)^{-1} A_2^T \Sigma^{-1/2} \right\} \left\{ \Sigma^{-1/2} n^{-3/2} U(\beta_0) \right\} + o_p(1).
$$

Using again the approximation adopted in Appendix C of [40], i.e., $n^{-3/2} U(\beta_0)$ asymptotically follows a normal distribution with mean 0 and variance $\Sigma$, and the fact that $I - \Sigma^{-1/2} A_2 (A_2^T \Sigma^{-1} A_2)^{-1} A_2^T \Sigma^{-1/2}$ is a symmetric and idempotent matrix with trace $q$, we have the desired result. □

**Proof of Theorem 2.3.** The proof of Theorem 2.3 is similar to that of Theorem 2.1. Thus we will only provide a sketch. First, straightforward algebraic simplification (cf. Lemma A.3 of [15]) shows that

$$
\sum_{i=1}^n V_i(\beta_0)^T V_i(\beta_0) / n = 4 \Sigma + o_p(1).
$$

and (ii) in Lemma A.2 can be verified in the same way with $W_i(\beta_0)$ (respectively) replaced by $V_i(\beta_0)$ (respectively). Next, we can show that the root of $\hat{g}_i(\lambda) = 0$ is $o_p(n^{-1/2})$ and has an asymptotic expression as $\Sigma^{-1/2} Q(\beta_0) / 4 + o_p(n^{-1})$. Finally, we expand $\hat{l}(\beta_0)$ in the same way as $l(\beta_0)$ to yield the desired result. □

**Proof of Theorem 2.4.** The proof of Theorem 2.4 is similar to that of Theorem 2.2 and is thus omitted. □

**References**


