Semiparametric inference for transformation models via empirical likelihood

Yichuan Zhao

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

**Abstract**

Recent advances in the transformation model have made it possible to use this model for analyzing a variety of censored survival data. For inference on the regression parameters, there are semiparametric procedures based on the normal approximation. However, the accuracy of such procedures can be quite low when the censoring rate is heavy. In this paper, we apply an empirical likelihood ratio method and derive its limiting distribution via \( U \)-statistics. We obtain confidence regions for the regression parameters and compare the proposed method with the normal approximation based method in terms of coverage probability. The simulation results demonstrate that the proposed empirical likelihood method overcomes the under-coverage problem substantially and outperforms the normal approximation based method. The proposed method is illustrated with a real data example. Finally, our method can be applied to general \( U \)-statistic type estimating equations.

**1. Introduction**

It is well known that the Cox [1] regression model is the most popular model used in survival analysis. The Cox model is semi-parametric, and its large sample inference properties have been demonstrated using martingale theory [2]. Moreover, practitioners have easy access to statistical software for this model. In practice, however, the proportional hazards assumption is often too restrictive, even for randomized clinical trials. In recent years, the transformation model has received a lot attention and provides a useful alternative to the Cox regression model in analyzing survival observations. Its simple structure and ease of interpretation make it an attractive method. The transformation model is becoming a valuable model for the analysis of survival data.

Let \( T \) be the failure time, i.e. the response variable, and \( Z \) a corresponding covariate vector. Suppose that we are interested in making inferences about the effect of \( Z \) on the response variable. If there are censored observations in the data, one usually uses the Cox model to examine the covariate effect. Let \( S_Z(t) \) be the survival function of \( T \) given \( Z \). Suppose that \( h(t) \) is a completely unspecified strictly increasing function, which maps the positive half-line onto the whole real line. Thus, a natural generalization of the Cox regression model is

\[
g[S_Z(t)] = h(t) + Z^T \beta,
\]

where \( g(\cdot) \) is a known decreasing function and \( \beta \) is a \( p \times 1 \) vector of unknown regression coefficients.

The model (1.1) includes the Cox model with \( g(x) = \log(-\log x) \), and the proportional odds model with \( g(x) = \logit(x) = \log(x/(1-x)) \) [3–6] as special cases. It is easy to see that (1.2) is equivalent to the linear transformation model:

\[
h(T) = -Z^T \beta + \varepsilon,
\]

where \( \varepsilon \) is a random error with distribution function \( F = 1 - g^{-1} \) and \( \varepsilon \) is independent of the covariate \( Z \).
For the Cox model, $F$ is the standard extreme value distribution, i.e. $F(t) = 1 - \exp(-\exp(t))$. If $F$ is the standard logistic distribution, then (1.2) is the proportional odds model. Inference procedures for $\beta$ under model (1.2) have been proposed by, for example, [7–12], among others. Cheng et al. [10] proposed and justified a general estimation method for linear transformation models with censored data using inverse-censoring-probability-weighted estimating equations. The method was further developed in [13–16], among others. A key step in their approach is the estimation of the survival function for the censoring variable by the Kaplan–Meier estimator. However, the accuracy of such a procedure could be low when the censoring proportion is high and needs to be improved. An appealing technique is the empirical likelihood method (EL). Furthermore, [17] studied EL inference for semiparametric linear transformation models based on martingale based estimating equations proposed by Chen et al. [12].

The EL is a nonparametric approach for constructing confidence regions, which was introduced by Owen [18,19] for the mean of a random vector based on i.i.d. complete data. Since then, the EL has been widely applied in different statistical areas to make inference. Some related work includes linear models [20], general estimating equations [21], confidence bands with right censoring [22–26] and the general plug-in EL [27], among others. Like the bootstrap and the jackknife, the EL method does not need to specify a family of distributions for the data. Furthermore, it holds some unique features, such as range-respecting, transformation-preserving, asymmetric confidence interval, etc. In recent years, the method has received much attention in the literature because of its excellent and well recognized small sample properties in terms of coverage probability.

In this paper, we apply an EL ratio method and derive its limiting distribution. However, the regular EL approach including general plug-in EL proposed by Hjort et al. [27] is not applicable for the $U$-statistic type estimating equation under the transformation model (cf. [10]). To overcome this difficulty, we adopt an empirical likelihood method based on pseudo observations proposed by Jing et al. [28]. The key idea is to turn it into a sample mean based on some pseudo observations. We derive the limiting distribution of the EL ratio, and find EL based confidence regions for the regression parameter. The simulation study demonstrates the proposed method outperforms the existing normal approximation method in terms of coverage probability. Furthermore, the main contribution of this paper is that the proposed method is used not only in the special semiparametric transformation model, but also in more general $U$-statistics type estimating equations.

The paper is organized as follows. In Section 2, we introduce a simple empirical likelihood method for regression parameter $\beta$. The proposed confidence region and main asymptotic result are presented in Section 2. In Section 3, we conduct a simulation study to compare the proposed method with the normal approximation based method and Lu and Liang’s method. A real data example is used to illustrate the EL method in Section 4. The conclusion is made in Section 5. Proofs are given in the Appendix.

2. Main results

2.1. Preliminaries

In this section, we introduce basic notations and known results given by Cheng et al. [10]. Let $T_i$ be the failure time for the $i$th patient ($i = 1, \ldots, n$). For $T_i$, one can only observe a bivariate vector $(X_i, \delta_i)$, where $X_i = \min(T_i, C_i)$ and $\delta_i = I(T_i < C_i)$. The censoring variable $C_i$ is assumed to be independent of $T_i$. Let a $p \times 1$ vector $Z_i$ be the corresponding covariate vector. Here as [10], we assume that the censoring variables $C_i$ are i.i.d. with the same survival function $G = P(C_i > t)$. In addition, the $Z_i$ and $C_i$ are independent. Denote $Z_{ij} = Z_i - Z_j$. Note that the error $\epsilon$ is independent of the covariate $Z$. Cheng et al. [10] proved that

$$E \left\{ \frac{\delta_i I(X_i \geq X_j)}{G^2(X_j)} \left| Z_i, Z_j \right. \right\} = \xi(Z_{ij}^T \beta_0), \quad (2.1)$$

where $\beta_0$ is the true value of $\beta$ and

$$\xi(s) = \int_{-\infty}^{\infty} \{1 - F(t + s)\}dF(t), \quad (2.2)$$

where $F$ is the completely specified distribution function of the error $\epsilon$. They proposed the following estimation equation:

$$U(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{n} w(Z_{ij}^T \beta) Z_{ij} \left\{ \frac{\delta_i I(X_i \geq X_j)}{G^2(X_j)} - \xi(Z_{ij}^T \beta) \right\}, \quad (2.3)$$

where $w(t)$ is a weights function such as $w(t) = 1$ which is similar to the usual linear regression and $w(t) = -\xi(t)/[\xi(t)(1 - \xi(t))]$ which mimics the quasi-likelihood approach for independent observations and $\hat{G}(t)$ is the Kaplan–Meier estimator of the survival function $G(t)$ of the censoring variable $C_i$. When the censoring variable depends on the covariate vector $Z$ and $Z$ takes only finitely many values, the alternative estimating equation is proposed similarly. This assumption of independence between the censoring time $C_i$ and the covariates $Z_i$ is strong. It can be weakened when the covariate is continuous. In fact, in that case it suffices to replace $G(t)$ by the conditional survival function of $C_i$ given $Z_i$, which can be estimated using smoothing techniques, see p. 838 of [10] and p. 338 of [29] for details.
Here, we take advantage of the estimating equation that the censoring variable depends on the covariate which takes finitely many values we can extend the result similarly.

\[
\Gamma_2 = 4 \int_0^\infty \frac{q(t)q^T(t)}{\pi(t)} \, d\Lambda_C(t),
\]

where

\[
e_0(\beta_0) = \frac{\delta I(X_i \geq X_j)}{G^2(X_i)} - \xi(Z_{ij}^T \beta_0),
\]

\[
\pi(t) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^n I(X_i \geq t) = P(X_i \geq t) \quad \text{in probability},
\]

\[
q(t) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n w(Z_{ij}^T \beta_0)Z_{ij} \frac{\delta I(X_i \geq X_j)}{G^2(X_i)} I(X_j \geq t) \quad \text{in probability},
\]

and \(\Lambda_C(\cdot)\) is the common cumulative hazard function of \(C_i\)'s. Denote \(\Gamma = \Gamma_1 - \Gamma_2\) and let

\[
\Lambda^{-1} = \lim_{n \to \infty} \left\{ n^{-2} \sum_{i=1}^n \sum_{j=1}^n w(Z_{ij}^T \beta_0) \xi'(Z_{ij}^T \beta_0)Z_{ij}^{\otimes 2} \right\} \quad \text{in probability},
\]

and \(v^{\otimes 2} = vv^T\) for a vector \(v\).

Cheng et al. [10] showed in Appendix 1 if the weights \(w(\cdot)\) are positive, then the equation \(U(\beta) = 0\) has, asymptotically, a unique solution \(\hat{\beta}\). Under certain conditions,

\[
n^{1/2}(\hat{\beta} - \beta_0) \overset{D}{\to} N(0, \Lambda \Gamma \Lambda). \quad (2.4)
\]

Let

\[
\hat{\Gamma}_1 = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1, k \neq j}^n \{ w(Z_{ij}^T \hat{\beta})\hat{e}_j(\beta) - w(Z_{ik}^T \hat{\beta})\hat{e}_k(\beta) \} \{ w(Z_{ik}^T \hat{\beta})\hat{e}_k(\beta) - w(Z_{ji}^T \hat{\beta})\hat{e}_i(\beta) \} Z_{ij}Z_{ik}^{\top},
\]

\[
\hat{\Gamma}_2 = 4 \frac{n^3}{n^2} \sum_{i=1}^n \left( \frac{1 - \delta_i}{\sum_{k=1}^n I(X_k \geq X_i)} \right)^2 \left\{ n^{-2} \sum_{i=1}^n \sum_{j=1}^n w(Z_{ij}^T \hat{\beta})Z_{ij} \frac{\delta I(X_i \geq X_j)}{G^2(X_i)} I(X_j \geq X_i) \right\}^{\otimes 2},
\]

\[
\hat{\xi}_0(\hat{\beta}) = \delta I(X_i \geq X_j)[G^2(X_i)]^{-1} - \xi(Z_{ij}^T \hat{\beta}).
\]

\[
\hat{\Lambda}^{-1} = n^{-2} \sum_{i=1}^n \sum_{j=1}^n w(Z_{ij}^T \hat{\beta})\xi'(Z_{ij}^T \hat{\beta})Z_{ij}^{\otimes 2}. \quad (2.5)
\]

Denote \(\hat{\Gamma} = \hat{\Gamma}_1 - \hat{\Gamma}_2\). From Lemma A.3 \(\Gamma\) is consistently estimated by \(\hat{\Gamma}\) (cf. [10]). Similarly as Lemma A.3 we can show that \(\Lambda\) is consistently estimated by \(\hat{\Lambda}\). Thus an asymptotic 100(1 - \(\alpha\))% confidence region for \(\beta\) based on the above normal approximation by Cheng et al. [10] 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 \(p\).

### 2.2. EL confidence region

In order to overcome the under-coverage problem for the normal approximation method proposed by Cheng et al. [10], we adopt the EL approach. We study the case where the censoring variable is independent of the covariate. When the censoring variable depends on the covariate which takes finitely many values we can extend the result similarly. Here, we take advantage of the estimating equation (2.3). Let \(U_i = (Z_i, X_i, \delta_i)\). Denote \(b(U_i, U_j; \beta) = \{ w(Z_{ij}^T \beta)Z_{ij}e_j(\beta) + w(Z_{ji}^T \beta)Z_{ji}e_i(\beta) \} \) and \(b_n(U_i, U_j; \beta) = \{ w(Z_{ij}^T \beta)Z_{ij}\hat{e}_j(\beta) + w(Z_{ji}^T \beta)Z_{ji}\hat{e}_i(\beta) \} \). Applying the idea of [30], we define

\[
W_i(\beta) = \frac{1}{n - 1} \sum_{j=1, j \neq i}^n \{ b(U_i, U_j; \beta) \},
\]

\[
W_m(\beta) = \frac{1}{n - 1} \sum_{j=1, j \neq i}^n \{ b_n(U_i, U_j; \beta) \}.
\]
under the above conditions is given in the Appendix and likelihood ratio, \( r \), the eigenvalues of Lemmas A.2 are estimated so that the above theorem can be used in parameter inference. We define \( \hat{\Sigma} \) as the sum of i.i.d. \( \Sigma \chi_{\alpha_i} \) for our main result and explain how it can be used to construct the confidence region for C.3. Both the matrix \( \Lambda \) and the matrix \( \Gamma \) are positive definite.

Recall that \( \beta_0 \) is the true value of \( \beta \). Throughout the paper, we define \( \Sigma = E[b(U_1, U_2; \beta_0) b^T(U_1, U_3; \beta_0)] \). Now we state our main result and explain how it can be used to construct the confidence region for \( \beta \).

**Theorem 2.1.** Under the above conditions 1–3, the EL statistic \( \hat{l}(\beta_0)/4 \) converges in distribution to \( r_1 \chi^2_{1,1} + \cdots + r_p \chi^2_{1,p} \), where \( \chi^2_{1,1}, \ldots, \chi^2_{1,p} \) are independent chi-square random variables with 1 degree of freedom and \( r_1, \ldots, r_p \) are the eigenvalues of \( \Sigma^{-1} \Gamma \).

The proof of Theorem 2.1 is given in the Appendix. We note that the limiting distribution of the EL ratio is a weighted sum of i.i.d. \( \chi^2_{1} \) instead of the standard \( \chi^2_{p} \) distribution. This is due to the fact that \( W_{ni} \) is dependent. A similar phenomenon occurs in various contexts, such as [31–35], among others.

Although the limiting distribution has nonstandard weighted sum expression, the weights involved can be readily estimated so that the above theorem can be used in parameter inference. We define \( \hat{\Sigma} = 1/n \sum_{i=1}^{n} W_{ni}(\beta) W_{ni}^T(\beta) \). From Lemmas A.2 and A.3 (ii), \( \Sigma \) is consistently estimated by \( \hat{\Sigma} \). Hence, the values of \( r_i \) can be estimated by those of \( \hat{r}_i \) which are the eigenvalues of \( \hat{\Sigma}^{-1} \Gamma \). An asymptotic 100(1 – \( \alpha \))% empirical likelihood (EL) confidence region for \( \beta \) is given by

\[
\{ \beta : \hat{l}(\beta)/4 \leq c_1(\alpha) \},
\]

(2.10)

where \( c_1(\alpha) \) is the upper \( \alpha \)-quantile of the distribution of \( \hat{r}_1 \chi^2_{1,1} + \cdots + \hat{r}_p \chi^2_{1,p} \) and can be obtained by simulation method.

Alternatively, the above EL approach can be adjusted to avoid the weighted sum expression. Let \( \rho(\beta) = p/\text{tr}(\Sigma^{-1}(\beta) \Gamma(\beta)) \) with \( \text{tr}(\cdot) \) denoting the trace of a matrix. Then, following [36], the distribution of \( \rho(\beta)(r_1 \chi^2_{1,1} + \cdots + r_p \chi^2_{1,p}) \) may be approximated by \( \chi^2_{p} \). This implies that the asymptotic distribution of the Rao–Scott adjusted empirical likelihood ratio, \( \hat{l}_{ad}(\beta)/4 = \hat{\rho}(\beta) \hat{l}(\beta)/4 \), may be approximated by \( \chi^2_{p} \), where the adjustment factor \( \hat{\rho}(\beta) \) is \( \rho(\beta) \) with \( \Sigma(\beta) \) and \( \Gamma(\beta) \) replaced by \( \hat{\Sigma}(\beta) \) and \( \hat{\Gamma}(\beta) \), respectively.
The adjusted EL approach was proposed by Wang and Rao [31], among others. We define an adjusted empirical likelihood ratio, by modifying \( \hat{\rho}(\beta) \) in \( \tilde{l}_{\text{ad}}(\beta)/4 \). Let \( \hat{S}(\beta) = \{ \sum_{i=1}^{n} W_{ni}(\beta)/n \} / \{ \sum_{i=1}^{n} W_{i}(\beta)/n \} \), and
\[
\hat{r}(\beta) = \frac{\text{tr} \left( \hat{\Sigma}^{-1}(\beta) \tilde{S}(\beta) \right)}{\text{tr} \left( \hat{\Sigma}^{-1}(\beta) \tilde{S}(\beta) \right)}.
\]
We define an adjusted empirical likelihood ratio by
\[
\tilde{l}_{\text{ad}}(\beta) = \hat{r}(\beta) \tilde{l}(\beta).
\]

**Theorem 2.2.** Under the above conditions 1–3, the EL statistic \( \tilde{l}_{\text{ad}}(\beta_0)/4 \) converges in distribution to \( \chi^2_p \).

Based on Theorem 2.2, an asymptotic 100(1 − α)% adjusted empirical likelihood (AEL) confidence region for \( \beta \) is given by
\[
R_3 = \left\{ \beta : \tilde{l}_{\text{ad}}(\beta)/4 \leq \chi^2_p(\alpha) \right\},
\]
where \( \chi^2_p(\alpha) \) is defined as before. The adjusted factor \( \hat{r}(\beta) \) involves \( \beta \). An updated \( \beta \) at each step is used instead of a fixed \( \hat{\beta} \) in the process of profile analysis for finding the confidence region.

3. **Simulation study**

An extensive simulation is conducted to compare the performance of the empirical likelihood procedure with the normal approximation based procedure (NA) and Lu and Liang’s method [17]. The NA is based on (2.6). The EL is based on (2.10) and the AEL is based on (2.11). We will compare the proposed EL approaches with NA based method and Lu and Liang’s method (LL) in terms of coverage probability in different settings. A similar set up as that in [10] is considered, i.e. one model corresponding to the proportional hazards model. As discussed in Section 2.1, the estimating equations based on inverse probability weighting technique require i.i.d. censoring assumption. Thus, both the NA and EL methods require that independence of \( C_i \) and \( T_i \), which may be restrictive in practice. Practitioners may be interested in the robustness of these methods against departure from the assumption. Hence, we also conduct some sensitivity analysis to evaluate the performance of the proposed method when the assumption is violated.

In the simulation study, the first model is a proportional hazards model with two independent covariates, the first one from a uniform variable on \([0, 1]\), and the second from a Bernoulli variable with success probability 0.5. The survival time is obtained with \( h \) the natural logarithm function and \( \epsilon \) having the standard extreme value distribution, and the censoring time is Uniform \([0, c]\), where \( c \) controls the censoring rate. Corresponding to \( \beta_0 = (0, 0)^T \), the censoring rates are approximately 10%, 20%, 30%, and 40% respectively, which represent light censoring, middle censoring, heavy censoring and very heavy censoring rates, respectively. The sample size is set to be 40, 60, 80, and 100, representing very small, relatively small, moderate, and large samples, respectively. The simulation results are reported in Table 1. Each entry of the table is based on 1000 simulated data sets.

Realistic situations usually deal with large values of \( p \). The second model is a proportional hazards model with three independent covariates, the first one from a uniform variable on \([0, 1]\), the second from a Bernoulli variable with success probability 0.5, and the third from a Bernoulli variable with success probability 0.7. The survival time is obtained with \( h \) the natural logarithm function and \( \epsilon \) having the standard extreme value distribution, and the censoring time is uniform \([0, c]\),
Table 2
Coverage probabilities for the regression parameter under model 2.

<table>
<thead>
<tr>
<th>CR</th>
<th>n</th>
<th>1 - α = 0.90</th>
<th>1 - α = 0.95</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NA</td>
<td>EL</td>
<td>AEL</td>
</tr>
<tr>
<td>10%</td>
<td>40</td>
<td>.796</td>
<td>.798</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>.840</td>
<td>.864</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>.874</td>
<td>.891</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>.873</td>
<td>.889</td>
</tr>
<tr>
<td>30%</td>
<td>40</td>
<td>.752</td>
<td>.744</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>.815</td>
<td>.818</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>.846</td>
<td>.855</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>.868</td>
<td>.874</td>
</tr>
</tbody>
</table>

Table 3
Coverage probabilities for the regression parameter under model 3.

<table>
<thead>
<tr>
<th>CR</th>
<th>n</th>
<th>1 - α = 0.90</th>
<th>1 - α = 0.95</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NA</td>
<td>EL</td>
<td>AEL</td>
</tr>
<tr>
<td>10%</td>
<td>40</td>
<td>.846</td>
<td>.848</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>.867</td>
<td>.881</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>.878</td>
<td>.895</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>.890</td>
<td>.899</td>
</tr>
<tr>
<td>30%</td>
<td>40</td>
<td>.852</td>
<td>.854</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>.850</td>
<td>.863</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>.872</td>
<td>.881</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>.891</td>
<td>.897</td>
</tr>
</tbody>
</table>

where \(c\) controls the censoring rate. Corresponding to \(\beta_0 = (0, 0, 0)^T\), the censoring rates are approximately 10% and 30% respectively. The sample size is set to be 40, 60, 80, and 100. The simulation results are reported in Table 2. Each entry of the table is based on 1,000 simulated data sets.

The third model is a proportional hazards model with two dimensional covariates \(Z = (Z_1, Z_2)^T\), the first one from \(Z_1 = U[0, 1]\), and the second from \(Z_2 = \text{Bernoulli}(p = 0.5)\). The survival time is obtained with \(h\) the natural logarithm function and \(e\) having the standard extreme valued distribution, and the censoring time is \(U[0, c] + Z_1 + Z_2\), where \(c\) controls the censoring rate. Corresponding to \(\beta_0 = (0, 0)^T\), the censoring rates are approximately 10% and 30% respectively. The sample size is set to be 40, 60, 80, and 100. The simulation results are reported in Table 3. Each entry of the table is based on 1,000 simulated data sets.

As shown in the Table 1, all the methods work reasonably well with right coverage probabilities of 90%, 95% when the sample size is large and the censoring rate is not heavy. The LL, the AEL, the EL and the NA based coverage probabilities tend to achieve the nominal levels with large sample sizes \((n = 100)\) when the censoring rate is light, while LL and AEL methods work well under censoring rate 10%, and 20%, respectively even for moderate sample. For very heavy censoring rate 40% and very small sample size \(n = 40\), LL is better than AEL and it demonstrates its efficiency compared to AEL. It also shows that Lu and Liang's method based on martingale based estimating equations is generally more efficient than the EL method based on inverse-censoring-probability-weighted estimating equations for very heavy censoring rate. For \(\beta_0 = (1, 1)^T\), simulation studies lead to very similar results, thus they are not displayed.

From Table 2, the simulation results show that the proposed AEL procedure performs well even for heavy censoring rate 30% and sample size \(n = 100\) except for small sample size when the number of parameters increases to three from two. Thus, the proposed AEL works well for \(p = 3\). Usually the accuracy of coverage probability will decrease when the number of parameters increases. The LL method has a similar trend as AEL did.

From Table 3, we find there is a similar pattern for the simulation results as Table 1. AEL and LL have good performance in terms of coverage accuracy. The performance of the proposed method is still good when the i.i.d. censoring assumption is violated. There is no evidence that this dependence assumption has a negative impact on the coverage accuracy of the proposed AEL method. This suggests that the method is robust against this type of departure from the independence assumption.

From Tables 1–3, we find that accuracy of coverage probabilities decreases as the censoring rate increases. At each nominal confidence level, the accuracy of coverage probabilities for four methods increases as the sample size \(n\) increases. But for very small sample size \((n = 40)\), the NA based method apparently has relatively larger under-coverage. In terms of statistical testing, it means that the type-I error is out of control and is larger than the required \(\alpha = 5\%\) or 10%. While the proposed EL confidence region has a slightly better coverage accuracy for nominal level 90%, 95% and the proposed AEL confidence region has a better coverage accuracy. The reason for this is that the NA confidence interval needs to estimate the variance and the estimates may fall outside the range for a small sample or heavy censoring rate. From Tables 1–3 we see that the coverage probability of the AEL confidence region \(\mathcal{N}_3\) based on (2.11) outperforms the NA and EL confidence regions. LL has the best performance when the censoring rate is very heavy and sample size is very small.
4. Application

In this section, we use multiple myeloma data to illustrate the proposed empirical likelihood methods and compare them with the normal approximation method. The data set is presented in the SAS/STAT Users guide (1999, pp. 2608–2617). The data come from a study on multiple myeloma in which researchers treated 65 patients with alkylating agents. Of those patients, 48 died during the study and 17 survived. The censoring rate is about 26%. For illustration, one covariate vector $Z = (Z_1, Z_2)^T$, consisting of $Z_1$: the logarithm of blood urea nitrogen, log(BUN) and $Z_2$: HGB (hemoglobin at diagnosis) is considered. We fit the data set by the proportional hazards model. The estimate of regression parameter $\beta$ is $(1.4079, -0.0983)^T$.

To further investigate the properties of the confidence regions proposed in Section 2, we make three contour plots simultaneously, each plot containing only loops from confidence regions $R_1$, $R_2$ and $R_3$ for the regression parameter $\beta$. In Fig. 1 we report the point estimate of $\beta$, 90% NA, EL and AEL confidence regions for $\beta$. From this, we see the EL and AEL confidence regions are almost overlapped, and the empirical likelihoods produce more similar or comparable confidence regions than normal approximation confidence regions. We note that the normal approximation confidence region has the symmetry property which is not desirable since the distribution of the parameter estimator may be skewed. The empirical likelihood confidence regions are not exactly symmetric about the point estimator, and the empirical likelihood method is able to pick up possible skewness in contrast to the normal approximation method.

5. Conclusion

In this paper, based on the estimating equation proposed by Cheng et al. [10], we have applied an empirical likelihood ratio method to the semiparametric transformation model with right censored data and derived the limiting distribution of the empirical likelihood ratio. We have proposed the unadjusted and adjusted empirical likelihood confidence regions for the unknown vector of regression parameters. One advantage of the adjusted empirical likelihood method is that it does not need simulation to obtain critical values compared to the unadjusted empirical likelihood method. Using the multiple myeloma data, we have illustrated how to implement our proposed method into real data analysis. The simulation results show that our proposed empirical likelihood methods perform well in terms of coverage probability. From Tables 1–3, we also find that the normal approximation based method does not always work well and has under-coverage problems for small samples. One reason is that the normal approximation based confidence region needs to estimate $\Gamma$ and $\Lambda$. The variance estimates are not very stable and may contain values outside their ranges. However, the proposed adjusted empirical likelihood confidence region $R_3$ holds superior properties. It is a competitive method which outperforms the NA method and can overcome the under-coverage probability problem for small sample size. Furthermore, it has the best coverage probability and the comparable area of confidence region. Thus in practice, we recommend the adjusted empirical likelihood method for transformation models with right censored data.

However, there is a kind of trade off between coverage accuracy and computation time. Indeed, EL techniques rely on the computation of Lagrange multiplier as well as on matrix inversion, which can be time consuming for typical high-dimensional covariates. In our simulation study, we consider the number of parameters $p = 2$ or 3. For a very high-dimensional case such as $p \gg n$, the proposed EL method does not work since it depends on the original estimation Eq. (2.3). Currently, EL with high-dimensional data is still being developed. There has been few research in this field such as [27,37], among others. The methods used in [27,37] could be used for regression settings with a growing number of
covariates. A dimension reduction with the LASSO penalty is a good choice, and some of the advantages of the EL could be applied in this context. The new direction on EL seems to be promising and we will explore the challenging issue for the $p \gg n$ case in the future.

Note that in Table 1, we find that for a very small sample size $n = 40$ and very heavy censoring rate 40% both the NA method and EL method perform worse. The reason is that estimators of regression parameters are asymptotically biased for the estimating equations. To ensure better finite sample performance and also the consistency of the proposed estimator, a tail restriction is usually needed. Fine et al. [14] investigated this important problem for the linear transformation model, and proposed a modification of the estimation procedures [10] for regression parameters. In this paper, we did not apply the tail restriction [14] to the estimation equation for EL inference. Thus, it may deteriorate the performance of the proposed EL method for very heavy censoring, see Table 1. It is worthwhile to investigate transformation models combining empirical likelihood and tail restriction in the future. In the future, we will study this interesting and important issue and hope to improve the performance for very heavy censoring rates substantially.

In addition, the proposed EL method can be applied to other general transformation models such as Fine, Cai et al., Subramanian and Kong et al. [15,16,29,38], among others. Furthermore, our proposed method could be applied to other models involving $U$-statistics estimating equations, e.g. the accelerated failure time model with right censoring (see [39,40]), among others. In this paper, we assumed that the censoring and the survival time are independent. This is a rather heavy assumption. In the regression context, one usually prefers to work with the assumption of conditional independence, given the values of the covariates. Since the latter assumption is more realistic in practice. Simulation results in Table 3 demonstrate that our proposed EL and AEL methods still work well even when the independence assumption is invalid. Thus, the method can be used by practitioners in practice. More recently, Chen [41] developed weighted Breslow-type and maximum likelihood estimation for semiparametric transformation models under very general conditions. We will investigate confidence regions for the regression parameter using the empirical likelihood approach.

Acknowledgments

The author would like to thank Professor Wenbin Lu for making available some computer programs. The author is grateful to the associate editor and the two referees for their very helpful suggestions and useful comments, which substantially improved the paper. The research is partially supported by a grant from the National Security Agency and the Research Initiation Grant, Georgia State University.

Appendix. Proofs of theorems

We need the following lemma in order to prove Theorem 2.1. The variance of $S_n$ can be found following the derivation of [42] or [43] or [44] for scalar $U$-statistics. Then we have

$$\text{var}(S_n(\beta_0)) = \frac{4}{n} + O\left(n^{-2}\right), \quad a.s. \tag{A.1}$$

From Theorem 7.1 of [42], the asymptotic normality for $p$-dimensional multivariate $U$-statistic of degree 2 is as follows.

**Lemma A.1.** Let $S_n(\beta_0)$ be an unbiased estimating function of the form (2.7) for $\beta_0 \in \mathbb{R}^p$. Assume that $E\{b(U_1, U_2; \beta_0)b^T(U_1, U_2; \beta_0)\} < \infty$ and the matrix $\Sigma$ is positive definite. Then as $n \to \infty$, we have

$$n^{1/2}S_n(\beta_0) \xrightarrow{D} N_p(0, 4\Sigma).$$

**Lemma A.2.** Let $\Sigma_n = 1/n \sum_{i=1}^n W_i(\beta_0)W_i^T(\beta_0)$. If $E\{b(U_1, U_2; \beta_0)b^T(U_1, U_2; \beta_0)\} < \infty$ for the vector $b$, we have

$$\Sigma_n = \Sigma + o(1), \quad a.s.$$  

**Proof.** Note that

$$\Sigma_n = \frac{1}{n} \sum_{i=1}^n W_i(\beta_0)W_i^T(\beta_0) = \frac{1}{n} \sum_{i=1}^n \{W_i(\beta_0) - S_n(\beta_0)\}\{W_i(\beta_0) - S_n(\beta_0)\}^T + S_n(\beta_0)S_n^T(\beta_0). \tag{A.2}$$

We let $\text{vár}(\text{jack})$ be the jackknife estimator of $\text{var}(S_n(\beta_0))$. Following the same argument of [45], p. 223–224) for 1-dimensional $U$-statistics and [44], we have that

$$\frac{1}{n} \sum_{i=1}^n \{W_i(\beta_0) - S_n(\beta_0)\}\{W_i(\beta_0) - S_n(\beta_0)\}^T = \frac{(n - 2)^2}{4(n - 1)} \text{vár}(\text{jack}). \tag{A.3}$$

Since $\text{vár}(\text{jack})$ is a consistent estimator of $\text{var}(S_n(\beta_0))$ in the sense that

$$n[\text{vár}(\text{jack}) - \text{var}(S_n(\beta_0))] \to 0, \quad a.s.$$
Then as \( n \to \infty \), from (A.1) and (A.3) we have that
\[
\frac{1}{n} \sum_{i=1}^{n} \{W_i(\beta_0) - S_n(\beta_0)\}^T \left[ \frac{(n-2)^2}{4(n-1)} \text{var}[S_n(\beta_0)] + o(n^{-1}) \right] = \left( \frac{4\Sigma}{n} + O(n^{-2}) \right)
\]
\[
= \Sigma + o(1), \quad \text{a.s.} \tag{A.4}
\]
By the strong law of large numbers for \( U \)-statistics we have that \( S_n(\beta_0) = o(1) \), a.s. Thus, by (A.2) and (A.4) we have that
\( \Sigma_n = \Sigma + o(1) \), a.s. \( \Box \)

Thus, \( \lim_{n \to \infty} \Sigma_n \) (in probability) exists and it is equal to \( \Sigma \).

**Lemma A.3.** Under the conditions of Theorem 2.1, we have

(i) \( \frac{1}{n} \sum_{i=1}^{n} W_i(\beta_0)W_i^T(\beta_0) \xrightarrow{p} \Sigma \),

(ii) \( \hat{\Sigma} \xrightarrow{p} \Sigma \),

(iii) \( \hat{I} \xrightarrow{p} I \).

**Proof.** Let
\[
\hat{\Sigma}_n = \frac{1}{n} \sum_{i=1}^{n} W_i(\beta_0)W_i^T(\beta_0).
\]
In order to prove (i), we only need to show \( \hat{\Sigma}_n = \Sigma_n + o_P(1) \). For any \( a \in \mathbb{R}^p \), the following decomposition holds:
\[
a^T(\hat{\Sigma}_n - \Sigma_n)a = \frac{1}{n} \sum_{i=1}^{n} \left[ a^T(W_i(\beta_0) - W_i(\beta_0)) \right]^2 + 2 \frac{1}{n} \sum_{i=1}^{n} \{a^T(W_i(\beta_0))\} \{a^T(W_i(\beta_0) - W_i(\beta_0))\}
\]
\[
=: l_1 + 2l_2. \tag{A.5}
\]
By Gill [46], we have that
\[
K_1 = \sup_{0 \leq x \leq X(n)} \left| \frac{G(x) - \hat{G}(x)}{G(x)} \right| = o_P(1). \tag{A.6}
\]
Denote
\[
\phi_i(\beta) = w(Z \beta X) \frac{\delta I(X_i \geq X_j)}{C(X_j)}, \quad \hat{\phi}_i(\beta) = w(Z \beta X) \frac{\delta I(X_i \geq X_j)}{C(X_j)}.
\]
We have
\[
|a^T \phi_i(\beta_0)| \leq |a^T (\hat{\phi}_i(\beta_0) - \phi_i(\beta_0))| + |a^T \phi_i(\beta_0)|
\]
\[
\leq 3K_1^2 |a^T \hat{\phi}_i(\beta_0)| + 2K_1 |a^T \phi_i(\beta_0)| + |a^T \phi_i(\beta_0)|.
\]
Thus, we have
\[
|a^T \hat{\phi}_i(\beta_0)| \leq \frac{|a^T \phi_i(\beta_0)|}{1 - 3K_1^2 - 2K_1}. \tag{A.7}
\]

Denote \( v(U_1, U_2; \beta) = |a^T \phi_i(\beta)| + |a^T \phi_i(\beta)| \) and \( \theta = E v(U_1, U_2; \beta_0) \). Put \( v_n(U_1, U_2; \beta_0) = |a^T \hat{\phi}_i(\beta_0)| + |a^T \hat{\phi}_i(\beta_0)| \). Let
\[
V_i(\beta) = \frac{1}{n-1} \sum_{j=1,j \neq i}^{n} \{v(U_i, U_j; \beta)\},
\]
\[
V_m(\beta) = \frac{1}{n-1} \sum_{j=1,j \neq i}^{n} \{v_n(U_i, U_j; \beta)\},
\]
for \( i = 1, \ldots, n \). The values of \( V_i \) are identically distributed. Then
\[
T_n(\beta) = \frac{1}{n} \sum_{i=1}^{n} V_i(\beta) \tag{A.8}
\]
is a \( U \)-statistic for fixed \( \beta \). By the strong law of large numbers for \( U \)-statistics, \( T_n(\beta_0) = \theta + o(1) \), a.s.
First note that

\[
\left| a^T \{ W_i(\beta_0) - W_i(\beta_0) \} \right| \leq \frac{1}{n-1} \sum_{j=1,j\neq i}^n \left\{ 3K^2 \left| a^T \hat{\phi}_j(\beta_0) \right| + 2K_1 \left| a^T \hat{\phi}_j(\beta_0) \right| \right\} \\
+ \frac{1}{n-1} \sum_{j=1,j\neq i}^n \left\{ 3K^2 \left| a^T \hat{\phi}_j(\beta_0) \right| + 2K_1 \left| a^T \hat{\phi}_j(\beta_0) \right| \right\} \\
= \frac{3K^2 + 2K_1}{1 - 3K^2 - 2K_1} \frac{1}{n-1} \sum_{j=1,j\neq i}^n \left\{ v(U_j; \beta_0) \right\}.
\]

We also have

\[
\left| a^T W_i(\beta_0) \right| \leq \frac{1}{n-1} \sum_{j=1,j\neq i}^n \left\{ \left| a^T \phi_j(\beta_0) \right| + \left| a^T \phi_j(\beta_0) \right| \right\} \\
+ \frac{1}{n-1} \sum_{j=1,j\neq i}^n \left\{ \left| a^T w(Z_i^T \beta_0)Z_j \xi(Z_i^T \beta_0) \right| + \left| a^T w(Z_i^T \beta_0)Z_j \xi(Z_i^T \beta_0) \right| \right\} \\
= v_i(\beta_0) + N_i(\beta_0).
\]

Recall that \( Z_i \) is bounded. Thus, \( w(Z_i^T \beta_0)Z_j \xi(Z_i^T \beta_0) \) is bounded. Let \( \left| a^T w(Z_i^T \beta_0)Z_j \xi(Z_i^T \beta_0) \right| \leq C_1/2 \). Hence \( N_i(\beta_0) \leq C_1 \).

Note that \( 1/n \sum_{i=1}^n \{ v_i(\beta_0; \beta_0) \} = O(1) \) by the similar argument of Lemma A.2. It follows from \( E \{ v_i(\beta_0) \} = \theta \) that

\[
\left| I_2 \right| \leq \frac{3K^2 + 2K_1}{1 - 3K^2 - 2K_1} \frac{1}{n-1} \sum_{i=1}^n \left\{ \left| v_i(\beta_0) \right| + C_1 \right\} \\
\leq \frac{3K^2 + 2K_1}{1 - 3K^2 - 2K_1} \frac{1}{n-1} \sum_{i=1}^n \left\{ 2V_i^2(\beta_0) + C_1^2 \right\} \\
\leq \frac{3K^2 + 2K_1}{1 - 3K^2 - 2K_1} \frac{2}{n-1} \sum_{i=1}^n \left\{ 2[V_i(\beta_0) - E[V_i(\beta_0)])]^2 + 2[E[V_i(\beta_0)])^2 + C_1^2 \right\} \\
\leq \frac{3K^2 + 2K_1}{1 - 3K^2 - 2K_1} \frac{2}{n-1} \sum_{i=1}^n \left\{ 2[V_i(\beta_0) - \theta)]^2 + 2\theta^2 + C_1^2 \right\} \\
= o_p(1). \tag{A.9}
\]

Similarly we can show that \( I_1 = o_p(1) \). Thus by (A.5), (A.9), we prove Lemma A.3(i).

In order to prove Lemma A.3(ii), we only need to show that \( \hat{\Sigma} = \hat{\Sigma}_n + o_p(1) \). Let

\[
J_i = W_i(\hat{\beta}) - W_i(\beta_0).
\]

Denote \( \hat{\beta} = (\hat{\beta}^1, \ldots, \hat{\beta}^p)^T, \beta = (\beta^1, \ldots, \beta^p)^T \), and \( \beta_0 = (\beta_0^1, \ldots, \beta_0^p)^T \), respectively. Applying the mean-value theorem we obtain the following equality:

\[
\eta_1(i, j) = w(Z_i^T \hat{\beta}) - w(Z_i^T \beta_0) \\
= \sum_{k=1}^p \frac{\partial w(Z_i^T \beta_0)}{\partial \beta^k} \frac{\partial}{\partial \beta^k} \{ Z_i^T (\beta_0^1, \ldots, \beta_0^{k-1}, \hat{\beta}_k, \hat{\beta}_{k+1}, \ldots, \hat{\beta}^p)^T \},
\]

where \( \hat{\beta}_k \) is between \( \beta_0^k \) and \( \hat{\beta}_k \) for \( 1 \leq k \leq p \). Combining the above equality, conditions 1–2, the consistency and the asymptotic normality of \( \hat{\beta} \) (cf. (2.4)), we have \( |\eta_1(i, j)| \leq L_1 = O_p(n^{-1/2}) \), for \( 1 \leq i \leq n, 1 \leq j \leq n \).

Recall that \( \hat{\xi}(\cdot) \) is continuous. Similarly as before, we can apply the mean-value theorem to

\[
\eta_2(i, j) = w(Z_i^T \hat{\beta}) \xi(Z_i^T \hat{\beta}) - w(Z_i^T \beta_0) \xi(Z_i^T \beta_0).
\]

Combining the above equality, conditions 1–2, the consistency and the asymptotic normality of \( \hat{\beta} \), we have \( |\eta_2(i, j)| \leq L_2 = O_p(n^{-1/2}) \), for \( 1 \leq i \leq n, 1 \leq j \leq n \).

Denote

\[
f_{ij} = a^T Z_i \delta I(X_i \geq X_j) / G^2(X_j), \quad \hat{f}_{ij} = a^T Z_i \delta I(X_i \geq X_j) / \hat{G}^2(X_j).
\]
As (A.7), we have
\[ |\hat{f}_{ij}| \leq \frac{1}{1 - 3K_1^2 - 2K_1} |f_{ij}|. \]

Combining the above equalities, conditions 1–2, the consistency and the asymptotic normality of \( \hat{\beta} \), we have
\[ |a^T \hat{J}_k| \leq \frac{1}{n - 1} \sum_{j=1,j\neq k}^{n} \left\{ L_1 (|\hat{f}_{ij}| + |\hat{f}_{ji}|) + L_2 (|a^T Z_{ij}| + |a^T Z_{ji}|) \right\} \]
\[ \leq \frac{L_1}{1 - 3K_1^2 - 2K_1} \frac{1}{n - 1} \sum_{j=1,j\neq k}^{n} \{ |f_{ij}| + |f_{ji}| \} + 2L_2C_2, \]
where \(|a^T Z_{ij}| \leq C_1\) for \(1 \leq i \leq n\) and \(1 \leq j \leq n\). Denote
\[ M_i = \frac{1}{n - 1} \sum_{j=1,j\neq i}^{n} (|f_{ij}| + |f_{ji}|). \]

The \(M_i\)'s are identically distributed. Then \(1/n \sum_{i=1}^{n} M_i\) is a \(U\)-statistic. Denote \( \alpha = E(|f_{12}|) \). By the strong law of large numbers for \(U\)-statistics, \(1/n \sum_{i=1}^{n} M_i = \alpha + o(1)\), a.s. Note that \(1/n \sum_{i=1}^{n} (M_i - \alpha)^2 = o(1)\) by the similar argument of Lemma A.2. It follows \(1/n \sum_{i=1}^{n} M_i^2 = o(1)\).

We have
\[ \frac{1}{n} \sum_{i=1}^{n} (a^T J_i)^2 \leq \frac{L_1^2}{(1 - 3K_1^2 - 2K_1)^2} \frac{2}{n} \sum_{i=1}^{n} \left\{ \frac{1}{n - 1} \sum_{j=1,j\neq i}^{n} (|f_{ij}| + |f_{ji}|) \right\}^{2} + \frac{8L_2^2}{n} \sum_{i=1}^{n} C_2^2 \]
\[ \leq \frac{2L_1^2}{(1 - 3K_1^2 - 2K_1)^2} \frac{1}{n} \sum_{i=1}^{n} M_i^2 + \frac{8L_2^2}{n} C_2^2 \]
\[ = O_p(n^{-1/2}). \]

From (A.7), we have
\[ V_{un}(\beta_0) \leq \frac{1}{(1 - 3K_1^2 - 2K_1)^2} V_i(\beta_0). \]

We have
\[ \frac{2}{n} \sum_{i=1}^{n} \left\{ |a^T J_i| \cdot |a^T W_{ni}(\beta_0)| \right\} \leq \frac{2}{n} \sum_{i=1}^{n} \left\{ \frac{L_1 M_i}{(1 - 3K_1^2 - 2K_1)} + 2L_2 C_2 \right\} \left\{ V_{ni}(\beta_0) + C_1 \right\} \]
\[ \leq \frac{L_1}{(1 - 3K_1^2 - 2K_1)} \frac{1}{n} \sum_{i=1}^{n} \left\{ M_i^2 + V_i^2(\beta_0) \right\} + 4L_2 C_1 C_2 \]
\[ + \frac{2L_1 C_1}{(1 - 3K_1^2 - 2K_1)} \frac{1}{n} \sum_{i=1}^{n} M_i + \frac{4L_2 C_2}{(1 - 3K_1^2 - 2K_1)} \frac{1}{n} \sum_{i=1}^{n} V_i(\beta_0) \]
\[ = O_p(n^{-1/2}). \]

We have
\[ |a^T (\hat{\Sigma} - \Sigma_0)| a \leq \frac{1}{n} \sum_{i=1}^{n} (a^T J_i)^2 + \frac{2}{n} \sum_{i=1}^{n} \left\{ |a^T J_i| \cdot |a^T W_{ni}(\beta_0)| \right\} \]
\[ = O_p(n^{-1/2}). \]

Therefore Lemma A.3(iii) follows. Following the same line as above, we can show Lemma A.3(iii). □

**Proof of Theorem 2.1.** Denote \( Y_n = \max_{1 \leq i \neq j \leq n} \|b(U_i, U_j; \beta_0)\| \). Following the proof of Lemma 3 of [28], we have \( Y_n = o(n^{1/2}) \), a.s. Note that
\[ \|W_i(\beta_0)\| \leq \frac{1}{n - 1} \sum_{j=1,j\neq i}^{n} \|b(U_i, U_j)\| \leq Y_n. \]
for any $1 \leq i \leq n$. Thus,

$$\max_{1 \leq i \leq n} \| W_i(\beta_0) \| = o(n^{1/2}), \quad \text{a.s.} \tag{A.10}$$

By Zhou [47], we have that

$$K_2 = \sup_{0 \leq x \leq N(n)} \left| \frac{G(x) - \hat{G}(x)}{\hat{G}(x)} \right| = O_p(1).$$

Then we have

$$\| W_m(\beta_0) - W_i(\beta_0) \| \leq \frac{1}{n-1} \sum_{i=1, i \neq j}^{n} \left\{ K_2^2 \| \phi_j(\beta_0) \| + 2K_2 \| \phi_j(\beta_0) \| \right\} + \frac{1}{n-1} \sum_{i=1, j \neq i}^{n} \left\{ K_2^2 \| \phi_j(\beta_0) \| + 2K_2 \| \phi_j(\beta_0) \| \right\} .$$

Thus as (A.10), it follows that

$$\max_{1 \leq i \leq n} \| W_m(\beta_0) - W_i(\beta_0) \| \leq 2K_2 \max_{1 \leq i \neq j \leq n} \| \phi_j(\beta_0) \| + 4K_2 \max_{1 \leq i \neq j \leq n} \| \phi_j(\beta_0) \|$$

$$= o_P(n^{1/2}).$$

Thus, we have

$$\max_{1 \leq i \leq n} \| W_m(\beta_0) \| \leq \max_{1 \leq i \leq n, 1 \leq j \leq n} \| W_m(\beta_0) - W_i(\beta_0) \| + \max_{1 \leq i \leq n} \| W_i(\beta_0) \|$$

$$= o_P(n^{1/2}).$$

We have $\hat{\Sigma}_n = \Sigma + o_P(1)$ from Lemmas A.2 and A.3. From the Appendix of [10], we have that $2U(\beta_0) = (n - 1) \sum_{i=1}^{n} W_m(\beta_0) = o_P(n^{3/2})$. Then, it follows from (2.9) and the argument used in [19] that

$$\| \lambda \| = O_P(n^{-1/2}). \tag{A.11}$$

Combining (A.11), Taylor’s expansion to (2.8) and the arguments as those in [19], we have that

$$\frac{1}{4} l(\beta_0) = \frac{1}{4} \sum_{i=1}^{n} \lambda^T W_m(\beta_0) + o_P(1)$$

$$= \frac{1}{4} \left\{ n^{-3/2} 2U(\beta_0) \right\}^T \left\{ n^{-1} \sum_{i=1}^{n} W_m(\beta_0) W_m^T(\beta_0) \right\}^{-1} \left\{ n^{-3/2} 2U(\beta_0) \right\} + o_P(1)$$

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

From the Appendix of [10], we have that $\Gamma^{-1/2} \left\{ n^{-3/2} U_n(\beta_0) \right\} \overset{D}{\to} N(0, I_p)$. Since $\Gamma^{-1/2} \Sigma^{-1} \Gamma^{1/2}$ and $\Sigma^{-1} \Gamma$ have the same eigenvalues, Theorem 1 follows.

**Proof of Theorem 2.2.** Combining Taylor’s expansion to (2.8) and the arguments as those in [19], we have that

$$\frac{1}{4} l_{ad}(\beta_0) = \left\{ \Gamma^{-1/2} n^{-3/2} U(\beta_0) \right\}^T \left\{ \Gamma^{-1/2} n^{-3/2} U(\beta_0) \right\} + o_P(1).$$

By $\Gamma^{-1/2} \left\{ n^{-3/2} U_n(\beta_0) \right\} \overset{D}{\to} N(0, I_p)$. Theorem 2.2 is complete.

**References**