EMPIRICAL LIKELIHOOD REGRESSION ANALYSIS FOR
RIGHT CENSORED DATA

Gang Li\textsuperscript{1} and Qi-Hua Wang\textsuperscript{2}

\textsuperscript{1}University of California and
\textsuperscript{2}Chinese Academy of Science and Heilongjiang University

Abstract: Linear models are useful alternatives to the Cox (1972) proportional hazards model for analyzing censored regression data. This article develops empirical likelihood methods for linear regression analysis of right censored data. An adjusted empirical likelihood is constructed for the vector of regression coefficients using a synthetic data approach. The adjusted empirical likelihood is shown to have a central chi-squared limiting distribution, which enables one to make inference using standard chi-square tables. We also derive an adjusted empirical likelihood method for linear combinations of the regression coefficients. In addition, we discuss how to incorporate auxiliary information. A small simulation study is carried out to highlight the performance of the adjusted empirical likelihood methods compared with the traditional normal approximation method. It shows that the empirical likelihood confidence intervals tend to have more accurate coverage probabilities than the normal theory intervals. An illustration is given using the Stanford Heart Transplant data.

Key words and phrases: Confidence intervals, linear models, right censoring, synthetic data.

1. Introduction

Owen (1991) and Chen (1993, 1994) derived empirical likelihood inference procedures for linear models. The empirical likelihood methods have sampling properties similar to those of the bootstrap and have wider validity than the usual parametric procedures. They also have the appealing feature that the shape and orientation of the resulting confidence regions are determined entirely by the data. In contrast, it would be difficult to determine the $1 - \alpha$ central fraction of a point cloud with the bootstrap method when the dimension is two or higher.

In survival analysis, the survival time of interest is often not completely observed. For example, a common situation is that of right censoring where, due to the end of follow-up or occurrence of competing events, the survival times of some individuals are not observed, but are known to be greater than some observed values. The purpose of this paper is to extend the empirical likelihood method to linear regression analysis for right censored survival data. Linear
models provide useful alternatives to the popular Cox (1972) model for analysis of survival data when the proportional hazards assumption fails to hold.

Specifically, assume that one observes \( n \) i.i.d. triples \((X_i, \tilde{Y}_i, \delta_i) = (X_i, Y_i \wedge C_i, I[Y_i \leq C_i]), i = 1, \ldots, n\), where for subject \( i \), \( Y_i \) is a known monotone transformation of the survival time of interest, \( C_i \) is the corresponding censoring time and \( X_i = (X_{i1}, \ldots, X_{ip})^T \) is a vector of \( p \) covariates. Consider the linear model

\[
Y_i = X_i^T \beta_0 + \epsilon_i, \quad i = 1, \ldots, n,
\]

where \( \beta_0 \) is an (unknown) column vector of regression coefficients, \( \epsilon_i = Y_i - E(Y_i|X_i) \), and \( C_i \) is independent of \((X_i, Y_i), i = 1, \ldots, n\). Note that the distribution of \( \epsilon_i \) is completely unknown and that \( \text{Var}(\epsilon|X = x) \) is allowed to depend on \( x \). In the absence of censoring, this model reduces to the linear model studied by Owen (1991) who gave a nice discussion of why empirical likelihood is adequate for linear models with heteroscedastic error terms (Owen (1991), Section 5.1). In this paper, we develop empirical likelihood inference for \( \beta_0 \) and its linear combinations based on right censored data \((X_i, \tilde{Y}_i, \delta_i), i = 1, \ldots, n\).

Apparently the results of Owen (1991) and Chen (1993, 1994) for complete data do not apply to censored data since the \( Y_i \)'s are not always observed. A possible solution is to consider a synthetic data approach, used by Koul, Susarla and Ryzin (1981), Leurgans (1987) Srinivasan and Zhou (1991) and Zhou (1992), to derive least squares estimates of \( \beta_0 \). For simplicity, we proceed with Koul et al.'s (1981) proposal. The basic idea is to first introduce a synthetic response variable whose expectation is close to that of \( Y_i \). This is done in Section 2. A complete data empirical likelihood is then constructed for \( \beta_0 \) from the synthetic data as if they were i.i.d. observations. Because the synthetic data are in fact dependent, standard chi-square tables do not directly apply. By examining an asymptotic expansion of the empirical likelihood, we introduce an empirical adjustment and show that the adjusted empirical log-likelihood has an asymptotic standard chi-square distribution. The adjustment factor reflects the information loss due to censoring. In the absence of censoring, our adjusted empirical likelihood reduces to the standard one of Owen (1991). It is worth noting that a similar technique was used by Kitamura (1997) who developed a so-called block empirical likelihood using an adjustment factor in a dependent process model.

We also consider the problem of making empirical likelihood inference for linear combinations of the regression coefficients. Examples of linear combinations include a single coefficient, a subset of coefficients, and contrasts. For the complete data problem, Chen (1994) showed that the empirical likelihood still has a standard chi-square limiting distribution after the nuisance parameters are profiled out. However, it is not clear whether or not the same can be said for censored data. Moreover, profiling the nuisance parameters involves constrained
optimization which may not be an easy task in high dimensional cases. Instead
of using a profile likelihood, we replace the nuisance parameters by their least
squares estimates and derive an adjusted empirical likelihood with an appropriate
adjustment factor.

We further extend the adjusted empirical likelihood method to situations
where there is available auxiliary information on $X$. The results are useful in
instances where some population characteristics of the covariate $X$ are known.
For example, one may know the mean or median of $X$, or that the population
distribution is symmetric about a known constant.

The use of a likelihood ratio in nonparametric settings dates back at least
to Thomas and Grunkemeier (1975) who derived nonparametric likelihood ratio-
based confidence intervals for survival probabilities. Its first theoretical develop-
ment was due to Owen (1988, 1990), who introduced empirical likelihood con-
fidence regions for the mean of a random vector based on i.i.d. observations.
During the last decade, empirical likelihood has been extended to a wide range
of applications including, among others, linear models (Owen (1991) and Chen
quantile estimation (Chen and Hall (1993), Zhou and Jing (2002)), biased sample
models (Qin (1993)), generalized estimating equations (Qin and Lawless (1994)),
truncation models (Li (1995a)), dependent process model (Kitamura (1997)),
partial linear models (Wang and Jing (1999)), mixture proportions (Qin (1999)),
random censorship models (Hollander, McKeague and Yang (1997), Li, Hollan-
Li and Van Keilegom (2002), Wang and Li (2002), Wang and Jing (1999) and
Wang and Wang (2001)), and confidence tubes for multiple quantile plots (Ein-
mahl and McKeague (1999)). Some nice discussion of properties of empirical
likelihood can be found in DiCiccio, Hall and Romano (1991), Hall (1992), and
Hall and Scala (1990), and elsewhere.

The paper is organized as follows. In Section 2, we derive an adjusted empir-
ical likelihood for making inference for $\beta_0$. A Wilks-type theorem is established.
It ensures that the resulting adjusted empirical likelihood confidence region has
asymptotically correct coverage probability. In Section 3, we develop an adjusted
empirical likelihood method for combinations of the regression coefficients. In
Section 4, we describe how to incorporate auxiliary information. In Section 5, we
conduct a small simulation study to compare the adjusted empirical likelihood
method with the traditional normal approximation method. An illustration is
given using the Stanford Heart Transplant data. In Section 6, we note some limi-
tations of the synthetic data approach and discuss possible extensions that could
lead to better empirical likelihood procedures. Proofs are given in the appendix.
2. Adjusted Empirical Likelihood for Global Inference

In this section we derive an adjusted empirical likelihood (ADEL) method to make global inference for $\beta_0$. From now on, we assume that $E(X_iX_i^T)$ is positive definite.

Define a synthetic variable $Y_{iG} = \tilde{Y}_i \delta_i/(1 - G(\tilde{Y}_i))$, $i = 1, \ldots, n$, where $G$ is the cumulative distribution function of the censoring time $C_i$. It can be verified that $E(Y_{iG} | X_i) = E(Y_i | X_i)$. Hence, under the linear model (1), we have

$$Y_{iG} = X_i^\top \beta_0 + e_i,$$

where $e_i = Y_{iG} - E(Y_{iG} | X_i)$.

It follows from (2) that $\beta_0 = (EX_iX_i^T)^{-1}E(X_iY_{iG})$, or $EX_i(Y_{iG} - X_i^\top \beta_0) = 0$. Therefore, for a given $\beta$, the problem of testing $H_0 : \beta = \beta_0$ is equivalent to testing $E(W_i(\beta)) = 0$ based on $n$ i.i.d. observations $W_i(\beta) = X_i(Y_{iG} - X_i^\top \beta)$, $i = 1, \ldots, n$.

If $G$ were known, one could test $EW_i(\beta) = 0$ using the empirical likelihood of Owen (1990):

$$l_n(\beta) = -2 \sup \left\{ \sum_{i=1}^{n} \log(np_i) \left| \sum_{i=1}^{n} p_i W_i(\beta) = 0, \sum_{i=1}^{n} p_i = 1, p_i \geq 0, i = 1, \ldots, n \right. \right\}.$$

It follows from Owen (1990) that, under $H_0 : \beta = \beta_0$, $l_n(\beta)$ has an asymptotic central chi-square distribution with $p$ degrees of freedom. An essential condition for this result to hold is that the $W_i(\beta)$'s in the linear constraint are i.i.d. random variables.

Unfortunately, the censoring distribution $G$ is generally unknown and thus $l_n(\beta)$ cannot be computed since it depends on $G$. A natural solution is to replace $G$ by its Kaplan-Meier (1958) estimator in $l_n(\beta)$. Specifically, let $W_{in}(\beta) = X_i(Y_{iGn} - X_i^\top \beta)$, where $G_n(t)$ is the Kaplan-Meier estimator of $G$ given by

$$1 - \hat{G}_n(t) = \prod_{i=1}^{n} \left[ \frac{n - i}{n - i + 1} \right]^{I[\hat{Y}_{(i)} \leq t, \delta(i) = 0]},$$

$\hat{Y}_{(1)} \leq \cdots \leq \hat{Y}_{(n)}$ are the order statistics of the $\hat{Y}$-sample, and $\delta(i)$ is the $\delta$ associated with $\hat{Y}_{(i)}$, $i = 1, \ldots, n$. An estimated empirical log-likelihood is defined by

$$\tilde{l}_n(\beta) = -2 \sup \left\{ \sum_{i=1}^{n} \log(np_i) \left| \sum_{i=1}^{n} p_i W_{in}(\beta) = 0, \sum_{i=1}^{n} p_i = 1, p_i \geq 0, i = 1, \ldots, n \right. \right\}.$$

It is easy to show that

$$\tilde{l}_n(\beta) = 2 \sum_{i=1}^{n} \log\{1 + \lambda^\top W_{in}(\beta)\},$$

(3)
where $\lambda$ is the solution of the equation

$$
\frac{1}{n} \sum_{i=1}^{n} \frac{W_{in}(\beta)}{1 + \lambda^2 W_{in}(\beta)} = 0. \tag{4}
$$

Because $W_{in}(\beta)$, $i = 1, \ldots, n$, are dependent, $\tilde{I}_n(\beta)$ no longer has an asymptotic standard chi-square distribution. In the appendix (Remark A.1), we show that $\tilde{I}_n(\beta_0)$ converges in distribution to $\sum_{i=1}^{p} w_i \chi^2_{1,i}$ where the $\chi^2_{1,i}$ are independent $\chi^2_1$ random variables, the weights $w_i$ are eigenvalues of $\Sigma^{-1}(\beta_0)(\Sigma_1(\beta_0) - \Sigma_2)$, and $\Sigma_1(\beta_0)$ and $\Sigma_2$ are defined by (C.1) and (C.2) in the appendix. Although the weights could be estimated from data, Monte Carlo simulations would be needed to compute percentiles of the limiting distribution even if the weights were known. Instead of estimating the distribution of $\tilde{I}_n(\beta)$ directly, we present another method by introducing an adjustment factor for $\tilde{I}_n(\beta)$ so that the adjusted empirical likelihood function has an asymptotic standard chi-square distribution.

The following notations are needed. Let $F$ denote the distribution of $Y_i$. Let $\hat{F}_n$ be the Kaplan-Meier estimator of $F$. Let $Q_n(s) = (\sum_{i=1}^{n} I[Y_i \leq s])/n$,

$$
H_n(s) = \frac{1}{n} \sum_{i=1}^{n} \frac{X_i Y_i \hat{G}_n I[Y_i \leq s]}{(1 - \hat{G}_n(s))(1 - \hat{F}_n(s))},
$$

$$
\Lambda_n(t) = \int_{0}^{t} \frac{1}{1 - \hat{G}_n(s)} ds \hat{G}_n(s) = \frac{1}{n} \sum_{i=1}^{n} \frac{(1 - \delta_i) I[Y_i \leq t]}{1 - Q_n(Y_i)},
$$

$$
\hat{\Sigma}_{1n}(\beta) = \frac{1}{n} \sum_{i=1}^{n} X_i X_i^T (Y_i \hat{G}_n - \hat{X}_i \beta)^2,
$$

$$
\hat{\Sigma}_{2n} = \frac{1}{n} \sum_{i=1}^{n} (1 - \delta_i) (H_n(Y_i) H_n^T(Y_i) - \Lambda_n(Y_i)),
$$

$$
\hat{\Sigma}_n(\beta) = \hat{\Sigma}_{1n}(\beta) - \hat{\Sigma}_{2n},
$$

$$
S_n(\beta) = \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_{in}(\beta) \right) \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_{in}(\beta) \right)^T.
$$

The adjusted empirical likelihood function is defined by

$$
\hat{I}_{n,ad}(\beta) = r_n(\beta) \tilde{I}_n(\beta), \tag{5}
$$

where $r_n(\beta) = \text{tr}(\Sigma^{-1}_n(\beta) S_n(\beta))/\text{tr}(\Sigma^{-1}_{1n}(\beta) S_n(\beta))$. The adjustment factor can be derived by examining the leading term in the asymptotic expansion of $\tilde{I}_n(\beta)$ given in the proof of Theorem 2.1 in the appendix. Another way of motivating $r_n(\beta)$ is to use a result of Rao and Scott (1981) who showed that the distribution of $\tilde{r}(\beta_0) \sum_{i=1}^{p} w_i \chi^2_{1,i}$ may be approximated by $\chi^2_p$, where $\tilde{r}(\beta_0) =$
For $\beta_0$ also be written as $\tilde{r}(\beta_0)$ can also be written as $\tilde{r}(\beta_0) = \frac{\text{tr}\{\Sigma^{-1}(\beta_0)\Sigma(\beta_0)\}}{\text{tr}\{\Sigma^{-1}(\beta_0)\Sigma(\beta_0)\}}$. Replacing $\Sigma^{-1}(\beta_0), \Sigma^{-1}(\beta_0)$ and $\Sigma(\beta_0)$ in $\tilde{r}(\beta_0)$ by the sample estimates $\hat{\Sigma}^{-1}(\beta_0), \hat{\Sigma}^{-1}(\beta_0)$ and $S_n(\beta_0)$, respectively, leads to the expression for $r_n(\beta_0)$.

**Theorem 2.1.** Assume that the conditions (C.XY), (C.FG), (C.\Sigma_1) and (C.\Sigma_2) listed in the appendix hold.

(a) As $n \to \infty$, $\hat{l}_{n,ad}(\beta_0) \xrightarrow{d} \chi^2_p$, where $\chi^2_p$ is a standard chi-square random variable with $p$ degrees of freedom.

(b) For $0 < \alpha < 1$, define $I_\alpha = \{ \beta : \hat{l}_{n,ad}(\beta) \leq \chi^2_{p,\alpha} \}$, where $\chi^2_{p,\alpha}$ is the upper $\alpha$th percentile of the $\chi^2_p$ distribution. Then $\lim_{n \to \infty} P(\beta_0 \in I_\alpha) = 1 - \alpha$.

(c) Define $I_\alpha = \{ \beta : \hat{l}_n(\beta) \leq \chi^2_{p,\alpha} \}$. Then $\lim_{n \to \infty} P(\beta_0 \in I_\alpha) > 1 - \alpha$.

**Remark 2.1.** It can be shown that, under mild conditions, $r_n(\beta)$ is always greater than or equal to 1. This implies that the adjusted interval $I_a$ is a subset of the unadjusted interval $I_e$. Furthermore, if $\delta_i = 1$ for all $i$, then $r_n(\beta) \equiv 1$ and $Y_{iG_n} = Y_i$ for all $i$. Therefore the adjusted empirical likelihood reduces to Owen’s (1991) empirical likelihood in the absence of censoring.

**Remark 2.2.** Theorem 2.1 (b)-(c) show that the asymptotic confidence level of $I_a$ is $1 - \alpha$ and that of $I_e$ exceeds $1 - \alpha$. In Section 5 we present a simulation study which indicates that, for small samples, the actual confidence level of $I_a$ tends to be lower than $1 - \alpha$, especially when there is heavy censoring. In such cases, $I_e$ showed smaller coverage probability errors than $I_a$ since it is wider than $I_a$. For large samples, however, $I_a$ is preferred to $I_e$ since it is narrower and is expected to have a smaller coverage error, Theorem 2.1 (b)-(c) and simulation.

### 3. Adjusted Empirical Likelihood for Linear Combinations of $\beta_0$

This section extends the adjusted empirical likelihood method to make inference for a vector of linear combinations $\theta_0 = C\beta_0$ of $\beta_0$, where $C = (C_1, C_2)$, $C_1$ is a $k \times k$ matrix and $C_2$ is a $k \times (p - k)$ matrix $(k \leq p - 1)$. For example, $\theta_0$ is the subvector of the first $k$ regression coefficients if $C_1 = I_k$ and $C_2 = 0$. If $k = 1$, then $\theta_0$ reduces to a single linear combination, which includes an individual regression coefficient and the mean response at a given $X$ level as special cases. Without loss of generality, we assume $C_1^{-1}$ exists.

Let $\gamma_0 = (\theta_0^\top, \beta_0^{(k)})^\top$, where $\beta_0^{(k)}$ denotes the column subvector of the last $p - k$ elements of $\beta_0$. Write $X_i = (X_{i1}^\top, X_{i2}^\top)^\top$, where $X_{i1}$ and $X_{i2}$ are $k \times 1$ and $(p - k) \times 1$ subvectors. Let $\tilde{X}_i = (\tilde{X}_{i1}^\top, \tilde{X}_{i2}^\top) = (X_{i1}^\top C_1^{-1}, \tilde{X}_{i2}^\top - \tilde{X}_{i1}^\top C_1^{-1} C_2)^\top$. Then, model (1) reduces to $Y_i = \tilde{X}_i^\top \gamma_0 + \epsilon_i$, $i = 1, \ldots, n$.

Let $\hat{\gamma}_n(G) = (\sum_{i=1}^n \tilde{X}_i \tilde{X}_i^\top)^{-1}(\sum_{i=1}^n \tilde{X}_i Y_i G)$ and let $\hat{\beta}_n^{(k)}(G)$ denote the subvector of the last $p - k$ elements of $\hat{\gamma}_n(G)$. Note that $E\{\tilde{X}_{i1} (Y_{iG} - \tilde{X}_{i1}^\top \theta_0 - \tilde{X}_{i2}^\top \beta_0^{(k)}) \} = 0$.
\( \hat{X}_{i2}^\top \hat{\beta}_{n(k)}(\hat{G}) \} = 0, \ i = 1, \ldots, n. \) Similar to the previous section, for a given \( \theta \), we introduce the auxiliary variables \( u_{in}(\theta) = \hat{X}_{i1}(Y_{i \hat{G}_n} - \hat{X}_{i1}^\top \theta - \hat{X}_{i2}^\top \hat{\beta}_{n(k)}(\hat{G}_n)), \ i = 1, \ldots, n, \) and define an estimated empirical likelihood function \( l_{nk}(\theta) = 2 \sum_{i=1}^n \log(1 + \lambda^\top u_{in}(\theta)), \) where \( \lambda \) satisfies \( \sum_{i=1}^n u_{in}(\theta)/[1 + \lambda^\top u_{in}(\theta)] = 0. \)

Again, an adjustment factor is needed for \( l_{nk}(\theta_0) \) to have a central chi-square limiting distribution. Write

\[
\frac{1}{n} \sum_{i=1}^n \hat{X}_i \hat{X}_i^\top = \left( \begin{array}{c} \frac{1}{n} \sum_{i=1}^n \hat{X}_{i1} \hat{X}_{i1}^\top, \ K_n^\top \\ K_n, \ P_n \end{array} \right).
\]

Let

\[
\eta_{ni} = \hat{X}_{i1} - \left( \frac{1}{n} \sum_{j=1}^n \hat{X}_{j1} \hat{X}_{j1}^\top \right) \left( \frac{1}{n} \sum_{j=1}^n \hat{X}_{j1} \hat{X}_{j1}^\top \right)^{-1} \hat{X}_{i1} + \left( \frac{1}{n} \sum_{j=1}^n \hat{X}_{j1} \hat{X}_{j1}^\top \right) \left( \frac{1}{n} \sum_{j=1}^n \hat{X}_{j1} \hat{X}_{j1}^\top - K_n^\top P_n^{-1} K_n \right)^{-1} \left( \hat{X}_{i1} - K_n^\top P_n^{-1} \hat{X}_{i2} \right),
\]

\[
H_{n0}(s) = \frac{1}{n} \sum_{i=1}^n \eta_{ni} Y_{i \hat{G}_n} I[s < \hat{Y}_i] \quad (1 - G_n(s))(1 - F_n(s - s)),
\]

\[
\hat{\Sigma}_{10n}(\theta) = \frac{1}{n} \sum_{i=1}^n \eta_{ni} \eta_{ni}^\top (Y_{i \hat{G}_n} - \hat{X}_{i1}^\top \theta - \hat{X}_{i2}^\top \hat{\beta}_{n(k)}(\hat{G}_n))^2,
\]

\[
\hat{\Sigma}_{20n}(\theta) = \frac{1}{n} \sum_{i=1}^n (1 - \delta_i) H_{n0}(\hat{Y}_i) H_{n0}(\hat{Y}_i)(1 - \Delta \hat{G}_n)(\hat{Y}_i),
\]

\[
\hat{\Sigma}_{n0}(\theta) = \hat{\Sigma}_{10n}(\theta) - \hat{\Sigma}_{20n},
\]

\[
\hat{\Sigma}_{n0}(\theta) = \frac{1}{n} \sum_{i=1}^n u_{in}(\theta) u_{in}(\theta)^\top,
\]

\[
S_{n0}(\theta) = \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n u_{in}(\theta) \right) \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n u_{in}(\theta) \right)^\top.
\]

An adjusted empirical likelihood is then defined by \( l_{nk,ad}(\theta) = r_{n0} l_{nk}(\theta) \), where \( r_{n0}(\theta) = \text{tr}(\hat{\Sigma}_{n0}^{-1}(\theta) S_{n0}(\theta)) / \text{tr}(\hat{\Sigma}_{n0}^{-1}(\theta) S_{n0}(\theta)). \)

**Theorem 3.1.** Assume that (C.XY) and (C.FG) given in the appendix hold. In addition, assume that \( \Sigma_{10}(\theta_0) = E[\eta_1 \eta_1^\top (Y_{1G} - \hat{X}_{11}^\top \gamma_0)^2] \) and \( \Sigma_{20} = \int_0^Q H_0(s) H_0(s) (1 - F(s)) (1 - \Delta G(s)) dG(s) \) are positive definite matrices, where

\[
H_0(s) = \frac{E[\eta_1 Y_{1G} I(s < \hat{Y}_1)]}{(1 - G(s))(1 - F(s - s))},
\]

\[
\eta_i = \hat{X}_{i1} - E(\hat{X}_{i1} \hat{X}_{i1}^\top) \{ E(\hat{X}_{i1} \hat{X}_{i1}^\top) \}^{-1} \hat{X}_{i1},
\]

\[
\hat{\Sigma}_{n0}(\theta) = \frac{1}{n} \sum_{i=1}^n u_{in}(\theta) u_{in}(\theta)^\top,
\]

\[
S_{n0}(\theta) = \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n u_{in}(\theta) \right) \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n u_{in}(\theta) \right)^\top.
\]
+E(\tilde{X}_{11}\tilde{X}_{11}^\tau)(E(\tilde{X}_{11}\tilde{X}_{11}^\tau) - K^\tau P^{-1} K)^{-1}(\tilde{X}_{11} - K^\tau P^{-1} \tilde{X}_{21}),
K = E(K_n), \quad P = E(P_n).

Then, \( l_{nk,ad}(\theta_0) \) is asymptotically standard chi-square distributed with \( k \) degrees of freedom.

4. Adjusted Empirical Likelihood with Auxiliary Information

In many applications, some auxiliary population characteristics of \( X \) are known. It has been shown in the literature that effective usage of the available auxiliary information can lead to more efficient inferences; see e.g., Chen and Qin (1993), Qin and Lawless (1994) and Zhang (1995, 1996). In this section we show how to incorporate auxiliary information of \( X \) using an adjusted empirical likelihood.

Assume that the available auxiliary information on \( X \) is given in the form
\( Eg(X) = 0 \), where \( g(x) = (g_1(x), \ldots, g_r(x))^\tau \), \( r \geq 1 \), is a vector of \( r \) known functions.

To make use of the auxiliary information, we maximize
\[
\prod_{i=1}^{n} p_i
\]  
subject to \( \sum_{i=1}^{n} p_i = 1 \), \( \sum_{i=1}^{n} p_i g(X_i) = 0 \) and \( \sum_{i=1}^{n} p_i \psi_{in}(\xi) = 0 \), where \( \psi_{in}(\xi) \) is \( W_{in}(\beta) \) or \( u_{in}(\theta) \) as in Sections 2 and 3, depending on the context.

Let \( A_{ni}(\xi) = (g^\tau(X_i), \psi^\tau_{in}(\xi))^\tau \). By the method of Lagrange multipliers, (6) is maximized at \( p_{im} = 1/[n(1 + \zeta^\tau_n A_{ni}(\xi))] \), \( i = 1, \ldots, n \), where \( \zeta_n \) satisfies \( \sum_{i=1}^{n} A_{ni}(\xi)/[n(1 + \zeta^\tau_n A_{ni}(\xi))] = 0 \). Hence, the empirical log-likelihood ratio function is given by \( l_{n,AU}(\xi) = -2 \sum_{i=1}^{n} \log np_{in} = 2 \sum_{i=1}^{n} \log(1 + \zeta^\tau_n A_{ni}(\xi)) \).

Just like \( \hat{l}_n(\beta) \) in Section 2, an adjustment factor is needed for \( l_{n,AU}(\xi) \) to have a standard chi-square asymptotic distribution. Let \( V_{n1}(\xi) = (\sum_{i=1}^{n} g(X_i) g^\tau(X_i))/n \), \( V_{n2}(\xi) = (\sum_{i=1}^{n} g(X_i) \psi_{in}(\xi))/n \) and \( V_{n3}(\xi) = (\sum_{i=1}^{n} \psi_{in}(\xi) \psi^\tau_{in}(\xi))/n \),

\[
V_{n1,AU}(\beta) = \begin{pmatrix} V_{n1}(\xi), & V_{n2}(\xi) \\ V_{n2}(\xi), & V_{n3}(\xi) \end{pmatrix} \text{ and } V_{n2,AU}(\beta) = \begin{pmatrix} V_{n1}(\xi), & 0 \\ 0, & \hat{S}(\xi) \end{pmatrix},
\]
where \( \hat{S}(\xi) \) is \( \hat{S}_n(\xi) \) defined in Section 2 or \( \hat{\sigma}^2_n(\theta) \) in Section 3.

Similar to (5), we define an adjusted empirical log-likelihood for \( \xi \) by \( \hat{l}_{n,AU}(\xi) = r_{n,AU}(\xi) l_{n,AU}(\xi) \), where \( r_{n,AU}(\xi) = \text{tr}(V_{n2,AU}(\xi) \Psi_{n}(\xi))/\text{tr}(V_{n1,AU}(\xi) \Psi_{n}(\xi)) \) and \( \Psi_{n}(\xi) = (\sum_{i=1}^{n} A_{ni}(\xi))/(\sum_{i=1}^{n} A_{ni}(\xi))^\tau \).

Theorem 4.1. Assume that \( Eg(X)g^\tau(X) \) is positive definite and that \( E\frac{g(X)X^\tau Y\delta}{1-G(Y)} \) exists.
(a) Let $\xi = \beta$ and $\psi_{in}(\xi) = W_{in}(\beta)$. Then, under the conditions of Theorem 2.1, 
\[ \hat{l}_{n,AU}(\beta_0) \xrightarrow{d} \chi^2_{p+r} \text{ as } n \to \infty. \]
(b) Let $\psi_{in}(\xi) = u_{in}(\theta)$, where $\xi = \theta$ is a vector of $k$ linear combinations of $\beta$ as in Section 3. Then, under conditions of Theorem 3.1, 
\[ \hat{l}_{n,AU}(\theta_0) \xrightarrow{d} \chi^2_{r+k} \text{ as } n \to \infty. \]

5. Example and Simulation

In this section we illustrate the adjusted empirical likelihood method and compare it to the normal approximation method using a real data set. We also present a small simulation study to compare the performance of empirical likelihood confidence intervals to the normal approximation method.

Consider the heart transplant data of Miller ((1976), Table 1). The data includes the lengths of survival (in days) after transplantation, ages at time of transplant, and T5 mismatch scores for 69 patients who received heart transplants at Stanford and were followed from October 1, 1967 to April 1, 1974. The T5 mismatch score is a measure of the degree of dissimilarity between the donor and recipient tissue. Twenty-four patients were still alive on April 1, 1974 and thus their survival times were censored.

Let $Y$ be the logarithm to base 10 of the length of survival from transplantation. The three models we consider are (I) regress $Y$ on the mismatch score $T_5$; (II) regress $Y$ on age; (III) regress $Y$ on both $T_5$ and age (Koul, Susarla and van Ryzin (1981)). As in Koul et al. (1981), regressions of survival on the mismatch score $T_5$ were performed with nonrejection related death being treated as censoring since the mismatch score is directed at the rejection phenomenon (Miller (1976)). Confidence intervals for the slope parameters based on the normal approximation method (cf. Koul et al. (1981) and Lai, Ying and Zheng (1995)) the adjusted empirical likelihood (ADEL) and the estimated likelihood are given in Table 1.

Table 1. 95% adjusted empirical likelihood (ADEL), estimated empirical likelihood (EEL), and normal confidence interval estimates for heart transplant data.

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter</th>
<th>Estimate</th>
<th>ADEL</th>
<th>EEL</th>
<th>Normal</th>
</tr>
</thead>
<tbody>
<tr>
<td>(I)</td>
<td>$\beta_{T_5}$</td>
<td>0.258</td>
<td>$[-0.512, 0.943]$</td>
<td>$[-0.518, 0.947]$</td>
<td>$[-0.466, 0.943]$</td>
</tr>
<tr>
<td>(II)</td>
<td>$\beta_{age}$</td>
<td>0.054</td>
<td>$[0.019, 0.108]$</td>
<td>$[0.016, 0.112]$</td>
<td>$[0.017, 0.096]$</td>
</tr>
<tr>
<td>(III)</td>
<td>$\beta_{T_5}$</td>
<td>0.052</td>
<td>$[-0.717, 0.759]$</td>
<td>$[-0.721, 0.762]$</td>
<td>$[-0.643, 0.746]$</td>
</tr>
<tr>
<td></td>
<td>$\beta_{age}$</td>
<td>0.077</td>
<td>$[0.042, 0.139]$</td>
<td>$[0.037, 0.150]$</td>
<td>$[0.039, 0.114]$</td>
</tr>
<tr>
<td>(IV)</td>
<td>$\beta_{T_5}$</td>
<td>$-0.108$</td>
<td>$[-0.691, 0.405]$</td>
<td>$[-0.696, 0.409]$</td>
<td>$[-0.601, 0.386]$</td>
</tr>
<tr>
<td></td>
<td>$\beta_{age}$</td>
<td>0.056</td>
<td>$[0.021, 0.109]$</td>
<td>$[0.018, 0.113]$</td>
<td>$[0.019, 0.093]$</td>
</tr>
</tbody>
</table>
It is seen that for the heart transplant data, the empirical confidence intervals are comparable to those of the normal approximation method. As expected, the estimated empirical likelihood (EEL) is the most conservative and gives larger intervals. It is observed that the empirical likelihood confidence intervals are asymmetric about the point estimate. They are shifted to the left for $\beta_{T_5}$ and to the right for $\beta_{\text{age}}$ compared to the normal confidence intervals. Recall that the traditional normal approximation method always imposes symmetry on the confidence interval. This is not a desirable property since the underlying distribution of the parameter estimate can be skewed. The empirical likelihood method is able to pick up possible skewness in the underlying distribution of the parameter estimate.

We carried out a small Monte Carlo simulation to examine coverage probabilities of the empirical likelihood confidence intervals compared to the normal approximation method. The data were generated from the following model: $Y = 1 + X + \epsilon$, where $X$ and $\epsilon$ are independent normal random variables with mean 0 and variance 0.25, the censoring time $C$ is a normal random variable with mean $\mu$ and standard deviation 4. We vary $\mu$ to produce different amounts of censoring. We also vary the sample size $n$. The simulated confidence levels of the empirical likelihood and normal confidence intervals for the slope parameter are given in Table 2. Each entry in the table was computed using 5000 Monte Carlo samples.

Table 2. Simulated coverage probabilities (%) of the normal, the adjusted empirical likelihood (ADEL) and the estimated empirical likelihood (EEL) confidence intervals for the slope parameter (nominal level = 1 - $\alpha$).

<table>
<thead>
<tr>
<th>Censoring Rate</th>
<th>Sample Size</th>
<th>$1 - \alpha = 90%$</th>
<th>$1 - \alpha = 95%$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Normal</td>
<td>ADEL</td>
<td>EEL</td>
</tr>
<tr>
<td>60%</td>
<td>50</td>
<td>77.9*</td>
<td>83.5</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>82.2*</td>
<td>86.5</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>90.4</td>
<td>90.9</td>
</tr>
<tr>
<td>32%</td>
<td>50</td>
<td>85.6*</td>
<td>87.8</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>89.4*</td>
<td>89.9</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>91.3</td>
<td>91.4</td>
</tr>
</tbody>
</table>

(* indicates a coverage probability that deviates the most from the nominal level among the three methods: “normal”, “ADEL” and “EEL”.)

Table 2 shows that the adjusted empirical likelihood confidence interval has more accurate coverage probabilities than the normal approximation method. The improvement of the empirical likelihood method is usually more pronounced for small samples (e.g., $n = 50$). Although the estimated empirical likelihood method is conservative for large samples (e.g., $n = 500$), it does improve the
Empirical likelihood methods are studied for linear regression analysis of right censored data. Our results are valid for the heteroscedastic model that allows the conditional variance of the response to vary with the level of the covariate. The proposed method for the coefficient vector $\beta_0$ reduces to that of Owen (1991) in the absence of censoring. However, our method for linear combinations of $\beta_0$ has not been seen in the literature even for complete data. The empirical likelihood method demonstrates better small sample performance than the normal approximation method in a small simulation study and the improvement could be more pronounced in higher dimensional cases.

The synthetic data used in this article has some limitations. As shown by our simulation results (Table 2), the coverage probability can be far below the nominal level when there is heavy censoring and the sample size is small, even though the adjusted empirical likelihood method produces some improvement over the normal approximation method. Other types of synthetic data have been suggested in the literature for right censored data; see, e.g., Leurgans (1987) and Lai et al. (1995) among others. We point out that the idea used in this article can be applied directly to derive adjusted empirical likelihood procedures based on other synthetic data. Furthermore, one could develop adjusted empirical likelihood along the same lines based on general estimating equations (Buckley and James (1979) and Lai et al. (1995)). The use of other types of synthetic data or general estimating equations may lead to more efficient empirical likelihood procedures. Further investigation of various adjusted empirical likelihood methods for right censored data will be carried out in another paper.

7. Appendix

The following conditions are needed in Theorem 2.1.

(C.XY). $E(XI[s < Y])$ exists for every $0 \leq s < \infty$.

(C.FG) (i). For all $s \leq \tau_Q \equiv \inf\{t : Q(t) = 1\}$, $G(s)$ and $F(s)$ have no common jumps, where $Q(t) = P(\tilde{Y} \leq t)$.

(ii). $E\left[\frac{\|X\|Y}{(1-G(y))(1-F(Y))}\right] < \infty$.

(iii). $\int_0^{\tau_Q} \|H(s)\|[\{1 - F(s)\}/[1-F(s-)]\]|dG(s)/(1-G(s)) < \infty$, where $H(s) = E[XYG(s < \tilde{Y})]/\{(1-G(s))(1-F(s-))\}$.

(C.\Sigma_1). $\Sigma_1(\beta_0) = E[XX^r(Y_G - X^r\beta_0)^2]$ is a positive definite matrix.

(C.\Sigma_2). $\Sigma_2 = \int_0^\infty \|H(s)\|H^r(s) (1 - F(s-))(1 - \Delta G(s)) dG(s)$ is a positive definite matrix, where
\[ \Lambda^G(t) = \int_{-\infty}^{t} \frac{1}{1 - G(s-)} dG(s). \]

The following lemma is needed to prove Theorem 2.1. For simplicity, we denote \( W_{in}(\beta_0) \) by \( W_{in} \).

**Lemma A.1.** Suppose that the assumptions (C.XY), (C.FG)(ii), (iii) and (C.\( \Sigma_1 \)) hold. Then \( n^{-\frac{1}{2}} \sum_{i=1}^{n} W_{in} \xrightarrow{\mathcal{L}} Z \), where \( Z \) is a \( p \)-variate normal \( N(0, \Sigma(\beta_0)) \) random vector, \( \Sigma(\beta_0) = \Sigma_1(\beta_0) - \Sigma_2 \), and \( \Sigma_1(\beta_0) \) and \( \Sigma_2 \) are defined in assumptions (C.\( \Sigma_1 \)) and (C.\( \Sigma_2 \)).

**Proof.** The proof is similar to that of Theorem 2 of Lai, Ying and Zheng (1995), and is omitted.

**Proof of Theorem 2.1.** (a) To prove part (a) of the theorem, we need to show that (i) \( \max_{1 \leq i \leq n} \|W_{in}\| = O_p(n^{-\frac{1}{2}}) \), and (ii) \( \lambda = O_p(n^{-\frac{1}{2}}) \). We first prove (i). It is seen that

\[
\begin{align*}
\max_{1 \leq i \leq n} \|W_{in}\| &\leq \max_{1 \leq i \leq n} \|X(Y_i \mathcal{G}_n - Y_i G)\| + \max_{1 \leq i \leq n} \|W_i\|, \\
\max_{1 \leq i \leq n} \|X(Y_i \mathcal{G}_n - Y_i G)\| &\leq \max_{1 \leq i \leq n} \|X Y_i G\| \sup_{0 \leq z \leq Y_n} \left| \frac{\hat{G}_n(z) - G(z)}{1 - \hat{G}_n(z)} \right|. 
\end{align*}
\]

(7) (8)

By Lemma 3 of Owen (1990), we have

\[ \max_{1 \leq i \leq n} \|W_i\| = o(n^{-\frac{1}{2}}) \]

(9)

under assumption (C.FG)(ii). Moreover, it follows from Zhou (1992) that

\[ \sup_{0 \leq z \leq Y_n} \left| \frac{\hat{G}_n(z) - G(z)}{1 - \hat{G}_n(z)} \right| = O_p(1). \]

(10)

Thus, (i) follows immediately from (7)–(10).

Next, we prove (ii). Let \( \lambda = \rho \theta \), where \( \rho \geq 0 \) and \( \|\theta\| = 1 \). Recall that \( \hat{\Sigma}_{1n}(\beta_0) = 1/n \sum_{i=1}^{n} W_{in} W_{in}^T \). Let

\[ \hat{\Sigma}_{1n}(\beta_0) = \frac{1}{n} \sum_{i=1}^{n} X_i X_i^T \left( \frac{\delta_i \hat{Y}_i}{1 - G(Y_i)} - X_i^T \beta_0 \right)^2. \]

It can be shown that

\[ \hat{\Sigma}_{1n}(\beta_0) = \hat{\Sigma}_{1n}(\beta_0) + o_p(1). \]

(11)

Let \( \sigma_p \) be the smallest eigenvalue of \( S = E[XX^T (\hat{Y}_G - X^T \beta_0)^2] \). Then, by Owen (1990),

\[ \theta^T \hat{\Sigma}_{1n}(\beta_0) \theta \geq \sigma_p + o_p(1). \]

(12)
Let $e_j$ be the unit vector in the $j$th coordinate direction. By Lemma A.1,

$$\left\| \frac{1}{n} \sum_{j=1}^{p} e_j' \sum_{i=1}^{n} W_{in} \right\| = O_p(n^{-\frac{1}{2}}). \quad (13)$$

It follows from (4), (11)–(13), and the arguments used in the proof of (2.14) of Owen (1990) that $\|\lambda\| = O_p(n^{-\frac{1}{2}})$. This proves (ii).

It follows from (i) and (ii) that $\max_{1 \leq i \leq n} |\lambda^T W_{in}| = O_p(n^{-\frac{1}{2}}) o_p(n^{\frac{1}{2}}) = o_p(1)$. Hence, by Taylor’s expansion, we have $\log(1 + \lambda^T W_{in}) = \lambda^T W_{in} - \frac{1}{2}(\lambda^T W_{in})^2 + \eta_i$, where, for some constant $C > 0$, $P(|\eta_i| \leq C|\lambda^T W_{in}|^3, 1 \leq i \leq n) \to 1$ as $n \to \infty$.

Therefore

$$\tilde{l}_n(\beta_0) = 2 \sum_{i=1}^{n} \log\{1 + \lambda^T W_{in}\} = 2 \sum_{i=1}^{n} \left( \lambda^T W_{in} - \frac{1}{2}(\lambda^T W_{in})^2 \right) + r_n, \quad (14)$$

$$P\left( |r_n| \leq C \sum_{i=1}^{n} |\lambda^T W_{in}|^3 \right) \to 1, \quad \text{as } n \to \infty. \quad (15)$$

Note that $\sum_{i=1}^{n} |\lambda^T W_{in}|^3 \leq \|\lambda\|^3 \max_{1 \leq i \leq n} \|W_{in}\| \sum_{i=1}^{n} \|W_{in}\|^2 = o_p(1)$, where the last step follows from (i), (ii), and the fact that

$$\frac{1}{n} \sum_{i=1}^{n} \|W_{in}\|^2 = O_p(1). \quad (16)$$

This, combined with (15), implies that

$$|r_n| = o_p(1). \quad (17)$$

Note that

$$0 = \frac{1}{n} \sum_{i=1}^{n} \frac{W_{in}}{1 + \lambda^T W_{in}}$$

$$= \frac{1}{n} \sum_{i=1}^{n} W_{in} \left[ 1 - \lambda^T W_{in} + \frac{(\lambda^T W_{in})^2}{1 + \lambda^T W_{in}} \right]$$

$$= \frac{1}{n} \sum_{i=1}^{n} W_{in} - \left( \frac{1}{n} \sum_{i=1}^{n} W_{in} W_{in}^T \right) \lambda + \frac{1}{n} \sum_{i=1}^{n} \frac{W_{in}(\lambda^T W_{in})^2}{1 + \lambda^T W_{in}}. \quad (18)$$

By (11), (18), (i) and (ii), we get

$$\lambda = \left( \sum_{i=1}^{n} W_{in} W_{in}^T \right)^{-1} \sum_{i=1}^{n} W_{in} + o_p(n^{-\frac{1}{2}}). \quad (19)$$
It is seen from (18) that
\[ 0 = \sum_{i=1}^{n} \frac{\lambda^\tau W_{in}}{1 + \lambda^\tau W_{in}} = \sum_{i=1}^{n} (\lambda^\tau W_{in}) - \sum_{i=1}^{n} (\lambda^\tau W_{in})^2 + \frac{1}{n} \sum_{i=1}^{n} \frac{\lambda^\tau W_{in}}{1 + \lambda^\tau W_{in}}. \] (20)

Moreover, by (i), (ii) and (16), we have
\[ \frac{1}{n} \sum_{i=1}^{n} \frac{\lambda^\tau W_{in}}{1 + \lambda^\tau W_{in}} = o_p(1). \] (21)

It is easy to see that (20) and (21) imply
\[ \sum_{i=1}^{n} \lambda^\tau W_{in} = \sum_{i=1}^{n} (\lambda^\tau W_{in})^2 + o_p(1). \] (22)

Combining (14), (17), (19) and (22) yields
\[ \hat{l}_n(\beta_0) = \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_{in} \right)^\tau \hat{\Sigma}_{1n}^{-1}(\beta_0) \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_{in} \right) + o_p(1). \] (23)

A direct argument can be used to prove
\[ \hat{\Sigma}_{1n}(\beta_0) \overset{p}{\rightarrow} \Sigma_1(\beta_0), \] (24)

where \( \Sigma_1(\beta_0) \) is defined in assumption (C.\Sigma_1) in the beginning of this section. Furthermore,
\[ \hat{\Sigma}_{2n}(\beta_0) = \int_0^\infty H_n(z)H_n^\tau(z)(1 - \triangle \hat{G}_n(z))(1 - Q_n(z))d\hat{G}_n(z) \overset{p}{\rightarrow} \Sigma_2(\beta_0), \]
by Stute and Wang (1993). Hence,
\[ \hat{\Sigma}_n(\beta_0) \overset{p}{\rightarrow} \Sigma(\beta_0). \] (25)

By Lemma A.1, (17), (23) and (25),
\[ \hat{I}_{n,ad}(\beta_0) = \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_{in} \right)^\tau \hat{\Sigma}_{1n}^{-1}(\beta_0) \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_{in} \right) + r_n(\beta_0) o_p(1) \overset{d}{\rightarrow} Z^\tau \Sigma^{-1}(\beta_0) Z \sim \chi^2_p. \]

This proves part (a).

(b). Part (a) implies immediately that \( P(\beta_0 \in I_\alpha) = P(\hat{I}_{n,ad}(\beta_0) \leq \chi^2_{p,\alpha}) \rightarrow 1 - \alpha. \)

(c). It follows from (23), (24) and Lemma A.1 that
\[ \bar{l}_n(\beta_0) = \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_{in} \right)^\tau \bar{\Sigma}_{1n}^{-1}(\beta_0) \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_{in} \right) + o_p(1) \overset{d}{\rightarrow} Z^\tau \Sigma_1^{-1}(\beta_0) Z. \] (26)
Therefore, \( P(\beta_0 \in I_{\alpha}) = P(\tilde{t}(\beta_0) \leq \chi^2_{p,\alpha}) = P(Z^T\Sigma^{-1}(\beta_0)Z \leq \chi^2_{p,\alpha}) > P(Z^T\Sigma^{-1}(\beta_0)Z \leq \chi^2_{p,\alpha}) = 1 - \alpha. \)

**Remark A.1.** It follows from (26) that the asymptotic distribution of \( \tilde{t}_n(\beta_0) \) is the same as that of \( \sum_{l=1}^{p} w_l \chi^2_i \) where \( w_l, l = 1, \ldots, p, \) are the eigenvalues of \( \Sigma^{-1}(\beta_0) \Sigma(\beta_0) \) and the \( \chi^2_i \)'s are independent standard chi-square random variables with one degree of freedom.

**Proof of Theorem 3.1.** Similar to (23) and (24), it can be shown that

\[
\hat{t}_{nk,ad}(\theta_0) = \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_{in}(\theta_0) \right) \hat{\Sigma}^{-1}_{0}(\theta_0) \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_{in}(\theta_0) \right) + o_p(1), \tag{27}
\]

\[
\hat{\Sigma}_{n0}(\theta_0) \xrightarrow{p} \Sigma_0(\theta_0), \tag{28}
\]

where \( \Sigma_0(\theta_0) = \Sigma_{10}(\theta_0) - \Sigma_{20}(\theta_0). \) Next we show that

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_{in}(\theta_0) \xrightarrow{L} N(0, \Sigma_0(\theta_0)). \tag{29}
\]

Let \( \hat{\gamma}_n = \hat{\gamma}_n(\hat{G}_n) \) and \( \hat{\theta}_n \) be the subvector of the first \( k \) elements of \( \hat{\gamma}_n. \) Using \( \hat{X}_i^\tau \hat{\gamma}_n = \hat{X}_i^\tau \hat{\theta}_n + \hat{X}_i^\tau \hat{\beta}_n(k)(\hat{G}_n), \) it can be verified that

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_{in}(\theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{X}_i(Y_{i\hat{G}_n} - \bar{X}_i^\tau \gamma_0) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{X}_i \bar{X}_i^\tau (\gamma_0 - \hat{\gamma}_n) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \bar{X}_i \bar{X}_i^\tau (\hat{\theta}_n - \theta_0). \tag{30}
\]

Denote the three terms on the right side of (30) by \( T_{n1}, T_{n2} \) and \( T_{n3} \) respectively. Then,

\[
T_{n2} = -(E[\hat{X}_{11} \hat{X}_{11}^\tau]) (E[\hat{X}_1 \hat{X}_1^\tau])^{-1} \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{X}_i(Y_{i\hat{G}_n} - \bar{X}_i^\tau \gamma_0) \right] + o_p(1), \tag{31}
\]

\[
T_{n3} = E(\hat{X}_{11} \hat{X}_{11}^\tau) \{ E(\hat{X}_1 \hat{X}_1^\tau) - K^\tau P^{-1} K \}^{-1} \times \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\hat{X}_i - \bar{X}_i^\tau K - \hat{X}_i^\tau \gamma_0) (Y_{i\hat{G}_n} - \bar{X}_i^\tau \gamma_0) \right] + o_p(1). \tag{32}
\]

It follows from (30), (31) and (32) that

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_{in}(\theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \eta_i(Y_{i\hat{G}_n} - \bar{X}_i^\tau \gamma_0) + o_p(1), \tag{33}
\]
where $\eta_i$ is defined in Theorem 3.1. This, together with the arguments in the proof of Theorem 2 of Lai, Ying and Zheng (1995) leads to (29).

Finally, combining (27), (28) and (29) completes the proof.

**Lemma A.2.**

(i). If $\psi_{in}(\xi) = W_{in}(\beta)$ and the conditions of Theorems 2.1 and 4.1 hold, then $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} A_{ni}(\beta_0) \xrightarrow{d} N(0, V_{2,\text{AU}}(\beta_0))$, where $V_{2,\text{AU}}(\beta_0) = \begin{pmatrix} V_1 & 0 \\ 0 & \Sigma(\beta_0) \end{pmatrix}$.

(ii). If $\psi_{in}(\xi) = u_{in}(\theta)$ and the conditions of Theorems 3.1 and 4.1 hold, then $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} A_{ni}(\theta_0)$ is asymptotically normal with mean 0 and variance-covariance matrix $V_{2,\text{AU}}(\theta_0) = \begin{pmatrix} V_1 & 0 \\ 0 & \sigma^2(\theta_0) \end{pmatrix}$.

**Proof.** Part (i) is a direct consequence of Lemma A.1 and the following facts: $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(X_i) \xrightarrow{d} N(0, V_1(\beta_0))$, $\text{Cov} \left( \frac{1}{\sqrt{n}} \sum g(X_i), \frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_{in}(\beta_0) \right) \rightarrow 0$. Part (ii) can be proved along the same lines.

**Proof of Theorem 4.1.** The theorem can be proved using Lemma A.2 and the same arguments used in the proof of Theorems 2.1 and 3.1. We omit the details.

**Acknowledgements**

The authors are grateful to the Editor, and an associate editor and two anonymous referees for their constructive comments that lead to significant improvement of the paper.

**References**


Department of Biostatistics, School of Public Health, Los Angeles, CA 90095-1772.

E-mail: gangli@sunlab.ph.ucla.edu

(Received March 2001; accepted July 2002)