Functional inference for interval-censored data in proportional odds model with covariate measurement error
FUNCTIONAL INFEERENCE FOR INTERVAL-CENSORED DATA IN PROPORTIONAL ODDS MODEL WITH COVARIATE MEASUREMENT ERROR

Chi-Chung Wen and Yi-Hau Chen

Tamkang University and Academia Sinica

Abstract: It is common in regression analysis of failure time data, such as the AIDS Clinical Trial Group (ACTG) 175 clinical trial data, that the failure time (AIDS incidence time) is subject to interval-censoring and the covariate (baseline CD4 count) is subject to measurement error. To perform valid analysis in this setting, we propose a functional inference method under the semiparametric proportional odds model. The new proposal utilizes the working independence strategy to handle general mixed case interval censorship, as well as the conditional score approach to handle mismeasured covariate without specifying the covariate distribution. The asymptotic theory, together with a stable computation procedure combining the Newton-Raphson and self-consistency algorithms, is established for the proposed estimation method. We illustrate the nice performance of the new proposal via simulation studies and analysis of ACTG 175 data.

Key words and phrases: Conditional score, Interval-censoring, Measurement error, Semiparametric, Survival analysis.

1. Introduction

Interval-censored failure time data are commonly encountered in medical studies in which the failure time of interest cannot be observed exactly but is known to fall in a time interval obtained from a sequence of examinations. Regression analysis for assessing the effects of covariates on the interval-censored failure time has been widely studied by, e.g., Huang and Wellner (1997), Huang and Rossini (1997), Betensky et al. (2001), and was comprehensively reviewed by Zhang and Sun (2010). These existing methods, however, are restricted to the setting where the covariates are accurately measured. Such a restriction hampers general applications of these methods given that the covariates are often measured with error. For example, in the AIDS Clinical Trial Group (ACTG)
175 clinical trial on HIV-infected patients (Hammer et al. (1996)), the effects of baseline CD4 cell counts and treatments (zidovudine alone, zidovudine plus didanosine, zidovudine plus zalcitabine, and didanosine alone) on time to the incidence of AIDS are of interest. However, data on the AIDS onset time are determined at intermittent clinic visits and hence subject to interval-censoring. In addition, measurements of the baseline CD4 counts are subject to error due to both instrumental contamination and biological variation (Song and Ma (2008)). In this application, it would be invalid to apply directly the methods mentioned above with the error on the CD4 count measurements fully ignored.

While there is considerable work on the covariate measurement error problem for right-censored failure time data (see, e.g., Hu et al. (1998), Tsiatis and Davidian (2001), Song and Huang (2005)), regression analysis for interval-censored data with mismeasured covariate has so far been less studied. Song and Ma (2008) used a multiple augmentation method to convert interval-censored data into right-censored data and then applied the conditional score method to the converted data. Wen (2012) proposed a full-likelihood method in this problem. Both the two methods rely on a parametric specification for the distribution of the unobserved error-prone covariate, hence belong to the structural modeling approach in the measurement error literature (Carroll et al. (2006)). Correctly specifying a parametric model for the covariate distribution may be difficult given that the covariate has been mismeasured, and misspecification of the covariate distribution can result in biased estimates for the regression parameters. Alternatively, a functional modeling method which makes no distributional assumption for the error-prone covariate was considered by Wen and Chen (2012), but only in the specific “case 1” interval-censoring setting wherein there is only one examination time for each subject.

In this article we describe a functional inference method under the semiparametric proportional odds model for interval-censored failure time data with mismeasured covariates. The semiparametric proportional odds model is considered since it is a flexible and popular model in regression analysis of failure time data, and it allows for a particularly effective application of the well-known conditional score approach for dealing with covariate measurement error. The new method extends the conditional score approach of Wen and Chen (2012) from
case 1 to general interval-censored data, by treating multiple examinations from
the same subjects as single examinations from different subjects, and then ap-
plying the conditional score correction method of Wen and Chen (2012) to each
examination. Dependence among different examinations of the same subject is
accounted for in standard error estimation. This idea of working independence
for interval-censored data has been adopted by Betensky et al. (2001) and Zhu et
al. (2008). As mentioned in Betensky et al. (2001), underlying such an approach
is the assumption of independence between the examination times and the failure
time given covariates, which is ensured when the examinations continue to occur
regardless of whether the failure has occurred, as observed in the ACTG 175
data. In general, data of this type arise in settings in which there are endpoints
of secondary interest, for which examinations continue even after the occurrence
of the primary endpoint.

This paper is organized as follows. Section 2 introduces the data structure
and the model. Section 3 describes the proposed working independence condi-
tional score method, and the theoretical properties of the estimator. Section 4
presents the computation algorithm. Section 5 evaluates the proposed method
through simulation studies and an analysis of the ACTG 175 data. Section 6
contains concluding remarks. Technical proofs of the asymptotic properties are
given in the Appendix.

2. The Conditional Score Estimator

Let $T$ and $(X,Z)'$ denote the failure time and covariate vector for a sub-
ject, where $Z$ is error-free and $X$ is error-prone. We assume $X$ is univariate
for brevity of derivation, but the idea is extendable to the case of multivariate
$X$. Instead of exact measurement of $X$, its replicated surrogate measurement
$W = (W_1, \ldots, W_m)$ is available with $W_j = X + e_j$, $j = 1, \ldots, m$, where the
measurement errors $e_j$’s are $N(0, \sigma^2)$ distributed, independently of each other
and of $X$. Throughout this work, we consider the mixed case interval-censoring
(Shick and Yu (2000)), where $T$ is not observed but is monitored by a triangle
array of random examination times $U = \{U_{K,l} : l = 1, \ldots, K, K = 1,2, \ldots\}$,
with $U_{K,1} < \ldots < U_{K,K}$ and the number of examinations $K$ being random. As-
sume that the examinations continue to occur regardless of whether the failure
has occurred, so that the examination times $(K,U)$ and the failure time $T$ are
independent given covariates \((X, Z)\). The variables we observe for one subject are thus \(O = \{(K, U_{K,l}, \Delta_{K,l}, W, Z) : l = 1, \ldots, K\}\), where \(\Delta_{K,l} = I(T \leq U_{K,l})\) indicates whether the failure time \(T\) precedes the examination time \(U_{K,l}\).

Given covariates \((X, Z)\), the failure time \(T\) is assumed to follow the proportional odds model, namely at time \(t\) the conditional survival function of \(T\) given covariates is of the form

\[
\Pr(T > t|X = x, Z = z) = \left\{ 1 + \exp(\beta_1 x + \beta_2^t z + H(t)) \right\}^{-1}, \tag{2.1}
\]

where \(\beta = (\beta_1, \beta_2')^t\) is an unknown regression vector and \(H\) is an unspecified, nondecreasing and continuous baseline log odds function. Assume that \((T, K, U)\) and \(W\) are conditionally independent given \((X, Z)\), i.e. the surrogate condition, and the conditional distribution of \((K, U)\) given \((X, Z)\) does not depend on parameters of interest. Let \(W = \sum_{j=1}^m W_j/m\) and \(\bar{\sigma}^2 = \sigma^2/m\). Then the conditional likelihood of \((\Delta_{K,l}, W)\), given \((K, U_{K,l}, X, Z)\) is proportional to

\[
\exp\{\Delta_{K,l}(\beta_1 X + \beta_2 Z + H(U_{K,l}))\} \over [1 + \exp\{\beta_1 X + \beta_2 Z + H(U_{K,l})\}][(\bar{\sigma}^2)^{1/2}] \exp\left\{ (-\bar{W} - X)^2 \over 2\bar{\sigma}^2 \right\} \\
= \exp\{\Delta_{K,l}(\beta_2 Z + H(U_{K,l}))\} \over [1 + \exp\{\beta_1 X + \beta_2 Z + H(U_{K,l})\}][(\bar{\sigma}^2)^{1/2}] \exp\left\{ (-\bar{W}^2 + X^2) \over 2\bar{\sigma}^2 \right\} \exp \left\{ X S_{K,l} \over \bar{\sigma}^2\right\},
\]

where \(S_{K,l} = S_{K,l}(\beta, \sigma^2) = \beta_1 \Delta_{K,l} \bar{\sigma}^2 + \bar{W}\) is a complete sufficient statistic for \(X\). Thus the conditional probability of \(\Delta_{K,l} = 1\) given \((K, U_{K,l}, S_{K,l}, Z)\) is

\[
\mathcal{E}_{K,l}(\theta)(O) = \exp\{\beta_1 S_{K,l} - \beta_2^2 \bar{\sigma}^2 / 2 + \beta_2^t Z + H(U_{K,l})\} \over [1 + \exp\{\beta_1 S_{K,l} - \beta_2^2 \bar{\sigma}^2 / 2 + \beta_2^t Z + H(U_{K,l})\}]
\]

where \(\theta = (\beta, H, \sigma^2)\).

The approach to estimation we proposed is based on treating examinations from the same subjects as if they were single examinations from different subjects. Under this "working independence" assumption, the conditional likelihood of \((\Delta_{K,1}, \ldots, \Delta_{K,K})\) given \(\{(K, U_{K,l}, S_{K,l}, Z) : l = 1, \ldots, K\}\) would take the form

\[
L(\theta)(O) = \prod_{l=1}^K \mathcal{E}_{K,l}(\theta)(O) \Delta_{K,l} \{1 - \mathcal{E}_{K,l}(\theta)(O)\}^{1-\Delta_{K,l}}. \tag{2.2}
\]

Let \(O_i = \{(K_{i,l}, U_{K,l}^{(i)}, \Delta_{K,l}^{(i)}, W_i, Z_i) : l = 1, \ldots, K_i\}, i = 1, \ldots, n\), be \(n\) i.i.d. copies of observed variable \(O\). Suppose that \(W_i = (W_{i1}, \ldots, W_{im_i})\) and define...
\begin{equation}
\bar{W}_i = \sum_{j=1}^{m_i} W_{ij}/m_i, \ \hat{\sigma}^2 = \sigma^2/m_i, \ \text{and} \ S_{K,l}^{(i)} = \beta_1 \Delta_{(K,l)}^{(i)} \hat{\sigma}^2 + \bar{W}_i \ \text{for} \ l = 1, \ldots, K_i.
\end{equation}

Following the idea of Stefanski and Carroll (1987), we can construct a conditional score (CS) estimator for \( \beta \) by solving the estimating equation

\begin{equation}
\sum_{i=1}^{n} \ell_1(\theta)(O_i) = 0,
\end{equation}

where \( \ell_1(\theta)(O) = \sum_{i=1}^{K} (S_{K,l} - \hat{\beta}_1 \hat{\sigma}^2, Z') \{ \Delta_{K,l} - E_{K,l}(\theta)(O) \} \) is obtained by differentiating the logarithm of conditional likelihood (2.2) with respect to \( \beta \), ignoring the dependence of \( S_{K,l} \)'s on \( \beta \).

For fixed \( \beta \) and \( \sigma^2 \), we propose to estimate the baseline log odds function \( H \) by maximizing the conditional likelihood

\begin{equation}
L_n(\theta)(O_1, \ldots, O_n) = \prod_{i=1}^{n} L(\theta)(O_i).
\end{equation}

It is easy to see that (2.4) depends on \( H \) only through \( \{ H(U_{K,i,l}^{(i)}) : l = 1, \ldots, K_i, i = 1, \ldots, n \} \). Therefore, in maximizing \( L_n \) we treat \( H \) as a nondecreasing step function with possible jumps only at the examination times \( U_{K,i,l}^{(i)} \)'s.

Usually the error variance \( \sigma^2 \) is also unknown and must be estimated. Based on the replicated measurement data \( \{ W_i : i = 1, \ldots, n \} \), \( \hat{\sigma}^2 \) can be consistently estimated by \( \hat{\sigma}^2 = \sum_{i,j} (W_{ij} - \bar{W}_i)^2/ \sum_i (m_i - 1) \), the solution of the estimating equation

\begin{equation}
\sum_{i=1}^{n} \phi(\sigma^2)(W_i) = 0,
\end{equation}

where \( \phi(\sigma^2)(W_i) = \sum_{i,j} (W_{ij} - \bar{W}_i)^2 - (m_i - 1)\sigma^2 \).

In summary, the proposed estimation procedure consists of the following two steps:

1. Obtain the estimate \( \hat{\sigma}^2 \) for the variance \( \sigma^2 \) of measurement error, using (2.5).

2. Substitute \( \hat{\sigma}^2 \) for \( \sigma^2 \) in (2.3) and (2.4), obtain the conditional score estimate \( (\hat{\beta}, \hat{H}) \) for \( (\beta, H) \) by solving (2.3) and maximizing (2.4).

We detail in Section 4 the computation algorithm used in step 2. In step 1, to obtain a consistent estimate for the error variance \( \sigma^2 \), we require replicates of
that the true value \( \theta \) evaluated at \( \beta, H \) is of order \( n^{1/2} \), assuming temporarily that \( \hat{\beta} \) is consistent for \( (\beta_0, H_0) \), the true value of \( (\beta, H) \). The convergence rate of \( \hat{\beta} \) is of order \( n^{1/2} \) but that of \( \hat{H} \) is of order \( n^{1/3} \) only, which are rates as obtained in general semiparametric analysis of interval-censored data; see, e.g., Huang (1996). Assuming temporarily that the true value \( \sigma^2 \) of error variance is known, we have approximately

\[
\sqrt{n}(\hat{\beta} - \beta_0) \overset{d}{=} \left[ -E \left\{ \frac{\partial}{\partial \beta} \ell^*(\theta_0) \right\} \right]^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \ell^*_i(\theta_0),
\]

where \( \theta_0 = (\beta_0, H_0, \sigma^2_0) \), \( \ell^*(\theta)(O) = \sum_{i=1}^{K} \left\{ (S_{K,l} - \beta_l \hat{\sigma}^2, Z')' - g^*(U_{K,l}) \right\} \{ \Delta_{K,l} - \varepsilon_{K,l}(\theta)(O) \} \}, \ell^*_i(\theta) = \ell^*(\theta)(O_i) \), and \( g^* \) is given by

\[
g^*(u) = \frac{\sum_{k=1}^{\infty} \sum_{i=1}^{K} f_{K,l}(k, u) E_{k,u} \left\{ (S_{K,l} - \beta_l \hat{\sigma}^2, Z')' \varepsilon_{K,l}(\theta)(O) \right\}}{\sum_{k=1}^{\infty} \sum_{i=1}^{K} f_{K,l}(k, u) E_{k,u} \left\{ \varepsilon_{K,l}(\theta)(O) \right\}}, \tag{3.1}
\]
evaluating at \( \theta = \theta_0 \) with \( \varepsilon_{K,l}(\theta) = \varepsilon_{K,l}(\theta)(1 - \varepsilon_{K,l}(\theta)) \}, f_{K,l} \) the density of \( (K, U_{K,l}) \) and \( E_{k,u}(\cdot) = E(\cdot | K = k, U_{K,l} = u) \), the conditional expectation given \( K = k, U_{K,l} = u \).

Recall that \( \sigma^2 \) is unknown in general and can be estimated through (2.5). Write \( \varphi_i(\sigma^2) = \varphi(\sigma^2)(W_i) \) in (2.5). To account for the extra estimation of \( \sigma^2 \), we thus have approximately

\[
\sqrt{n} \left[ \frac{\hat{\beta} - \beta_0}{\sigma^2 - \sigma^2_0} \right] \overset{d}{=} \left[ I^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \ell^*_i(\theta_0) \right] \left[ \varphi_i(\sigma^2_0) \right],
\]
which is asymptotically distributed as $N(0, \mathcal{I}^{-1}\Sigma(\mathcal{I}^{-1})')$, where $\Sigma$ is the covariance matrix of $(\ell^*(\theta_0)', \varphi(\sigma_0^2))'$, and
\[
\mathcal{I} = -E\begin{bmatrix}
\frac{\partial}{\partial \beta} \ell^*(\theta_0) & \frac{\partial}{\partial \sigma^2} \ell^*(\theta_0) \\
0 & \frac{\partial}{\partial \sigma^2} \varphi(\sigma_0^2)
\end{bmatrix}.
\] (3.2)

The asymptotic variance can be estimated by replacing the population quantities with their empirical counterparts. In particular, according to Huang (1996), the quantity $g^*$ can be estimated by using the nonparametric regression technique with kernel smoothing. Based on the asymptotic normality and the estimated variance, hypothesis testing and confidence interval for the regression parameter $\beta$ can be simply performed. For example, the test of the significance of $\beta$, i.e., the test of $H_0: \beta = 0$, can be performed via the Wald test statistic $\hat{\beta}'\{\hat{\text{var}}(\hat{\beta})\}^{-1}\hat{\beta}$, which has an asymptotic $\chi^2$ distribution with $d = \dim(\beta)$ degrees of freedom under $H_0$, where $\hat{\text{var}}(\hat{\beta})$ is the submatrix of $\hat{\mathcal{I}}^{-1}\hat{\Sigma}(\hat{\mathcal{I}}^{-1})'$ corresponding to $\hat{\beta}$.

4. Computation Algorithm

The parameter $H$ is non-parametric and its size is of the order of the sample size. The maximization of the conditional likelihood is thus a high-dimensional optimization problem. We propose a self-consistency algorithm for estimation of $H$, which modifies the algorithm in Wen and Chen (2012) developed under case 1 interval-censoring.

Let $u_1 < \ldots < u_N$ denote the distinct ordered values of $\{U_{K_i,l}^{(i)}: l = 1, \ldots, K_i, i = 1, \ldots, n\}$. The function $H$ can be expressed as its jump sizes by $h = (h_1, \ldots, h_N)'$, where $h_1 = H(u_1)$ and $h_j = H(u_j) - H(u_{j-1})$ is the jump size of $H$ at $u_j$ for $j \geq 2$. We treat all the $h_j$’s as non-zero parameters even though some of them may approach 0, and obtain the estimate $\hat{h}_j$ for $h_j$ by differentiating the conditional likelihood (2.4) with respect to $h_j$, $j = 1, \ldots, N$. Accordingly, for fixed $(\beta, \sigma^2)$, $\hat{h} = (\hat{h}_1, \ldots, \hat{h}_N)'$ is obtained as the solution to the system of equations
\[
\frac{\partial}{\partial h_j} \log L_n(\beta, \hat{h}, \sigma^2) = \sum_{i=1}^{n} \sum_{l=1}^{K_i} \Delta_{K_i,l}^{(i)} I[U_{K_i,l}^{(i)} \geq u_j] - \sum_{i=1}^{n} \sum_{l=1}^{K_i} \mathcal{E}_{K_i,l}(\beta, \hat{h}, \sigma^2)(O_i) I[U_{K_i,l}^{(i)} \geq u_j] = a_j - b_j(\beta, \hat{h}, \sigma^2)
\] (4.1)
By definition in (4.1), \( a_j = b_j(\beta, \hat{h}, \sigma^2) \), and \( a_j + M_0 = b_j(\beta, \hat{h}, \sigma^2) + M_0 \) for any constant \( M_0 \). For fixed \( \beta \) and \( \sigma^2 \), we thus consider the self-consistency algorithm for solving \( \hat{h} \):

\[
h_j^{(k+1)} = D_j(\beta, h^{(k)}, \sigma^2), \quad j = 1, \ldots, N,
\]

with

\[
D_j(\beta, h^{(k)}, \sigma^2) = h_j^{(k)} \cdot \left\{ \frac{a_j + M_0}{b_j(\beta, h^{(k)}, \sigma^2) + M_0} \right\}, \quad j = 2, \ldots, N,
\]

\[
D_1(\beta, h^{(k)}, \sigma^2) = h_1^{(k)} + \log \left\{ \frac{a_1 + M_0}{b_1(\beta, h^{(k)}, \sigma^2) + M_0} \right\},
\]

where the superscript \( (k) \) denotes the \( k \)th iteration of the algorithm. It is easy to check that if \( h^{(k)} = \hat{h} \) in (4.2) then \( h^{(k+1)} = \hat{h} = h^{(k)} \), that is, the estimate \( \hat{h} \) is a fixed point of the algorithm. The non-negative constant \( M_0 \) is used here to aid convergence of the algorithm; more explanation is given below. In fact, the value of \( M_0 \) is not crucial and the convenient choice of \( M_0 = 0 \) usually works well in our numerical studies.

For fixed \( (H, \sigma^2) \), the CS estimate for \( \beta \) can be obtained by traditional methods such as the Newton-Raphson algorithm. The estimate \( \hat{\sigma}^2 \) for \( \sigma^2 \) can be obtained separately by solving (2.5). We therefore compute the CS estimate \( (\hat{\beta}, \hat{h}) \) via a hybrid algorithm, consisting of a Newton-Raphson algorithm for solving \( \beta \) and a self-consistency algorithm for solving \( h \). The hybrid algorithm proposed is depicted in the following. Starting from initial values \( \beta^{(0)} \) and \( h^{(0)} \) and fixing \( \sigma^2 = \hat{\sigma}^2 \) throughout, for \( k = 0, 1, \ldots \), we iterate between steps 1 and 2 below until some convergence criterion is met:

1. fix \( h = h^{(k)} \), update \( \beta^{(k)} \) to \( \beta^{(k+1)} \) by solving (2.3) with the one-step Newton-Raphson algorithm;

2. fix \( \beta = \beta^{(k+1)} \), update \( h^{(k)} \) to \( h^{(k+1)} \) by (4.2).

Remark 2. As other conditional score methods, the proposed conditional score estimating equation may have multiple solutions. To better locate the consistent solution, as in Tsiatis and Davidian (2001), we solve (2.3) with the initial \( (\beta, H) \)
values given by the naive estimates maximizing the standard likelihood (5.1) where the true covariate is imputed by the mean of the surrogate measurements. This strategy works well in our numerical studies. Another feasible strategy is to choose the consistent solution as the one minimizing the least squares or other goodness-of-fit criteria; see Heyde and Morton (1998) for details.

**Remark 3.** The role of $M_0$ in the algorithm is explained as follows. Since $a_l$, $b_l$, and $e_l(\beta, h, \sigma^2) \equiv (\partial/\partial h_l)b_l(\beta, \hat{h}, \sigma^2)$ are all positive, and $a_l = b_l(\beta, \hat{h}, \sigma^2)$, a sufficiently large $M_0$ ensures $|\partial D_1/\partial h_1| = |1 - e_1/(b_1 + M_0)| \in (0, 1)$ and $|\partial D_l/\partial h_l| = |(a_l + M_0)/(b_l + M_0) - h_l e_l(a_l + M_0)/(b_l + M_0)^2| \in (0, 1)$ for $l \geq 2$, when $h$ is near $\hat{h}$. This means that the self-consistency algorithm is locally contractive and will converge by the contraction principle (Rudin (1973)).

**Remark 4.** By definition $h_1$ may be negative while $h_j$ are positive for $j \geq 2$. To accommodate this, the type of iterative equation used for $h_j$ for $j \geq 2$ is applied to $e^{h_1}$ to obtain the iterative equation for $h_1$, as shown in equations below (4.2).

### 5. Numerical Studies

#### 5.1 Simulations

We perform simulation studies to assess the performance of the proposed conditional score estimator and examine the adequacy of the normal approximation.

In simulations, the error-free covariate $Z$ in model (2.1) is Bernoulli(0.5) and the error-prone covariate $X$ is $N(0, 1)$. The true regression coefficient $(\beta_{10}, \beta_{20})$ is set to be $(0.5, -0.5)$ or $(1, -0.5)$, and the baseline log odds $H_0(t)$ is taken to be $\log(e^t - 1)$. Two surrogate measurements $W = (W_1, W_2)$ of $X$ are made per subject with error variance $\sigma^2 = 0.25, 0.5$ or 0.75. The number of examinations $K$ is randomly selected from \{3, 4, 5\}, and given $K$, the examination time points $U_{K,1} < \ldots < U_{K,K}$ are generated to be the ordered statistics of a random sample of size $K$ from Uniform(0, 1). The sample size $n = 150$ or 300, and the simulation replication is 400 in each study.

To evaluate the performance of the CS estimator, the bias, standard deviation (SD), average of estimated standard errors (ASE), and the coverage probability of the 95% confidence intervals (CP) are calculated over simulation replicates and summarized in Table 1. Note that, since the simulation replication is 400,
if the true coverage is 95%, then 80% of the simulations would have simulated coverage between 93.5% and 96.25%.

For comparison, Table 1 also includes results from the naive analysis, which naively substitutes the mean of the surrogate measurements \( \bar{W} = (W_1 + W_2)/2 \) for the true covariate \( X \) in the standard proportional odds regression analysis. Namely, the naive estimator maximizes over the parameter \((\beta, H)\) the likelihood

\[
\tilde{L}(\beta, H) = \prod_{i=1}^{n} \prod_{l=1}^{K_i+1} \left\{ F_i(U_{K_i,l-1}^{(i)}) - F_i(U_{K_i,l}^{(i)}) \right\} \tilde{\Delta}_{K_i,l}^{(i)}
\]

with \( F_i(t) = \{1 + \exp(\beta_1 \bar{W}_i + \beta_2 Z_i + H(t))\}^{-1} \), where \( U_{K_i,0}^{(i)} = 0, U_{K_i,K_i+1}^{(i)} = \infty \), and \( \tilde{\Delta}_{K_i,l}^{(i)} \equiv I[U_{K_i,l}^{(i)} < T_i \leq U_{K_i,l+1}^{(i)}] \).

Results in Table 1 indicate that the CS estimator performs well in the finite sample setting considered. The bias of the CS is fairly small compared with the standard deviation, and it decreases further with decreases in the covariate effect or error variance, or with increases in the sample size. The proposed standard error estimate is adequate and close to simulation standard deviation. The normal approximation works well, as reflected in the correct coverage probabilities of the resulting confidence intervals. In the simulation scenario with \((\beta_1, \beta_2) = (1, -0.5), \sigma^2 = 0.5, \) and \( n = 300 \), Figure 1 shows the Q-Q plots comparing the standardized CS estimates \((\hat{\text{var}}(\hat{\beta}))^{-1/2}(\hat{\beta} - \beta_0)\) with the standard normal variate, which agrees well with a straight line and hence further reveals the adequacy of the normal approximation theory for the CS estimator. In contrast, from Table 1 we understand that the naive estimator, obtained by using the mean surrogate values in the standard proportional odds analysis, performs poorly for \( \beta_1 \), the coefficient corresponding to the error-prone covariate.

We have also conducted additional simulations with \( X \) following a uniform distribution, i.e., \( X \sim \sqrt{3}U(-1, 1) \), and other specifications unchanged. The results are quite similar to those presented above, hence are omitted here.

### 5.2 Application to ACTG 175 data

We apply the proposed inference procedure to ACTG 175 data introduced in Section 1. The primary goal of the analysis is to address the effects of the baseline CD4 count and treatments on the time to the AIDS incidence in antiretroviral-naive patients. To this end, we consider a proportional odds model with two
### Table 5.1: Simulation Results

(a) Conditional score method

<table>
<thead>
<tr>
<th>$(\beta_{10}, \beta_{20})$</th>
<th>$\sigma_{20}^2$</th>
<th>$n$</th>
<th>$\beta_1$ estimate</th>
<th>$\beta_2$ estimate</th>
<th>$\beta_1$ estimate</th>
<th>$\beta_2$ estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(0.5, -0.5)$</td>
<td>0.25</td>
<td>150</td>
<td>0.025</td>
<td>0.222</td>
<td>0.215</td>
<td>0.195</td>
</tr>
<tr>
<td></td>
<td></td>
<td>300</td>
<td>0.017</td>
<td>0.135</td>
<td>0.136</td>
<td>0.136</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>150</td>
<td>0.028</td>
<td>0.240</td>
<td>0.210</td>
<td>0.200</td>
</tr>
<tr>
<td></td>
<td></td>
<td>300</td>
<td>0.018</td>
<td>0.144</td>
<td>0.146</td>
<td>0.146</td>
</tr>
<tr>
<td></td>
<td>0.75</td>
<td>150</td>
<td>0.033</td>
<td>0.263</td>
<td>0.228</td>
<td>0.228</td>
</tr>
<tr>
<td></td>
<td></td>
<td>300</td>
<td>0.019</td>
<td>0.154</td>
<td>0.156</td>
<td>0.156</td>
</tr>
<tr>
<td>$(1, -0.5)$</td>
<td>0.25</td>
<td>150</td>
<td>0.073</td>
<td>0.258</td>
<td>0.238</td>
<td>0.238</td>
</tr>
<tr>
<td></td>
<td></td>
<td>300</td>
<td>0.040</td>
<td>0.168</td>
<td>0.165</td>
<td>0.165</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>150</td>
<td>-0.054</td>
<td>0.175</td>
<td>0.165</td>
<td>0.165</td>
</tr>
<tr>
<td></td>
<td></td>
<td>300</td>
<td>-0.054</td>
<td>0.113</td>
<td>0.114</td>
<td>0.114</td>
</tr>
<tr>
<td></td>
<td>0.75</td>
<td>150</td>
<td>-0.142</td>
<td>0.153</td>
<td>0.147</td>
<td>0.147</td>
</tr>
<tr>
<td></td>
<td></td>
<td>300</td>
<td>-0.142</td>
<td>0.100</td>
<td>0.102</td>
<td>0.102</td>
</tr>
</tbody>
</table>

(b) Naive method

<table>
<thead>
<tr>
<th>$(\beta_{10}, \beta_{20})$</th>
<th>$\sigma_{20}^2$</th>
<th>$n$</th>
<th>$\beta_1$ estimate</th>
<th>$\beta_2$ estimate</th>
<th>$\beta_1$ estimate</th>
<th>$\beta_2$ estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(0.5, -0.5)$</td>
<td>0.25</td>
<td>150</td>
<td>-0.054</td>
<td>0.175</td>
<td>0.165</td>
<td>0.165</td>
</tr>
<tr>
<td></td>
<td></td>
<td>300</td>
<td>-0.054</td>
<td>0.113</td>
<td>0.114</td>
<td>0.114</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>150</td>
<td>-0.103</td>
<td>0.163</td>
<td>0.155</td>
<td>0.155</td>
</tr>
<tr>
<td></td>
<td></td>
<td>300</td>
<td>-0.102</td>
<td>0.106</td>
<td>0.107</td>
<td>0.107</td>
</tr>
<tr>
<td></td>
<td>0.75</td>
<td>150</td>
<td>-0.142</td>
<td>0.153</td>
<td>0.147</td>
<td>0.147</td>
</tr>
<tr>
<td></td>
<td></td>
<td>300</td>
<td>-0.141</td>
<td>0.100</td>
<td>0.102</td>
<td>0.102</td>
</tr>
<tr>
<td>$(1, -0.5)$</td>
<td>0.25</td>
<td>150</td>
<td>-0.102</td>
<td>0.182</td>
<td>0.190</td>
<td>0.190</td>
</tr>
<tr>
<td></td>
<td></td>
<td>300</td>
<td>-0.119</td>
<td>0.128</td>
<td>0.130</td>
<td>0.130</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>150</td>
<td>-0.208</td>
<td>0.168</td>
<td>0.176</td>
<td>0.176</td>
</tr>
<tr>
<td></td>
<td></td>
<td>300</td>
<td>-0.222</td>
<td>0.120</td>
<td>0.121</td>
<td>0.121</td>
</tr>
<tr>
<td></td>
<td>0.75</td>
<td>150</td>
<td>-0.292</td>
<td>0.157</td>
<td>0.165</td>
<td>0.165</td>
</tr>
<tr>
<td></td>
<td></td>
<td>300</td>
<td>-0.304</td>
<td>0.113</td>
<td>0.113</td>
<td>0.113</td>
</tr>
</tbody>
</table>
covariates, log(CD4) (X) and a treatment indicator (Z = 1 for zidovudine alone and 0 for any of the other three therapies). The surrogates (W₁, W₂) for the error-prone covariate X are defined to be the last two measurements of the CD4 count observed prior to treatment, and are standardized to have mean 0 and variance 1. As a crude diagnosis of the normality assumption for the measurement error, Figure 2 depicts the Q-Q plot of the deviations from the average of standardized log(CD4) versus a standard normal variate, showing the measurement error is nearly normally distributed. The mean number of examination times K per subject is 9.06 with the range 1–15 among the total 1014 patients. The proposed conditional score method is applied to account for both the measurement error and interval-censoring present in this data.

The results of applying both the CS and the naive methods are given in Table 2. As expected, the baseline CD4 count is negatively associated with the incidence time of the AIDS. The zidovudine alone treatment is significantly worse than the other three treatments in preventing the onset of AIDS. Compared with the proposed CS estimate, the method naively using the mean surrogate CD4
Table 5.2: Analysis of ACTG data

<table>
<thead>
<tr>
<th>Method</th>
<th>log(CD4)</th>
<th>treatment (zidovudine)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>S.E.</td>
</tr>
<tr>
<td>CS</td>
<td>-1.0669</td>
<td>0.1907</td>
</tr>
<tr>
<td>Naive</td>
<td>-0.7776</td>
<td>0.1452</td>
</tr>
</tbody>
</table>

Figure 5.2: Q-Q plots of the deviations from the mean of standardized log(CD4).

measurements yields a remarkably attenuated estimates for the effects of true baseline CD4 count and treatment. The estimate of the error variance $\sigma^2$ given by the CS method is 0.2928 with a standard error of 0.0174 (on the standardized log(CD4) scale).

*Remark 5.* We use the last two measurements prior to treatment as two replicates of the surrogate baseline CD4 count. This is done because for all study subjects, their last two CD4 count measurements prior to treatment were measured consecutively within one month, and hence the two CD4 measurements were close with the relative difference, $|W_2 - W_1|/W_1 < 30\%$ for more than 80\% of the study subjects.

6. Conclusion

Motivated by the well-known AIDS Clinical Trial Group 175 data, where the failure time (AIDS incidence time) is subject to interval-censoring and the
covariate (baseline CD4 count) is subject to measurement error, we have developed a functional inference method under the semiparametric proportional odds model for failure time regression analysis. To accommodate general mixed case interval censorship as well as mismeasured covariates, while avoiding the need to specify the covariate distribution, we utilize the strengths from the conditional score approach proposed by Wen and Chen (2012) as well as the working independence idea similar to that in Betensky et al. (2001). To our best knowledge, formal statistical methodologies applicable in the same setting are still lacking in the literature. Results from simulation studies and an analysis of ACTG 175 data reveals the utilities of the proposed method.

The current work stimulates further research topics in the area of interval-censored failure time data with mismeasured covariates. First, it is quite promising to extend the proposed method to more general regression models than the proportional odds model, such as the semiparametric transformation model (Zeng and Lin (2007)) and additive hazards model (Zeng et al. (2006)). Second, to improve estimation efficiency, it is worthwhile to develop a functional modeling approach without the working independence assumption for the examination times in the same subject.

Acknowledgment The authors are grateful to the very helpful comments from the Associate Editor and Referees. This research is supported by National Science Council of Republic of China (NSC 101-2118-M-032-005-MY2 and NSC 101-2119-M-001-002-MY3).

Appendix

Assumptions

Let \( P_0 \) denote the true underlying distribution, \( P_n f = \sum_{i=1}^{n} f(O_i)/n \) and \( P_0 f = Ef(O) \) for a measurable function \( f \). Consider \( H \) in \( \mathcal{H} \), the set of right-continuous non-decreasing functions that are uniformly bounded on the study period \([0, \tau]\).

For simplicity the proofs are presented under the simpler setting where the distribution of \((K, U_{K,l}, l = 1, \ldots, K)\) is independent of \((W, Z)\), though the proposed method can allow the dependence case. Let \( \ell(\theta) = \log L(\theta) \). The asymptotic theories are based on the following regularity assumptions, which have been similarly made in the context of interval-censoring studies (e.g., Huang and Wellner (1997), Zeng et al. (2006), Ma (2010)).
(C1) There exists a positive $\xi$ such that $P(U_{K,l} - U_{K,l-1} \geq \xi) = 1$ for $l = 1, \ldots, K$.

(C2) Given $K$, each $U_{K,l}$ has a continuous density and the union of the supports for conditional distributions $\{U_{K,l}, l = 1, \ldots, K\}$ is an interval $[\tau_1, \tau_2]$ with $0 < \tau_1 < \tau_2 < \tau$.

(C3) The true parameter $(\beta_0, \sigma_0^2)$ lies in the interior of a compact parameter set $B \times Q$; $H_0$ is continuously differentiable on $[\tau_1, \tau_2]$ and satisfies $-M < H_0(\tau_1) < H_0(\tau_2) < M$.

(C4) For $\theta$ near $\theta_0$, $P_0\{\ell(\theta) - \ell(\theta_0)\} \leq -\{\|H - H_0\|_2^2 + \|\beta - \beta_0\|^2 + \|\sigma^2 - \sigma_0^2\|^2\}$, where $\|\cdot\|$ is the Euclidean norm, $\|H\|_2^2 = \int \sum_{k=1}^{\infty} \sum_{l=1}^{K} f_{K,l}(k,u)H^2(u)du$, and $f_{K,l}$ denotes the density of $(K, U_{K,l})$. The notation $\leq$ means smaller than, up to a constant.

(C5) The function $g^*$ given in (3.1) is differentiable with a bounded derivative on $[\tau_1, \tau_2]$.

(C6) The information matrix $I$ defined in (3.2) is invertible.

Note that the condition (C1) rules out accurately observed failure time and makes the number of examination times $K$ bounded. Below we establish the consistency and rate of convergence of proposed CS estimators.

**Theorem 1** (Consistency and rate of convergence)

The estimator $\hat{\beta}$ is consistent; that is, $\hat{\beta} \xrightarrow{p} \beta_0$. The rate of convergence of $\hat{H}$ is of order $n^{-1/3}$; that is, $\|\hat{H} - H_0\|_2 = O_p(n^{-1/3})$. The consistency of $\hat{H}$ can be obtained from its rate of convergence.

**Proof.** Denote $\hat{H}_{(\beta, \sigma^2)}$ the maximizer of $L_n$ with $(\beta, \sigma^2)$ fixed. We first apply Theorem 5.7 of van der Vaart (1998) to establish the consistency of $\hat{H}_{(\beta_0, \sigma_0^2)}$. Define $w(\theta) = \log\{[L(\theta) + L(\theta_0)]/2\}$. Since the class of monotone and uniformly bounded functions is a Donsker class, by Theorem 2.10.6 of van der Vaart and Wellner (1996), we know that the class $\{w(\beta_0, H, \sigma_0^2) | H \in \mathcal{H}\}$ is Donsker and hence Glivenko-Cantelli. Further, by the concavity of $r(u) \equiv \log((u + 1)/2)$ and Jensen’s inequality, we have

$$P_0[w(\beta_0, H, \sigma_0^2) - w(\theta_0)] = P_0r\left(\frac{L(\beta_0, H, \sigma_0^2)}{L(\theta_0)}\right) \leq r\left(P_0\left(\frac{L(\beta_0, H, \sigma_0^2)}{L(\theta_0)}\right)\right) = 0,$$
and the equality holds only if \( H = H_0 \) on \((\tau_1, \tau_2)\). This indicates that

\[
\sup_{\|H - H_0\|_2 > \epsilon} P_0 w(\beta_0, H, \sigma^2_0) < P_0 w(\theta_0).
\]

Furthermore, note that

\[
P_n w(\beta_0, \hat{H}(\beta_0, \sigma^2), \sigma^2_0) + o_p(1) = P_n w(\beta_0, \hat{H}(\beta_0, \sigma^2), \hat{\sigma}^2) \geq P_n w(\beta_0, H_0, \hat{\sigma}^2) = P_n w(\theta_0) + o_p(1),
\]

where the inequality follows from the definition of \( \hat{H}(\beta, \sigma^2) \), and two equalities are obtained by the mean value theorem and the consistency of \( \hat{\sigma}^2 \). Therefore, by Theorem 5.7 of van der Vaart (1998), we have \( \|\hat{H}(\beta_0, \sigma^2) - H_0\|_2 \xrightarrow{p} 0 \).

Using Theorem 2.10.6 of van der Vaart and Wellner (1996) and conditions (C1)-(C3), we can show that the class \( \{\ell_1(\beta, H, \sigma^2) | (\beta, H, \sigma^2) \in B \times H \times Q\} \) is Dosker and hence Glivenko-Cantelli. By the consistency of \( \hat{H}(\beta_0, \sigma^2), \hat{\sigma}^2 \) shown above and the fact that \( P_0 \ell_1(\beta_0, H_0, \sigma^2_0) = 0 \), we have \( P_n \ell_1(\beta_0, \hat{H}(\beta_0, \sigma^2), \hat{\sigma}^2) = o_p(1) \). This together with condition (C6) implies the existence of a consistent solution of \( \beta \) to the CS estimating equation \( P_n \ell_1(\beta, \hat{H}(\beta, \sigma^2), \hat{\sigma}^2) = 0 \).

Next, we shall prove

\[
\|\hat{H}(\beta, \sigma^2) - H_0\|_2 = O_P(\|\beta - \beta_0\| + \|\sigma^2 - \sigma^2_0\| + n^{-1/3}),
\]

by verifying the conditions (3.5) and (3.6) in Theorem 3.2 of Murphy and van der Vaart (1999). The rate of convergence of \( \hat{H} \) can then be obtained by the consistency of \( \hat{\beta}, \hat{\sigma}^2 \). A Taylor series argument gives \( P_0 \{\ell(\theta_0) - \ell(\beta, H_0, \sigma^2)\} \geq -\|\beta - \beta_0\|^2 + \|\sigma^2 - \sigma^2_0\|^2 \). This together with condition (C4) can verify

\[
P_0 \{\ell(\theta) - \ell(\beta, H_0, \sigma^2)\} \leq -\|H - H_0\|^2 + \|\beta - \beta_0\|^2 + \|\sigma^2 - \sigma^2_0\|^2,
\]

which is condition (3.5) of Murphy and van der Vaart (1999).

We introduce some definitions from van der Vaart (1998). Given two functions \( l \) and \( u \), the bracket \([l, u]\) is the set of all functions \( f \) with \( l \leq f \leq u \). An \( \varepsilon \)-bracket in \( L_2(P) = \{f : Pf < \infty\} \) is a bracket \([l, u]\) with \( P(u - l)^2 < \varepsilon^2 \). For a subclass \( C \) of \( L^2(P) \), the bracketing number \( N_\alpha(\varepsilon, C, L_2(P)) \) is the minimum number of \( \varepsilon \)-bracket need to cover \( C \).
Let $\Psi = \{\ell(\theta) : \theta \in \mathcal{B} \times \mathcal{H} \times \mathcal{Q}\}$. It is easy to see that each element in $\Psi$ is uniformly bounded and satisfies $P_0(\ell(\theta) - \ell(\beta, H_0, \sigma^2))^2 \leq \|H - H_0\|_2^2 + \|\beta - \beta_0\|^2 + \|\sigma^2 - \sigma_0^2\|^2$. Lemma 1 below gives the bracketing integral $J(\delta, \Psi, L_2(P))$, defined as $\int_0^\delta (1 + \log N(\varepsilon, \Psi, L_2(P)))^{1/2} d\varepsilon$, is $O(\delta^{1/2})$. It then follows from Lemma 3.3 of Murphy and van der Vaart (1999) that their condition (3.6) is satisfied for $\phi_n(\delta) = \delta^{1/2}$. This completes the proof.

**Lemma 1.** $\log N(\varepsilon, \Psi, L_2(P_0)) = O(1/\varepsilon)$.

**Proof.** For fixed $\theta$, the functions in $\Psi$ depend on $H$ monotonically for $\Delta_{K,l} = 1$ and $\Delta_{K,l} = 0$ separately. Thus, given a $\varepsilon$-bracket $H^L \leq H \leq H^U$, it follows from monotonicity of $\mathcal{E}_{K,l}$ in $H$ that we can get a bracket $(\ell^L, \ell^U)$ for $\ell(\theta)$ where

$$\ell^L = \log \prod_{l=1}^K [\mathcal{E}_{K,l}(\beta, H^L, \sigma^2)(O) \Delta_{K,l} \{1 - \mathcal{E}_{K,l}(\beta, H^L, \sigma^2)(O)\}^{1 - \Delta_{K,l}}];$$

$$\ell^U = \log \prod_{l=1}^K [\mathcal{E}_{K,l}(\beta, H^U, \sigma^2)(O) \Delta_{K,l} \{1 - \mathcal{E}_{K,l}(\beta, H^L, \sigma^2)(O)\}^{1 - \Delta_{K,l}}].$$

Further, by the mean value theorem, we have $|\ell^L - \ell^U|^2 \leq \sum_{l=1}^K (H^U - H^L)^2(U_{K,l})$. Thus brackets for $H$ of $\|\cdot\|_2$-size $\varepsilon$ can translate into brackets for $\ell(\theta)$ of $L_2(P_0)$-size proportional to $\varepsilon$. By Example 19.11 of van der Vaart (1998), we can cover the set of all $H$ by $\exp(C/\varepsilon)$ brackets of size $\varepsilon$ for some constant $C$. Next we allow $\zeta = (\beta', \sigma^2)'$ to vary freely as well. Because $\mathcal{B} \times \mathcal{Q}$ is finite-dimensional and $(\partial/\partial \zeta) \ell(\theta)(O)$ is uniformly bounded in $(\theta, O)$, this increases the entropy only slightly. This completes the proof.

Consider a parametric path $H_\varepsilon$ in $\mathcal{H}$ through $H$, that is, $H_\varepsilon \in \mathcal{H}$ and $H_0 = H$ when $\varepsilon = 0$. Let $\mathcal{H} = \{g : (\partial/\partial \varepsilon)|_{\varepsilon = 0} H_\varepsilon = g\}$. Then the score for $H$ along the direction $g$, define by $(\partial/\partial \varepsilon)|_{\varepsilon = 0} \ell(\beta, H_\varepsilon, \sigma^2)$, has the form

$$\ell_{2}(\theta)[g](O) = \sum_{l=1}^K g(U_{K,l})\{\Delta_{k,l} - \mathcal{E}_{K,l}(\theta)(O)\}.$$ 

Also define

$$\ell_{12}(\theta)[g] = (\partial/\partial \varepsilon)|_{\varepsilon = 0} \ell_{1}(\beta, H_\varepsilon, \sigma^2).$$
and \( \ell_{22}(\theta)[\tilde{g}, g] = (\partial/\partial \varepsilon)|_{\varepsilon=0} \ell_2(\beta, \tilde{H}, \sigma^2)[\tilde{g}] \), where \( g \) and \( \tilde{g} \) are in \( \mathcal{H} \). They have forms

\[
\ell_{12}(\theta)[g] = -\sum_{l=1}^{K} g(U_{K,l})\{S_{K,l} - \beta_1 \tilde{\sigma}^2, Z'\}V_{K,l}(\theta)(O),
\]

\[
\ell_{22}(\theta)[g, \tilde{g}] = -\sum_{l=1}^{K} g(U_{K,l})\tilde{g}(U_{K,l})V_{K,l}(\theta)(O).
\]

Following semiparametric M-estimator theories (e.g., Korosok (2008)), the function \( \ell^* \) given in Section 3 is defined as

\[
\ell^*(\theta) = \ell_{12}(\theta) - \ell_{22}(\theta)[g^*],
\]

where \( g^* \) is the \( d \)-dimensional (\( d = \dim(\beta) \)) vector-valued function satisfying

\[
P_0(\ell_{12}(\theta_0) - \ell_{22}(\theta_0)[g^*, g]) = 0,
\]

for all \( g \) in \( \mathcal{H} \). Note that (6.1) can written as

\[
\int \sum_{k=1}^{K} \sum_{l=1}^{k} f_{K,l}(k, u)g(u)E[\{S_{K,l} - \beta_1 \tilde{\sigma}^2, Z'\}V_{K,l}(\theta)(O)|K = k, U_{K,l} = u]du
\]

\[
= \int \sum_{k=1}^{K} \sum_{l=1}^{k} f_{K,l}(k, u)g^*(u)E[V_{K,l}(\theta)(O)|K = k, U_{K,l} = u]du,
\]

which is implies that \( g^* \) is given by (3.1). Below we establish the asymptotic theory of the CS estimator.

**Theorem 2** (Asymptotic normality) The estimator \((\hat{\beta}, \hat{\sigma}^2)\) is asymptotically normal; that is,

\[
\sqrt{n}\begin{bmatrix}
\hat{\beta} - \beta_0 \\
\hat{\sigma}^2 - \sigma_0^2
\end{bmatrix} = \mathcal{I}^{-1}\sqrt{n}\mathbb{P}_n\begin{bmatrix}
\ell^*(\theta_0) \\
\varphi(\sigma_0^2)
\end{bmatrix} + o_P(1) \sim N(0, \mathcal{I}^{-1}\Sigma(\mathcal{I}^{-1})'),
\]

where \( \Sigma = P_0[\ell^*(\theta_0)'\varphi(\sigma_0^2)'] \) and

\[
\mathcal{I} = -E\begin{bmatrix}
\frac{\partial}{\partial \beta} \ell^*(\theta_0) & \frac{\partial}{\partial \sigma^2} \ell^*(\theta_0) \\
0 & \frac{\partial}{\partial \sigma_0^2} \varphi(\sigma_0^2)
\end{bmatrix}.
\]

**Proof.** We first verify

\[
\sqrt{n}P_0\ell^*(\beta_0, \tilde{H}, \sigma_0^2) = o_P(1).
\]
Apply a Taylor expansion to $\ell^*(\beta_0, H, \sigma_0^2)(O)$ at the point $(H_0(U_{K,1}), \ldots, H_0(U_{K,K}))$ to get

$$P_0\ell^*(\beta_0, H, \sigma_0^2) = \left[ P_0\ell^*(\theta_0) + P_0\left\{ \ell_{12}(\theta_0)[H - H_0] - \ell_{22}(\theta_0)[g^*, H - H_0] \right\} \right] + O_p(\|H - H_0\|_{2}^2).$$

(6.3)

Using the fact that $P_0\ell^*(\theta_0) = 0$, (6.1), and applying the rate of convergence on $\hat{H}$ to (6.3), we get (6.2).

It is known that the class of uniformly bounded functions of bounded variations is a Donsker class. Applying condition (C5) and Theorem 2.10.6 of van der Vaart and Wellner (1996), it can be verified that $\{\ell^*(\theta)|\theta \in B \times H \times Q\}$ and $\{\varphi(\sigma^2)|\sigma^2 \in Q\}$ are uniformly bounded Donsker classes; the proof of which is technical and hence omitted here. Combining this with the consistency of $\hat{\theta}$ leads to

$$\sqrt{n}(P_n - P_0) \left[ \frac{\ell^*(\hat{\theta}) - \ell^*(\theta_0)}{\varphi(\sigma_0^2) - \varphi(\hat{\sigma}_0^2)} \right] = o_p(1).$$

By the mean value theorem, there exists $(\hat{\beta}, \hat{\sigma}^2)$ lying between $(\beta_0, \sigma_0^2)$ and $(\hat{\beta}, \hat{\sigma}^2)$ such that

$$-\sqrt{n}P_0 \left[ \frac{\partial}{\partial \beta} \ell^*(\hat{\beta}, \hat{\sigma}^2) \frac{\partial}{\partial \sigma^2} \ell^*(\hat{\beta}, \hat{\sigma}^2) \right] \left( \beta_0 - \beta \right) \left( \sigma_0^2 - \hat{\sigma}_0^2 \right)$$

$$= \sqrt{n}P_n \left[ \ell^*(\theta_0) \varphi(\sigma_0^2) \right] + o_p(1).$$

By the consistency of $(\hat{\beta}, \hat{\sigma}^2)$ and condition (C6), we have

$$\sqrt{n} \left[ \begin{array}{c} \hat{\beta} - \beta_0 \\ \hat{\sigma}^2 - \sigma_0^2 \end{array} \right] = I^{-1} \sqrt{n}P_n \left[ \begin{array}{c} \ell^*(\theta_0) \\ \varphi(\sigma_0^2) \end{array} \right] + o_p(1) \xrightarrow{d} N(0, I^{-1}\Sigma(I^{-1})').$$

This completes the proof.

References


Department of Mathematics, Tamkang University, New Taipei 25137, Taiwan.
E-mail: ccwen@mail.tku.edu.tw

Institute of Statistical Science, Academia Sinica, Taipei 11529, Taiwan.
E-mail: yhchen@stat.sinica.edu.tw