Statistica Sinica **24** (2014), 1301-1317
doi:http://dx.doi.org/10.5705/ss.2013.081

FUNCTIONAL INFERENCE 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 Trail 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 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 computational procedure combining the Newton-Raphson and self-consistency algorithms, is established for the proposed estimation method. We illustrate the performance of the 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 Rossin (1997), Betensky, Rabinowitz, and Tsiatis (2001), and was comprehensively reviewed by Zhang and Sun (2010). Existing methods, however, are restricted to the setting where the covariates are accurately measured, and such a restriction often does not apply. 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)).

While there is considerable work on the covariate measurement error problem for right-censored failure time data (see, e.g., Hu, Tsiatis, and Davidian (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. Both 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.

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 applying 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, Rabinowitz, and Tsiatis (2001) and Zhu, Tong, and Sun (2008). As mentioned in Betensky, Rabinowitz, and Tsiatis (2001), underlying such an approach is the assumption of independence between the examination times and the failure time given covariates. This 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 conditional score method and the properties of the estimator. Section 4 presents the computational algorithm. Section 5 evaluates the proposed method through simulation studies and an analysis of the ACTG 175 data. Section 6 contains concluding remarks. Proofs of the asymptotic properties are given in the Appendix.

2. The Conditional Score Estimator

Let T and $(X, \mathbf{Z}')'$ denote the failure time and covariate vector for a subject, where $\mathbf{Z}$ 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 $\mathbf{W} = (W_1, \ldots, W_m)$ 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 we consider mixed case interval-censoring (Schick and Yu (2000)) in which T is not observed but is monitored by a triangle array of random examination times $\mathbf{U} = \{U_{K,l} : l = 1, \ldots, K, K = 1, 2, \ldots\}$, with $U_{K,1} < \ldots < U_{K,K}$, the number of examinations K being random. Assume that the examinations continue to occur regardless of whether or not failure has occurred, so that the examination times $(K, \mathbf{U})$ and the failure time T are independent given covariates $(X, \mathbf{Z})$. The variables we observe for one subject are thus $O = \{(K, U_{K,l}, \Delta_{K,l}, \mathbf{W}, \mathbf{Z}) : l = 1, \ldots, K\}$, where $\Delta_{K,l} = I(T \le U_{K,l})$ indicates whether the failure time T precedes the examination time $U_{K,l}$.

Given covariates $(X, \mathbf{Z})$, we assume that at time t the conditional survival function of T given covariates is of the form

$$\Pr(T > t | X = x, \mathbf{Z} = \mathbf{z}) = \{1 + \exp(\beta_1 x + \boldsymbol{\beta}_2' \mathbf{z} + H(t))\}^{-1}, \tag{2.1}$$

where $\boldsymbol{\beta} = (\beta_1, \boldsymbol{\beta}_2')'$ is an unknown regression vector and H is an unspecified, nondecreasing and continuous baseline log odds function. Assume that $(T, K, \mathbf{U})$ and $\mathbf{W}$ are conditionally independent given $(X, \mathbf{Z})$, the surrogate condition, and the conditional distribution of $(K, \mathbf{U})$ given $(X, \mathbf{Z})$ does not depend on parameters of interest. Let $\bar{W} = \sum_{j=1}^m W_j / m$ and $\tilde{\sigma}^2 = \sigma^2 / m$. Then the conditional likelihood of $(\Delta_{K,l}, \bar{W})$, given $(K, U_{K,l}, X, \mathbf{Z})$ is proportional to

$$\begin{aligned} &\frac{\exp\{\Delta_{K,l}(\beta_1 X + \boldsymbol{\beta}_2' \mathbf{Z} + H(U_{K,l}))\}}{[1 + \exp\{\beta_1 X + \boldsymbol{\beta}_2' \mathbf{Z} + H(U_{K,l})\}](\tilde{\sigma}^2)^{1/2}} \exp\left\{-\frac{(\bar{W} - X)^2}{2\tilde{\sigma}^2}\right\} \\ = &\frac{\exp\{\Delta_{K,l}(\boldsymbol{\beta}_2' \mathbf{Z} + H(U_{K,l}))\}}{[1 + \exp\{\beta_1 X + \boldsymbol{\beta}_2' \mathbf{Z} + H(U_{K,l})\}](\tilde{\sigma}^2)^{1/2}} \exp\left(-\frac{\bar{W}^2 + X^2}{2\tilde{\sigma}^2}\right) \exp\left\{\frac{X S_{K,l}}{\tilde{\sigma}^2}\right\}, \end{aligned}$$

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

$$\mathcal{E}_{K,l}(\boldsymbol{\theta})(O) = \frac{\exp\{\beta_1 S_{K,l} - \beta_1^2 \tilde{\sigma}^2/2 + \boldsymbol{\beta}_2' \mathbf{Z} + H(U_{K,l})\}}{1 + \exp\{\beta_1 S_{K,l} - \beta_1^2 \tilde{\sigma}^2/2 + \boldsymbol{\beta}_2' \mathbf{Z} + H(U_{K,l})\}},$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}, H, \sigma^2)$.

The approach to estimation here 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 $\{(K, U_{K,l}, S_{K,l}, \mathbf{Z}) : l = 1, \ldots, K\}$ takes the form

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

Let $O_i = \{(K_i, U^{(i)}_{K_i,l}, \Delta^{(i)}_{K_i,l}, \mathbf{W}_i, \mathbf{Z}_i) : l = 1, \ldots, K_i\}$, $i = 1, \ldots, n$, be n i.i.d. copies of observed variable O. Suppose that $\mathbf{W}_i = (W_{i1}, \cdots, W_{im_i})$ and take $\bar{W}_i = \sum_{j=1}^{m_i} W_{ij}/m_i$, $\tilde{\sigma}_i^2 = \sigma^2/m_i$, and $S^{(i)}_{K_i,l} = \beta_1 \Delta^{(i)}_{(K,l)} \tilde{\sigma}_i^2 + \bar{W}_i$ for $l = 1, \ldots, K_i$. Following Stefanski and Carroll (1987), we can construct a conditional score (CS) estimator for $\boldsymbol{\beta}$ by solving the estimating equation

$$\sum_{i=1}^{n} \boldsymbol{\ell}_1(\boldsymbol{\theta})(O_i) = 0, \tag{2.3}$$

where $\boldsymbol{\ell}_1(\boldsymbol{\theta})(O) = \sum_{l=1}^{K} (S_{K,l} - \beta_1 \tilde{\sigma}^2, \mathbf{Z}')' \{\Delta_{K,l} - \mathcal{E}_{K,l}(\boldsymbol{\theta})(O)\}$ is obtained by differentiating the logarithm of conditional likelihood (2.2) with respect to $\boldsymbol{\beta}$, ignoring the dependence of $S_{K,l}$'s on $\boldsymbol{\beta}$.

For fixed $\boldsymbol{\beta}$ and σ^2, we propose to estimate the baseline log odds function H by maximizing the conditional likelihood

$$L_n(\boldsymbol{\theta})(O_1, \ldots, O_n) = \prod_{i=1}^{n} L(\boldsymbol{\theta})(O_i). \tag{2.4}$$

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

Usually the error variance σ^2 is also unknown and must be estimated. Based on the replicated measurement data $\{\mathbf{W}_i : i = 1, \ldots, n\}$, σ^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

$$\sum_{i=1}^{n} \varphi(\sigma^2)(\mathbf{W}_i) = 0, \tag{2.5}$$

where $\varphi(\sigma^2)(\mathbf{W}_i) = \sum_{j=1}^{m_i} (W_{ij} - \bar{W}_i)^2 - (m_i - 1)\sigma^2$.

The proposed estimation procedure thus consists of the following steps.

1. Obtain the estimate $\hat{\sigma}^2$ for the variance σ^2 of measurement error, using (2.5).
2. Substitute $\hat{\sigma}^2$ for σ^2 in (2.3) and (2.4), obtain the conditional score estimate $(\widehat{\boldsymbol{\beta}}, \widehat{H})$ for $(\boldsymbol{\beta}, H)$ by solving (2.3) and maximizing (2.4).

We detail in Section 4 the computational algorithm used in Step 2. In Step 1, to obtain a consistent estimate for the error variance σ^2, we require replicates of W_i, i.e., $m_i \geq 2$. In fact, this is also required in Step 2 to obtain the consistent conditional score estimate $(\widehat{\boldsymbol{\beta}}, \widehat{H})$, owing to the fact that the estimating equations derived from (2.3) and (2.4) are not orthogonal to that for the parameter σ^2.

Remark 1. As suggested by the Associate Editor, the examination times $U_{K,1} < \ldots < U_{K,K}$ can be alternatively viewed as the recurrence times generated by a point process. Treating the point process as a whole as a random component is a more direct way to examine the relationship between failure times and examination times. This is a promising topic for our future research.

3. Asymptotic Theories and Variance Estimation

The asymptotic theories for the CS estimator $(\widehat{\boldsymbol{\beta}}, \widehat{H})$ can be established using empirical process and semiparametric M-estimator theories (Korosok (2008)). Details of the theories, together with the assumptions and proofs, are relegated to the Appendix. Briefly, the CS estimator $(\widehat{\boldsymbol{\beta}}, \widehat{H})$ is consistent for $(\boldsymbol{\beta}_0, H_0)$, the true value of $(\boldsymbol{\beta}, H)$. The convergence rate of $\widehat{\boldsymbol{\beta}}$ is of order $n^{1/2}$, but that of $\widehat{H}$ is of order $n^{1/3}$ only; these are rates obtained in general semiparametric analysis of interval-censored data; see, e.g., Huang (1996). Assuming temporarily that the true value σ_0^2 of error variance σ^2 is known, we have

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \doteq \left[-E\left\{ \frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{\ell}^*(\boldsymbol{\theta}_0) \right\} \right]^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \boldsymbol{\ell}_i^*(\boldsymbol{\theta}_0),$$

where $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0, H_0, \sigma_0^2)$, $\boldsymbol{\ell}^*(\boldsymbol{\theta})(O) = \sum_{l=1}^{K} \{(S_{K,l} - \beta_1 \tilde{\sigma}^2, \mathbf{Z}')' - \mathbf{g}^*(U_{K,l})\}\{\Delta_{K,l} - \mathcal{E}_{K,l}(\boldsymbol{\theta})(O)\}$, $\boldsymbol{\ell}_i^*(\boldsymbol{\theta}) = \boldsymbol{\ell}^*(\boldsymbol{\theta})(O_i)$, and $\mathbf{g}^*$ is given by

$$\mathbf{g}^*(u) = \frac{\sum_{k=1}^{\infty} \sum_{l=1}^{k} f_{K,l}(k,u) E_{k,u} \left[\{S_{K,l} - \beta_1 \tilde{\sigma}^2, \mathbf{Z}'\}' \mathcal{V}_{k,l}(\boldsymbol{\theta})(O) \right]}{\sum_{k=1}^{\infty} \sum_{l=1}^{k} f_{K,l}(k,u) E_{k,u} \left[\mathcal{V}_{k,l}(\boldsymbol{\theta})(O) \right]}, \tag{3.1}$$

evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ with $\mathcal{V}_{K,l}(\boldsymbol{\theta}) = \mathcal{E}_{K,l}(\boldsymbol{\theta})(1 - \mathcal{E}_{K,l}(\boldsymbol{\theta}))$, $f_{K,l}$ the density of $(K, U_{K,l})$ and $E_{k,u}(\cdot)$, the conditional expectation given $K = k, U_{K,l} = u$.

In general σ^2 is unknown and can be estimated through (2.5). Write $\varphi_i(\sigma^2) = \varphi(\sigma^2)(\mathbf{W}_i)$ in (2.5). To account for the extra estimation of σ^2, we thus have

$$\sqrt{n} \begin{bmatrix} \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \\ \widehat{\sigma}^2 - \sigma_0^2 \end{bmatrix} \doteq \mathcal{I}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \begin{bmatrix} \boldsymbol{\ell}_i^*(\boldsymbol{\theta}_0) \\ \varphi_i(\sigma_0^2) \end{bmatrix},$$

which is asymptotically distributed as $N(0, \mathcal{I}^{-1}\Sigma(\mathcal{I}^{-1})')$, where Σ is the covariance matrix of $(\boldsymbol{\ell}^*(\boldsymbol{\theta}_0)', \varphi(\sigma_0^2))'$, and

$$\mathcal{I} = -E\begin{bmatrix} \frac{\partial}{\partial\boldsymbol{\beta}}\boldsymbol{\ell}^*(\boldsymbol{\theta}_0) & \frac{\partial}{\partial\sigma^2}\boldsymbol{\ell}^*(\boldsymbol{\theta}_0) \\ 0 & \frac{\partial}{\partial\sigma^2}\varphi(\sigma_0^2) \end{bmatrix}. \tag{3.2}$$

The asymptotic variance can be estimated by replacing the population quantities with their empirical counterparts. In particular, according to Huang (1996), the quantity $\mathbf{g}^*$ 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 $\boldsymbol{\beta}$ can be simply performed. For example, the test of $H_0 : \boldsymbol{\beta} = 0$ can be performed via the Wald test statistic $\widehat{\boldsymbol{\beta}}'\{\widehat{\text{var}}(\widehat{\boldsymbol{\beta}})\}^{-1}\widehat{\boldsymbol{\beta}}$, which has an asymptotic χ^2 distribution with $d = \dim(\boldsymbol{\beta})$ degrees of freedom under H_0, where $\widehat{\text{var}}(\widehat{\boldsymbol{\beta}})$ is the submatrix of $\widehat{\mathcal{I}}^{-1}\widehat{\Sigma}(\widehat{\mathcal{I}}^{-1})'$ corresponding to $\widehat{\boldsymbol{\beta}}$.

4. Computational 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, that 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 $\mathbf{h} = (h_1, \ldots, h_N)'$, where $h_1 = H(u_1)$ and $h_j = H(u_j) - H(u_j-)$ is the jump size of H at u_j for $j \geq 2$. We treat the h_j's as non-zero parameters even though some of them may approach 0, and obtain the estimate $\widehat{h}_j$ for h_j by differentiating the conditional likelihood (2.4) with respect to h_j, $j = 1, \ldots, N$. Accordingly, for fixed $(\boldsymbol{\beta}, \sigma^2)$, $\widehat{\mathbf{h}} = (\widehat{h}_1, \ldots, \widehat{h}_N)'$ is obtained as the solution to the system of equations

$$\begin{aligned} \frac{\partial}{\partial h_j}\log L_n(\boldsymbol{\beta}, \widehat{\mathbf{h}}, \sigma^2) &= \sum_{i=1}^{n}\sum_{l=1}^{K_i}\Delta_{K_i,l}^{(i)} I[U_{K_i,l}^{(i)} \geq u_j] \\ &\quad - \sum_{i=1}^{n}\sum_{l=1}^{K_i}\mathcal{E}_{K,l}(\boldsymbol{\beta}, \widehat{\mathbf{h}}, \sigma^2)(O_i) I[U_{K_i,l}^{(i)} \geq u_j] \\ &\equiv a_j - b_j(\boldsymbol{\beta}, \widehat{\mathbf{h}}, \sigma^2) \\ &= 0. \quad (j = 1, \ldots, N) \end{aligned} \tag{4.1}$$

In (4.1), by definition, $a_j = b_j(\boldsymbol{\beta}, \widehat{\mathbf{h}}, \sigma^2)$ and $a_j + M_0 = b_j(\boldsymbol{\beta}, \widehat{\mathbf{h}}, \sigma^2) + M_0$ for any constant M_0. For fixed $\boldsymbol{\beta}$ and σ^2, we consider the self-consistency algorithm for solving $\widehat{\mathbf{h}}$:

$$h_j^{(k+1)} = D_j(\boldsymbol{\beta}, \mathbf{h}^{(k)}, \sigma^2), \quad j = 1, \ldots, N, \tag{4.2}$$

with

$$D_j(\boldsymbol{\beta}, \mathbf{h}^{(k)}, \sigma^2) = h_j^{(k)} \cdot \left\{ \frac{a_j + M_0}{b_j(\boldsymbol{\beta}, \mathbf{h}^{(k)}, \sigma^2) + M_0} \right\}, \text{ for } j = 2, \ldots, N, \tag{4.3}$$

$$D_1(\boldsymbol{\beta}, \mathbf{h}^{(k)}, \sigma^2) = h_1^{(k)} + \log \left\{ \frac{a_1 + M_0}{b_1(\boldsymbol{\beta}, \mathbf{h}^{(k)}, \sigma^2) + M_0} \right\}, \tag{4.4}$$

where the superscript (k) denotes the kth iteration of the algorithm. It is easy to check that if $\mathbf{h}^{(k)} = \widehat{\mathbf{h}}$ in (4.2), the estimate $\widehat{\mathbf{h}}$ 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, σ^2), the CS estimate for $\boldsymbol{\beta}$ can be obtained by traditional methods such as the Newton-Raphson algorithm. The estimate $\widehat{\sigma}^2$ for σ^2 can be obtained separately by solving (2.5). We therefore compute the CS estimate $(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{h}})$ via a hybrid algorithm, consisting of a Newton-Raphson algorithm for solving $\boldsymbol{\beta}$ and a self-consistency algorithm for solving $\mathbf{h}$. The hybrid algorithm proposed is as follows. Starting from initial values $\boldsymbol{\beta}^{(0)}$ and $\mathbf{h}^{(0)}$ and fixing $\sigma^2 = \widehat{\sigma}^2$ throughout, for $k = 0, 1, \ldots$, we iterate between Steps 1 and 2 below until some convergence criterion is met.

1. Fix $\mathbf{h} = \mathbf{h}^{(k)}$, update $\boldsymbol{\beta}^{(k)}$ to $\boldsymbol{\beta}^{(k+1)}$ by solving (2.3) with the one-step Newton-Raphson algorithm.
2. Fix $\boldsymbol{\beta} = \boldsymbol{\beta}^{(k+1)}$, update $\mathbf{h}^{(k)}$ to $\mathbf{h}^{(k+1)}$ by (4.2).

Remark 2. As with other conditional score methods, there may be multiple solutions. To better locate the consistent solution, as in Tsiatis and Davidian (2001), we solve (2.3) with the initial $(\boldsymbol{\beta}, H)$ 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 has worked 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 $e_l(\boldsymbol{\beta}, \mathbf{h}, \sigma^2) \equiv (\partial/\partial h_l) b_l(\boldsymbol{\beta}, \mathbf{h}, \sigma^2)$ are all positive, and $a_l = b_l(\boldsymbol{\beta}, \widehat{\mathbf{h}}, \sigma^2)$, 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 $\mathbf{h}$ is near $\widehat{\mathbf{h}}$. This means that the self-consistency algorithm is locally contractive and converges by the contraction principle (Rudin (1973)).

Remark 4. By definition, h_1 may be negative while h_j is positive for $j \geq 2$. To accommodate this, the type of iterative equation used for h_j, $j \geq 2$, 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 performed 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) was Bernoulli(0.5) and the error-prone covariate X was $N(0,1)$. The true regression coefficient (β_{10}, β_{20}) was $(0.5, -0.5)$ or $(1, -0.5)$, and the baseline log odds $H_0(t)$ was $\log(e^t - 1)$. Two surrogate measurements $\mathbf{W} = (W_1, W_2)$ of X were made per subject with error variance $\sigma^2 = 0.25, 0.5$, or 0.75. The number of examinations K was randomly selected from $\{3, 4, 5\}$ and, given K, the examination time points $U_{K,1} < \ldots < U_{K,K}$ were generated as the ordered statistics of a random sample of size K from Uniform$(0, 1)$. The sample size was $n = 150$ or 300, and the simulation replication was 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) were calculated over simulation replicates and are summarized in Table 1. Since the simulation replication was 400, if the true coverage were 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, that 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. Thus the naive estimator maximizes, over the parameter $(\boldsymbol{\beta}, H)$, the likelihood

$$\tilde{L}(\boldsymbol{\beta}, H) = \prod_{i=1}^{n} \prod_{l=1}^{K_i+1} \left\{ F_i(U_{K_i,l-1}^{(i)}) - F_i(U_{K_i,l}^{(i)}) \right\}^{\tilde{\Delta}_{K_i,l}^{(i)}} \tag{5.1}$$

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

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 the simulation standard deviation.

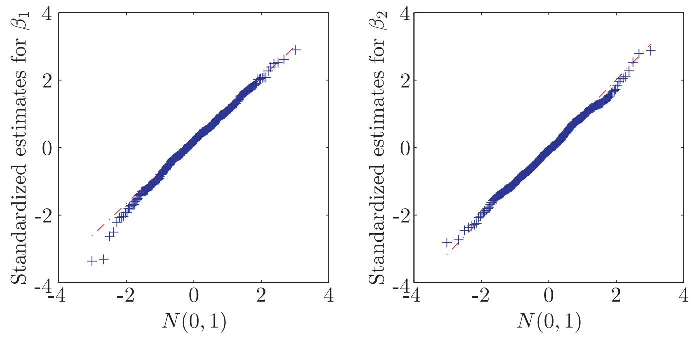

Figure 1. The Q-Q plot of standardized CS estimates versus standard normal variate for the simulation scenario with $(\beta_1, \beta_2) = (1, -0.5), \sigma^2 = 0.5$ and $n = 300$.

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 $(\widehat{\text{var}}(\widehat{\beta}))^{-1/2}(\widehat{\beta} - \beta_0)$ with the standard normal variate. It agrees well. In contrast, from Table 1 we see that the naive estimator, obtained by using the mean surrogate values in the standard proportional odds analysis, performed poorly for β_1, the coefficient corresponding to the error-prone covariate.

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

5.2. Application to ACTG 175 data

We applied 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 considered a proportional odds model with covariates log(CD4) (X) and a treatment indicator ($Z = 1$ for zidovudine alone and 0 for any of the other three therapies). The surrogates (W_1, W_2) for the error-prone covariate X were taken to be the last two measurements of the CD4 count observed prior to treatment, 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 measurement error nearly normal. The mean number of examination times K per subject was 9.06 with the range 1–15 among the total 1014 patients. The proposed condi-

Table 1. Simulation Results
(a) Conditional score method

			β_1 estimate				β_2 estimate			
(β_{10}, β_{20})	σ_{e0}^2	n	Bias	SD	ASE	CP	Bias	SD	ASE	CP
(0.5,-0.5)	0.25	150	0.025	0.222	0.195	93.00	-0.028	0.353	0.348	95.00
		300	0.017	0.135	0.136	95.25	-0.014	0.232	0.242	95.75
	0.50	150	0.028	0.240	0.210	93.75	-0.029	0.355	0.350	94.75
		300	0.018	0.144	0.146	95.00	-0.014	0.233	0.243	95.75
	0.75	150	0.033	0.263	0.228	95.00	-0.030	0.358	0.353	95.25
		300	0.019	0.154	0.156	95.00	-0.014	0.235	0.245	95.75
(1,-0.5)	0.25	150	0.073	0.258	0.238	94.50	-0.038	0.376	0.369	94.75
		300	0.040	0.168	0.165	93.75	-0.021	0.248	0.256	95.25
	0.50	150	0.082	0.277	0.267	95.00	-0.040	0.385	0.377	94.25
		300	0.045	0.186	0.183	95.00	-0.022	0.252	0.261	95.75
	0.75	150	0.098	0.312	0.303	95.50	-0.045	0.396	0.388	94.75
		300	0.052	0.207	0.204	96.25	-0.023	0.257	0.266	95.75

(b) Naive method

			β_1 estimate				β_2 estimate			
(β_{10}, β_{20})	σ_{e0}^2	n	Bias	SD	ASE	CP	Bias	SD	ASE	CP
(0.5,-0.5)	0.25	150	-0.054	0.175	0.165	93.25	-0.016	0.316	0.327	94.75
		300	-0.054	0.113	0.114	93.00	-0.008	0.216	0.228	95.25
	0.50	150	-0.103	0.163	0.155	88.25	-0.011	0.314	0.326	95.25
		300	-0.102	0.106	0.107	84.25	-0.004	0.215	0.228	95.25
	0.75	150	-0.142	0.153	0.147	80.25	-0.007	0.313	0.326	95.75
		300	-0.141	0.100	0.102	72.00	-0.000	0.214	0.228	95.50
(1,-0.5)	0.25	150	-0.102	0.182	0.190	91.00	-0.011	0.334	0.339	96.75
		300	-0.119	0.128	0.130	82.75	-0.003	0.226	0.236	96.50
	0.50	150	-0.208	0.168	0.176	77.25	0.003	0.331	0.337	96.50
		300	-0.222	0.120	0.121	54.25	0.009	0.225	0.235	96.25
	0.75	150	-0.292	0.157	0.165	54.50	0.013	0.329	0.335	96.50
		300	-0.304	0.113	0.113	22.75	0.018	0.224	0.234	96.75

tional score method was applied to account for both the measurement error and interval-censoring present.

The results of applying both the CS and the naive methods are given in Table 2. 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 naive method using the mean surrogate CD4 measurements yields a remarkably attenuated estimates for the effects of true baseline CD4 count and treatment. The estimate of the error variance σ^2 given by the CS method is

Table 2. Analysis of ACTG data

	Effects			
	log(CD4)		treatment (zidovudine)	
Method	Estimate	S.E.	Estimate	S.E.
CS	-1.0669	0.1907	0.8032	0.2733
Naive	-0.7776	0.1452	0.7585	0.2789

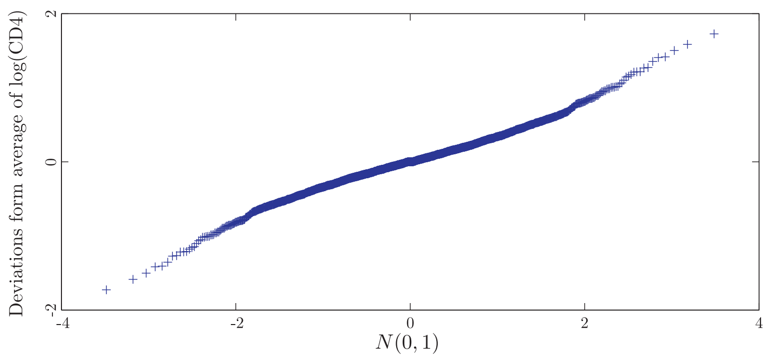

Figure 2. Q-Q plots of the deviations from the mean of standardized log(CD4).

0.2928 with a standard error of 0.0174 (on the standardized log(CD4) scale).

Remark 5. We used the last two measurements prior to treatment as two replicates of the surrogate baseline CD4 count. This was done because for the study subjects, the last two CD4 count measurements prior to treatment were measured consecutively within one month, hence 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 AIDS Clinical Trial Group 175 data, where the failure time (AIDS incidence time) was subject to interval-censoring and the covariate (baseline CD4 count) was 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 utilized the strengths of the conditional score approach of Wen and Chen (2012) and a working independence idea similar to that in Betensky, Rabinowitz, and Tsiatis (2001). Results from simulation studies and an analysis of ACTG 175 data reveals the utilitiy the proposed method.

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, Cai, and Shen (2006)). To improve estimation efficiency, it is also worthwhile to develop a functional modeling approach without the working independence assumption for the examination times on the same subject.

Acknowledgement

The authors are grateful for helpful comments from an associate editor and the 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, $\mathbb{P}_n f = \sum_{i=1}^{n} f(O_i)/n$, 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, proofs are presented under the simpler setting where the distribution of $(K, U_{K,l}, l = 1,\ldots,K)$ is independent of $(\mathbf{W}, Z)$, though the proposed method can accommodate the dependence case. Let $\ell(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta})$. The asymptotic theories are based on the regularity assumptions that have been similarly made in the context of interval-censoring studies (e.g., Huang and Wellner (1997); Zeng, Cai, and Shen (2006); Ma (2010)).

(C1) There exists a positive ξ 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 $(\boldsymbol{\beta}_0, \sigma_0^2)$ lies in the interior of a compact parameter set $\mathcal{B} \times \mathcal{Q}$; H_0 is continuously differentiable on $[\tau_1,\tau_2]$ and satisfies $-M < H_0(\tau_1) < H_0(\tau_2) < M$.

(C4) For $\boldsymbol{\theta}$ near $\boldsymbol{\theta}_0$, $P_0\{\ell(\boldsymbol{\theta}) - \ell(\boldsymbol{\theta}_0)\} \preceq -\{\|H - H_0\|_2^2 + \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|^2 + \|\sigma^2 - \sigma_0^2\|^2\}$, 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 $\preceq$ means smaller than, up to a constant.

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

(C6) The information matrix $\mathcal{I}$ defined as (3.2) is invertible.

The condition (C1) rules out accurately observed failure time and makes the number of examination times K bounded.

Theorem A.0 (Consistency and rate of convergence)**.** *The estimator $\widehat{\boldsymbol{\beta}}$ is consistent, $\widehat{\boldsymbol{\beta}} \xrightarrow{P} \boldsymbol{\beta}_0$. The rate of convergence of $\widehat{H}$ is of order $n^{-1/3}$, $\|\widehat{H} - H_0\|_2 = O_p(n^{-1/3})$.*

Proof. Let $\widehat{H}_{(\boldsymbol{\beta},\sigma^2)}$ be the maximizer of L_n with $(\boldsymbol{\beta}, \sigma^2)$ fixed. Let $w(\boldsymbol{\theta}) = \log\{[L(\boldsymbol{\theta}) + L(\boldsymbol{\theta}_0)]/2\}$. 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), the class $\{w(\boldsymbol{\beta}_0, H, \sigma_0^2) \mid H \in \mathcal{H}\}$ is Donsker and hence Glivenko-Cantelli. Further, by the concavity of $r(u) \equiv \log((u+1)/2)$ and Jensen's inequality, we have

$$P_0[w(\boldsymbol{\beta}_0, H, \sigma_0^2) - w(\boldsymbol{\theta}_0)] = P_0 r\left(\frac{L(\boldsymbol{\beta}_0, H, \sigma_0^2)}{L(\boldsymbol{\theta}_0)}\right) \leq r(P_0\left(\frac{L(\boldsymbol{\beta}_0, H, \sigma_0^2)}{L(\boldsymbol{\theta}_0)}\right)) = 0,$$

and the equality holds only if $H = H_0$ on (τ_1, τ_2). This indicates that

$$\sup_{\|H-H_0\|_2 > \varepsilon} P_0 w(\boldsymbol{\beta}_0, H, \sigma_0^2) < P_0 w(\boldsymbol{\theta}_0).$$

Furthermore,

$$\begin{aligned}\mathbb{P}_n w(\boldsymbol{\beta}_0, \widehat{H}_{(\boldsymbol{\beta}_0,\widehat{\sigma}^2)}, \sigma_0^2) + o_p(1) &= \mathbb{P}_n w(\boldsymbol{\beta}_0, \widehat{H}_{(\boldsymbol{\beta}_0,\widehat{\sigma}^2)}, \widehat{\sigma}^2) \\ &\geq \mathbb{P}_n w(\boldsymbol{\beta}_0, H_0, \widehat{\sigma}^2) \\ &= \mathbb{P}_n w(\boldsymbol{\theta}_0) + o_p(1),\end{aligned}$$

where the inequality follows from the definition of $\widehat{H}_{(\boldsymbol{\beta},\sigma^2)}$, and the equalities are obtained by the Mean Value Theorem and the consistency of $\widehat{\sigma}^2$. By Theorem 5.7 of van der Vaart (1998), we have $\|\widehat{H}_{(\boldsymbol{\beta}_0,\widehat{\sigma}^2)} - H_0\|_2 \xrightarrow{P} 0$.

Using Theorem 2.10.6 of van der Vaart and Wellner (1996) and conditions (C1)−(C3), we can show that the class $\{\boldsymbol{\ell}_1(\boldsymbol{\beta}, H, \sigma^2) | (\boldsymbol{\beta}, H, \sigma^2) \in \mathcal{B} \times \mathcal{H} \times \mathcal{Q}\}$ is Dosker and hence Glivenko-Cantelli. By the consistency of $(\widehat{H}_{(\boldsymbol{\beta}_0,\widehat{\sigma}^2)}, \widehat{\sigma}^2)$ and the fact that $P_0\boldsymbol{\ell}_1(\boldsymbol{\beta}_0, H_0, \sigma_0^2) = 0$, we have $\mathbb{P}_n\boldsymbol{\ell}_1(\boldsymbol{\beta}_0, \widehat{H}_{(\boldsymbol{\beta}_0,\widehat{\sigma}^2)}, \widehat{\sigma}^2) = o_p(1)$. This together with (C6) implies the existence of a consistent solution of $\boldsymbol{\beta}$ to the CS estimating equation $\mathbb{P}_n\boldsymbol{\ell}_1(\boldsymbol{\beta}, \widehat{H}_{(\boldsymbol{\beta},\widehat{\sigma}^2)}, \widehat{\sigma}^2) = 0$.

We prove

$$\|\widehat{H}_{(\boldsymbol{\beta},\sigma^2)} - H_0\|_2 = O_P(\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| + \|\sigma^2 - \sigma_0^2\| + n^{-1/3})$$

by verifying conditions (3.5) and (3.6) in Theorem 3.2 of Murphy and van der Vaart (1999). The rate of convergence of $\widehat{H}$ can then be obtained by the consistency of $(\widehat{\boldsymbol{\beta}}, \widehat{\sigma}^2)$. A Taylor series argument gives $P_0\{\ell(\boldsymbol{\theta}_0) - \ell(\boldsymbol{\beta}, H_0, \sigma^2))\} \preceq \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|^2 + \|\sigma^2 - \sigma_0^2\|^2$. This together with (C4) can verify that

$$P_0\{\ell(\boldsymbol{\theta}) - \ell(\boldsymbol{\beta}, H_0, \sigma^2)\} \preceq -\|H - H_0\|_2^2 + \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|^2 + \|\sigma^2 - \sigma_0^2\|^2,$$

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

Given two functions l and u, the bracket $[l, u]$ is the set of all functions f with $l \leq f \leq u$. An ε-bracket in $L_2(P) = \{f : Pf^2 < \infty\}$ is a bracket $[l, u]$ with $P(u-l)^2 < \varepsilon^2$. For a subclass $\mathcal{C}$ of $L^2(P)$, the bracketing number $N_{[\ ]}(\varepsilon, \mathcal{C}, L_2(P))$ is the minimum number of ε-bracket need to cover $\mathcal{C}$.

Let $\Psi = \{\ell(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \mathcal{B} \times \mathcal{H} \times \mathcal{Q}\}$. It is easy to see that each element in Ψ is uniformly bounded and satisfies $P_0\{\ell(\boldsymbol{\theta}) - \ell(\boldsymbol{\beta}, H_0, \sigma^2)\}^2 \preceq \|H - H_0\|_2^2 + \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|^2 + \|\sigma^2 - \sigma_0^2\|^2$. Lemma 1 gives the bracketing integral $J(\delta, \Psi, L_2(P)) = \int_0^\delta \{1 + \log N_{[\ ]}(\varepsilon, \Psi, L_2(P))\}^{1/2} d\varepsilon$ as $O(\delta^{1/2})$. It then follows from Lemma 3.3 of Murphy and van der Vaart (1999) that their (3.6) is satisfied for $\phi_n(\delta) = \delta^{1/2}$. This completes the proof.

Lemma A.1. $\log N_{[\ ]}(\varepsilon, \Psi, L_2(P_0)) = O(1/\varepsilon)$.

Proof. For fixed $\boldsymbol{\theta}$, the functions in Ψ depend on H monotonically for $\Delta_{K,l} = 1$ and $\Delta_{K,l} = 0$ separately. Thus, given a ε-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 (ℓ^L, ℓ^U) for $\ell(\boldsymbol{\theta})$ where

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

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

Suppose $H_\varepsilon \in \mathcal{H}$ and $H_\varepsilon = H$ when $\varepsilon = 0$. Let $\dot{\mathcal{H}} = \{g : (\partial/\partial\varepsilon)|_{\varepsilon=0} H_\varepsilon = g\}$. Then the score for H along the direction g, $(\partial/\partial\varepsilon)|_{\varepsilon=0}\ell(\boldsymbol{\beta}, H_\varepsilon, \sigma^2)$, has the form

$$\ell_2(\boldsymbol{\theta})[g](O) = \sum_{l=1}^K g(U_{K,l})\{\Delta_{k,l} - \mathcal{E}_{K,l}(\boldsymbol{\theta})(O)\}.$$

Take $\ell_{12}(\boldsymbol{\theta})[g] = (\partial/\partial\varepsilon)|_{\varepsilon=0}\ell_1(\boldsymbol{\beta}, H_\varepsilon, \sigma^2)$ and $\ell_{22}(\boldsymbol{\theta})[\tilde{g}, g] = (\partial/\partial\varepsilon)|_{\varepsilon=0}\ell_2(\boldsymbol{\beta}, H_\varepsilon, \sigma^2)[\tilde{g}]$, where g and $\tilde{g}$ are in $\dot{\mathcal{H}}$. Then

$$\ell_{12}(\boldsymbol{\theta})[g] = -\sum_{l=1}^K g(U_{K,l})\{S_{K,l} - \boldsymbol{\beta}_1\tilde{\sigma}^2, Z'\}'\mathcal{V}_{K,l}(\boldsymbol{\theta})(O),$$

$$\ell_{22}(\boldsymbol{\theta})[g,\tilde{g}] = -\sum_{l=1}^{K} g(U_{K,l})\tilde{g}(U_{K,l})\mathcal{V}_{K,l}(\boldsymbol{\theta})(O).$$

Following semiparametric M-estimator theories (e.g., Korosok (2008)), the function $\boldsymbol{\ell}^*$ given in Section 3 is $\boldsymbol{\ell}^*(\boldsymbol{\theta}) = \ell_1(\boldsymbol{\theta}) - \ell_2(\boldsymbol{\theta})[\mathbf{g}^*]$, where $\mathbf{g}^*$ is the d-dimensional ($d = \dim(\boldsymbol{\beta})$) vector-valued function satisfying

$$P_0(\ell_{12}(\boldsymbol{\theta}_0) - \ell_{22}(\boldsymbol{\theta}_0)[\mathbf{g}^*, g]) = 0, \tag{A.1}$$

for all g in $\dot{\mathcal{H}}$. Note that (A.1) can written as

$$\int \sum_{k=1}^{\infty}\sum_{l=1}^{k} f_{K,l}(k,u)g(u)E[\{S_{K,l} - \beta_1\tilde{\sigma}^2, Z'\}'\mathcal{V}_{K,l}(\boldsymbol{\theta})(O)|K=k, U_{K,l}=u]du$$
$$= \int \sum_{k=1}^{\infty}\sum_{l=1}^{k} f_{K,l}(k,u)g(u)\mathbf{g}^*(u)E[\mathcal{V}_{K,l}(\boldsymbol{\theta})(O)|K=k, U_{K,l}=u]du,$$

which implies that $\mathbf{g}^*$ is given by (3.1).

Theorem A.1 (Asymptotic normality)**.** The estimator $(\widehat{\boldsymbol{\beta}}, \widehat{\sigma}^2)$ satisfies

$$\sqrt{n}\begin{bmatrix}\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \\ \widehat{\sigma}^2 - \sigma_0^2\end{bmatrix} = \mathcal{I}^{-1}\sqrt{n}\mathbb{P}_n\begin{bmatrix}\boldsymbol{\ell}^*(\boldsymbol{\theta}_0) \\ \varphi(\sigma_0^2)\end{bmatrix} + o_P(1) \xrightarrow{d} N(0, \mathcal{I}^{-1}\Sigma(\mathcal{I}^{-1})'),$$

where $\Sigma = P_0\{[\boldsymbol{\ell}^*(\boldsymbol{\theta}_0)', \varphi(\sigma_0^2)]'[\boldsymbol{\ell}^*(\boldsymbol{\theta}_0)', \varphi(\sigma_0^2)]\}$ and

$$\mathcal{I} = -E\begin{bmatrix}\frac{\partial}{\partial\boldsymbol{\beta}}\boldsymbol{\ell}^*(\boldsymbol{\theta}_0) & \frac{\partial}{\partial\sigma^2}\boldsymbol{\ell}^*(\boldsymbol{\theta}_0) \\ 0 & \frac{\partial}{\partial\sigma^2}\varphi(\sigma_0^2)\end{bmatrix}.$$

Proof. We first verify

$$\sqrt{n}P_0\boldsymbol{\ell}^*(\boldsymbol{\beta}_0, \widehat{H}, \sigma_0^2) = o_p(1). \tag{A.2}$$

Apply a Taylor expansion to $\boldsymbol{\ell}^*(\boldsymbol{\beta}_0, H, \sigma_0^2)(O)$ at the point $(H_0(U_{K,1}), \ldots, H_0(U_{K,K}))$ to get

$$P_0\boldsymbol{\ell}^*(\boldsymbol{\beta}_0, H, \sigma_0^2) = P_0\boldsymbol{\ell}^*(\boldsymbol{\theta}_0) + P_0\{\ell_{12}(\boldsymbol{\theta}_0)[H - H_0] - \ell_{22}(\boldsymbol{\theta}_0)[\mathbf{g}^*, H - H_0]\}$$
$$+ O_p(\|H - H_0\|_2^2). \tag{A.3}$$

Using the fact that $P_0\boldsymbol{\ell}^*(\boldsymbol{\theta}_0) = 0$, (A.1), and applying the rate of convergence on $\widehat{H}$ to $(A.3)$, we get $(A.2)$.

It is known that the class of uniformly bounded functions of bounded variations is a Donsker class. Applying (C5) and Theorem 2.10.6 of van der Vaart and Wellner (1996), it can be verified that $\{\boldsymbol{\ell}^*(\boldsymbol{\theta})|\boldsymbol{\theta} \in \mathcal{B}\times\mathcal{H}\times\mathcal{Q}\}$ and $\{\varphi(\sigma^2)|\sigma^2 \in \mathcal{Q}\}$

are uniformly bounded Donsker classes; the proof is technical and is omitted here. Combining this with the consistency of $\widehat{\boldsymbol{\theta}}$ leads to

$$\sqrt{n}(\mathbb{P}_n - P_0)\begin{bmatrix}\boldsymbol{\ell}^*(\widehat{\boldsymbol{\theta}}) - \boldsymbol{\ell}^*(\boldsymbol{\theta}_0)\\ \varphi(\widehat{\sigma}^2) - \varphi(\sigma_0^2)\end{bmatrix} = o_p(1).$$

Adding (A.2) to the first row of here and using the facts that $P_0\boldsymbol{\ell}^*(\boldsymbol{\theta}_0) = 0$ and $\mathbb{P}_n\boldsymbol{\ell}^*(\widehat{\boldsymbol{\theta}}) = \mathbb{P}_n\varphi(\widehat{\sigma}^2) = 0$,

$$-\sqrt{n}P_0\begin{bmatrix}\boldsymbol{\ell}^*(\widehat{\boldsymbol{\theta}}) - \boldsymbol{\ell}^*(\boldsymbol{\beta}_0, \widehat{H}, \sigma_0^2)\\ \varphi(\widehat{\sigma^2}) - \varphi(\sigma_0^2)\end{bmatrix} = \sqrt{n}\mathbb{P}_n\begin{bmatrix}\boldsymbol{\ell}^*(\boldsymbol{\theta}_0)\\ \varphi(\sigma_0^2)\end{bmatrix} + o_p(1).$$

By the Mean Value Theorem, there exists $(\tilde{\boldsymbol{\beta}}, \tilde{\sigma^2})$ lying between $(\widehat{\boldsymbol{\beta}}, \widehat{\sigma^2})$ and $(\boldsymbol{\beta}_0, \sigma_0^2)$ such that

$$-\sqrt{n}P_0\begin{bmatrix}\frac{\partial}{\partial\boldsymbol{\beta}}\boldsymbol{\ell}^*(\tilde{\boldsymbol{\beta}}, \widehat{H}, \tilde{\sigma^2}) & \frac{\partial}{\partial\sigma^2}\boldsymbol{\ell}^*(\tilde{\boldsymbol{\beta}}, \widehat{H}, \tilde{\sigma^2})\\ 0 & \frac{\partial}{\partial\sigma^2}\varphi(\tilde{\sigma}^2)\end{bmatrix}\begin{pmatrix}\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\\ \widehat{\sigma}^2 - \sigma_0^2\end{pmatrix} = \sqrt{n}\mathbb{P}_n\begin{bmatrix}\boldsymbol{\ell}^*(\boldsymbol{\theta}_0)\\ \varphi(\sigma_0^2)\end{bmatrix} + o_p(1).$$

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

$$\sqrt{n}\begin{bmatrix}\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\\ \widehat{\sigma}^2 - \sigma_0^2\end{bmatrix} = \mathcal{I}^{-1}\sqrt{n}\mathbb{P}_n\begin{bmatrix}\boldsymbol{\ell}^*(\boldsymbol{\theta}_0)\\ \varphi(\sigma_0^2)\end{bmatrix} + o_P(1) \xrightarrow{d} N(0, \mathcal{I}^{-1}\Sigma(\mathcal{I}^{-1})').$$

This completes the proof.

References

Betensky, R. A., Rabinowitz, D. and Tsiatis, A. A. (2001). Computationally simple accelerated failure time regression for interval-censored data. *Biometrika* **88**, 703-711.

Carroll, R. J., Ruppert, D., Stefanski, L. A. and Crainiceanu, C. M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*, 2nd edn. Chapman & Hall/CRC Press, Boca Raton.

Hammer, S. M., Katzenstein, D. A., Hughes, M. D., Gundacker, H., Schooley, R. T., Haubrich, R. H., Henry, W. K., Lederman, M. M., Phair, J. P., Niu, M., Hirsch, M. S. and Merigan, T. C. (1996). A trial comparing nucleoside monotherapy with combination therapy in HIV-infected adults with CD4 cell counts from 200 to 500 per cubic millimeter. *New Engl. J. Medicine* **335**, 1081-1089.

Heyde, C. C. and Morton, R. (1998). Multiple roots in general estimating equations. *Biometrika* **85**, 954-959.

Hu, P., Tsiatis, A. A. and Davidian, M. (1998). Estimating the parameters in the Cox model when covariate variables are measured with error. *Biometrics* **54**, 1407-1419.

Huang, J. (1996). Efficient estimation for the Cox model with interval censoring. *Ann. Statist.* **24**, 540-568.

Huang, J. and Rossin, A. J. (1997). Sieve estimation for the proportional odds failure-time regression model with interval censoring. *J. Amer. Statist. Assoc.* **92**, 960-967.

Huang, J. and Wellner, A. J. (1997). Interval censored survival data: a review of recent progress. In *Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis* (Edited by D. Lin and T. Fleming), 123-169. Springer-Verlag, New York.

Korosok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer, New York.

Ma, S. (2010). Mixed case interval censored data with a cured subgroup. *Statist. Sinica* **17**, 1165-1181.

Murphy, S. A. and van der Vaart, A. W. (1999). Observed information in semiparametric models. *Bernoulli* **5**, 381-412.

Rudin, W. (1973). *Functional Analysis*. McGraw-Hill, New York.

Schick, A. and Yu, Q. (2000). Consistency of the GMLE with mixed case interval-censored data. *Scand. J. Statist.* **27**, 45-55.

Song, X. and Huang, Y. (2005). On corrected score approach for proportional hazards model with covariate measurement error. *Biometrics* **61**, 702-714.

Song, X. and Ma, S. (2008). Multiple augmentation for interval-censored data with measurement error. *Statist. Medicine* **27**, 3178-3190.

Stefanski, L. A. and Carroll, R. J. (1987). Conditional scores and optimal scores in generalized linear measurement error models. *Biometrika* **69**, 331-342.

Tsiatis, A. A. and Davidian, M. (2001). A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika* **88**, 447-458.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.

Wen, C.-C. (2012). Cox regression for mixed case interval-censored data with covariate errors. *Lifetime Data Anal.* **18**, 321-338.

Wen, C.-C. and Chen, Y.-H. (2012). Conditional score approach to errors-in-variable current status data under the proportional odds model. *Scand. J. Statist.* **39**, 635-644.

Zeng, D., Cai, J. and Shen, Y. (2006). Semiparametric additive risks model for interval censored data. *Statist. Sinica* **16**, 287-302.

Zeng, D. and Lin, D. Y. (2007). Maximum likelihood estimation in semiparametric regression models with censored data (with Discussion). *J. Roy. Statist. Soc. Ser. B* **69**, 507-564.

Zhang, Z. and Sun, J. (2010). Interval censoring. *Statist. Methods Med. Res.* **19**, 53-70.

Zhu, L., Tong, X., and Sun, J. (2008). A transformation approach for the analysis of interval-censored failure time data. *Lifetime Data Anal.* **14** 161-178.

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

(Received April 2013; accepted August 2013)