Statistica Sinica **16**(2006), 861-881

# ERRORS-IN-COVARIATES EFFECT ON ESTIMATING FUNCTIONS: ADDITIVITY IN LIMIT AND NONPARAMETRIC CORRECTION

Yijian Huang and C. Y. Wang

*Emory University and Fred Hutchinson Cancer Research Center*

*Abstract:* We consider Poisson, logistic and Cox regressions when some covariates are not accurately ascertainable but contaminated with additive errors. Huang and Wang (1999, 2000, 2001) showed that the slope parameters can be consistently estimated via nonparametric correction, without imposing distributional assumptions on both the underlying true covariates and the errors. However, certain instrumental variables, particularly replicated error-contaminated covariates, are required. In this article, we discover that the error effect is additive in the limit on some properly formulated estimating functions. This finding gives rise to a new nonparametric correction technique that accommodates a broad variety of practically important, internal and external error-assessment data. Simulations for Cox regression with external reliability data are conducted, and the application to an AIDS study is presented as an illustration.

*Key words and phrases:* Cox regression, external data, functional modeling, generalized linear model, logistic regression, measurement error, nonlinear model, Poisson regression, proportional hazards model, reliability study.

## 1. Introduction

Covariates of regression analysis are often measured with error in medical research. Indeed, many measures are not even accurately ascertainable; examples include CD4 lymphocyte count in HIV/AIDS studies and blood pressure in cardiovascular disease research. Both imperfect instrumentation and biological fluctuation may contribute to the error. As is well known, naive analysis by treating mismeasured covariates as the truth in conventional inference procedures may result in substantial estimation bias; see, for example, Carroll, Ruppert and Stefanski (1995).

Various methods exist to address nonlinear regression when some covariates are contaminated with additive errors. Many of them can be found in the monograph of Carroll, Ruppert and Stefanski (1995). Regression calibration (e.g., Prentice (1982), Carroll and Stefanski (1990) and Gleser (1990)) and simulation extrapolation (Cook and Stefanski (1994), Stefanski and Cook (1995), and Carroll, Küchenhoff, Lombard and Stefanski (1996)) are two popular approaches in practice. However, they both yield approximate but inconsistent estimation in general. For consistent estimation, methods have been developed under either structural or functional modeling, i.e., with or without parametric distributional assumptions imposed on the true covariates (Carroll, Ruppert and Stefanski (1995, Sec. 1.2)). Whereas the maximum likelihood principle may be employed under structural modeling, functional modeling demands innovation. Two general functional modeling methods are the conditional score (Stefanski and Carroll (1985)) and parametric correction (Nakamura (1990, 1992), Stefanski (1989), Huang and Wang (2001)). Note that parametric distributional assumptions are required on the measurement errors for the aforementioned approaches, structural and functional modeling alike.

In recent years, weakened assumptions on the error distribution have been pursued under functional modeling. Buzas (1997) investigated errors with a symmetric distribution around zero. Most recently, Huang and Wang (1999, 2000, 2001) completely eliminated distributional assumptions, except for mild regularity conditions, on both the true covariates and the errors. They proposed a nonparametric correction technique which achieves consistent slope estimation for several generalized linear models. The notion of nonparametric correction is appealing for obvious reasons. However, this technique requires a second mismeasured replicate for each error-prone covariate or, more generally, instrumental variable, i.e., a measure related to the true covariate besides the mismeasured one (cf., Carroll, Ruppert and Stefanski (1995, Chap. 5)). For covariate measurement error analysis in general, one needs information about the measurement error in addition to the primary data, i.e., those necessary for the naive analysis. Additional data serving this purpose are termed error-assessment data in this article, and the aforementioned instrumental variable is one form of such data. Common types of error-assessment data that the technique of Huang and Wang (1999, 2000, 2001) can not accommodate include

(i) Reliability data external to the primary data;
(ii) Reliability data internal to the primary data, with the complication that the error-assessment subset may depend on the underlying true covariates and the response.

To broaden the applicability of nonparametric correction, this article further develops the methodology and proposes a new technique.

Focusing on widely employed Poisson, logistic and Cox regressions, we show that the errors-in-covariates effect is additive in the limit for some properly formulated estimating functions. This main finding gives rise to a new nonparametric correction technique that accommodates various internal and external error-assessment data. Section 2 presents the estimating functions and reveals the additive errors-in-covariates effect. Nonparametric estimation of the additive effect using error-assessment data is dealt with in Section 3. As an illustration, we present a nonparametric correction procedure in Section 4 for Cox regression with external error-assessment data, along with simulation results and application to an AIDS study. Section 5 concludes with discussion. Technical details are deferred to the appendices.

## 2. Estimating Functions and Errors-in-Covariates Effects

For Poisson, logistic, or Cox regression, let $\mathbf{X}$ be the true covariate vector and $Y$ the response. We first formulate estimating functions for each regression with an iid sample of $\{\mathbf{X}, Y\}$. The errors-in-covariates effect on these estimating functions is then investigated when $\mathbf{X}$ is replaced by its surrogate $\mathbf{W}$. Specifically, the error contamination of $\mathbf{X}$ follows the classical additive model:

$$\left.\begin{aligned} \mathbf{W} &= \mathbf{X} + \boldsymbol{\varepsilon}, \\ \boldsymbol{\varepsilon} &\perp \{\mathbf{X}, Y\}, \end{aligned}\right\} \tag{1}$$

where $\boldsymbol{\varepsilon}$ is the error vector and $\perp$ denotes independence. Neither the distribution of $\mathbf{X}$ nor that of $\boldsymbol{\varepsilon}$ is specified. The error $\boldsymbol{\varepsilon}$ may not have a mean of zero, may not be symmetrically distributed, and its components can be correlated. In case some covariates are accurately measured, the corresponding components of $\boldsymbol{\varepsilon}$ are zeros.

### 2.1. Poisson regression

Poisson regression is widely applied in the analysis of count data. Write $\mathcal{E}$ as expectation. Response $Y$ is a frequency measure and the model postulates

$$\mathcal{E}(Y \mid \mathbf{X}) = \exp(\alpha + \boldsymbol{\beta}^T \mathbf{X}), \tag{2}$$

where $\alpha$ is the intercept and $\boldsymbol{\beta}$ the slope vector. The standard estimation procedure for $(\alpha, \boldsymbol{\beta}^T)^T$ is to solve the (normalized) score function in terms of $(a, \mathbf{b}^T)^T$:

$$\widehat{\mathcal{E}}\left[\{Y - \exp(a + \mathbf{b}^T \mathbf{X})\} \begin{pmatrix} 1 \\ \mathbf{X} \end{pmatrix}\right],$$

where $\widehat{\mathcal{E}}$ represents the empirical sample average. In this article, we treat $\alpha$ as a nuisance parameter, for two reasons. First, typically $\alpha$ is not of as much interest as $\boldsymbol{\beta}$ in practice. Second, for generalized linear measurement-error models, even with replicated mismeasured covariates, $\alpha$ is not identifiable unless certain distributional assumptions are imposed on $\boldsymbol{\varepsilon}$ (see Huang and Wang (1999; 2001, Sec. 4.1)). Incidentally, one assumption sufficient for identifiability is that $\boldsymbol{\varepsilon}$ is symmetrically distributed around $\mathbf{0}$. Therefore, we algebraically reduce the above estimating function to the following one for $\boldsymbol{\beta}$ only:

$$\begin{aligned}\widetilde{\boldsymbol{\Psi}}^{\mathrm{P}}(\mathbf{b}) &= \widehat{\mathcal{E}}[\{\mathbf{X}-\widehat{\mathcal{E}}(\mathbf{X})\}Y] - \widehat{\mathcal{E}}(Y)\left[\frac{\widehat{\mathcal{E}}\{\mathbf{X}\exp(\mathbf{b}^T\mathbf{X})\}}{\widehat{\mathcal{E}}\{\exp(\mathbf{b}^T\mathbf{X})\}} - \widehat{\mathcal{E}}(\mathbf{X})\right] \\ &\equiv \widehat{\mathcal{E}}[\{\mathbf{X}-\widehat{\mathcal{E}}(\mathbf{X})\}Y] - \widehat{\mathcal{E}}(Y)\widehat{\boldsymbol{\phi}}_1(\mathbf{b};\mathbf{X}),\end{aligned} \tag{3}$$

where

$$\widehat{\boldsymbol{\phi}}_H(\mathbf{b};\mathbf{G}) \equiv \frac{\widehat{\mathcal{E}}\{H\mathbf{G}\exp(\mathbf{b}^T\mathbf{G})\}}{\widehat{\mathcal{E}}\{H\exp(\mathbf{b}^T\mathbf{G})\}} - \widehat{\mathcal{E}}(\mathbf{G}) \tag{4}$$

for random vector $\mathbf{G}$ and non-negative random variable $H$.

### 2.2. Logistic regression

Logistic regression is popular given Bernoulli outcomes. With binary response $Y$, the model postulates

$$\Pr(Y=1 \mid \mathbf{X}) = \frac{\exp(\alpha+\boldsymbol{\beta}^T\mathbf{X})}{1+\exp(\alpha+\boldsymbol{\beta}^T\mathbf{X})}. \tag{5}$$

Again, $\alpha$ is considered as a nuisance and $\boldsymbol{\beta}$ is the parameter of interest. Given that the score function is not correction-amenable (Stefanski (1989)), Huang and Wang (2001) suggested a pair of weighted score functions for $(\alpha,\boldsymbol{\beta}^T)^T$:

$$\begin{aligned}&\widehat{\mathcal{E}}\left[\{Y-1+Y\exp(-a-\mathbf{b}^T\mathbf{X})\}\begin{pmatrix}1\\ \mathbf{X}\end{pmatrix}\right],\\ &\widehat{\mathcal{E}}\left[\{Y+(Y-1)\exp(a+\mathbf{b}^T\mathbf{X})\}\begin{pmatrix}1\\ \mathbf{X}\end{pmatrix}\right].\end{aligned}$$

These reduce to estimating functions for $\boldsymbol{\beta}$:

$$\left.\begin{aligned}\widetilde{\boldsymbol{\Psi}}^{\mathrm{L}}_{-}(\mathbf{b}) &= \widehat{\mathcal{E}}[\{\mathbf{X}-\widehat{\mathcal{E}}(\mathbf{X})\}(Y-1)] - \widehat{\mathcal{E}}(Y-1)\widehat{\boldsymbol{\phi}}_Y(-\mathbf{b};\mathbf{X}),\\ \widetilde{\boldsymbol{\Psi}}^{\mathrm{L}}_{+}(\mathbf{b}) &= \widehat{\mathcal{E}}[\{\mathbf{X}-\widehat{\mathcal{E}}(\mathbf{X})\}Y] - \widehat{\mathcal{E}}(Y)\widehat{\boldsymbol{\phi}}_{Y-1}(\mathbf{b};\mathbf{X}),\end{aligned}\right\} \tag{6}$$

as being the basis to form a class of consistent estimators. The approach of Huang and Wang (2001) can be used to obtain an efficient estimator in this class.

### 2.3. Cox regression

The proportional hazards model (Cox (1972)) is one of the most widely applied models for censored survival data. Of interest is the relationship between survival time $T^0$ and covariate vector $\mathbf{X}$, but $T^0$ is subject to censoring by time $C$ and thus is not fully observed. The response $Y$ consists of observed variables

$T \equiv T^0 \wedge C$ and $\Delta \equiv I(T^0 \leq C)$, i.e., $Y \equiv \{T, \Delta\}$, where $\wedge$ denotes minimum and $I(\cdot)$ is the indicator function. The model postulates a semiparametric formulation of the cumulative hazard function $\Lambda(\cdot \mid \mathbf{X})$ of $T^0$ given $\mathbf{X}$, and a conditional independence censorship:

$$\left.\begin{aligned} \Lambda(dt \mid \mathbf{X}) &= \Lambda(dt) \exp(\boldsymbol{\beta}^T \mathbf{X}), \\ T^0 &\perp C \mid \mathbf{X}, \end{aligned}\right\} \tag{7}$$

where $\boldsymbol{\beta}$ is the parameter of interest and $\Lambda(\cdot)$ an unspecified baseline cumulative hazard function. In fact, $\Lambda(\cdot)$ is analogous to the intercept in the Poisson- and logistic-regression models. Thus, not surprisingly, in the presence of covariate measurement error it may not be identifiable even when replicated mismeasured covariates are available (Huang and Wang (2000)). With the functional representation of Huang and Wang (2000), the (normalized) partial-score function can be written as

$$\widetilde{\boldsymbol{\Psi}}^{\mathrm{C}}(\mathbf{b}) = \widehat{\mathcal{E}}[\{\mathbf{X} - \widehat{\mathcal{E}}(\mathbf{X})\}\Delta I(T \leq \tau)] - \int_0^\tau \widehat{\boldsymbol{\phi}}_{I(T \geq t)}(\mathbf{b}; \mathbf{X})\, d\widehat{\mathcal{E}}\{\Delta I(T \leq t)\} \tag{8}$$

with time limit $\tau$ satisfying $\Pr(T \geq \tau) > 0$.

Cox regression can accommodate time-varying covariates as well, with proper generalizations of (7) and (8). Then, in the presence of measurement error, the error is a time-indexed process and the measurement error model (1) can be generalized accordingly. This would pose little additional difficulty for the development in this article except for notation. To focus on main ideas, we restrict our attention to the case of time-independent covariates.

### 2.4. Errors-in-covariates effects

A normalized estimating function is root-consistent if its limit has a unique zero-crossing at the parameter of interest. This is the case with a standard estimating function with $\mathbf{X}$ observed. When $\mathbf{X}$ is prone to error contamination, the naive estimating function, i.e., with $\mathbf{X}$ replaced by $\mathbf{W}$, has its limit with a shifted zero-crossing in general, resulting in bias. To pursue consistent estimation, both parametric correction (Nakamura (1990) and Stefanski (1989)) and the nonparametric correction technique of Huang and Wang (1999, 2000, 2001) take the strategy of constructing an estimating function to achieve the same limit, or at least the same limiting zero-crossing, as the one in the absence of measurement error. In the literature, the assessment of an errors-in-covariates effect on the limit of an estimating function is mostly qualitative due to mathematical intractability. Here we discover that a quantitative assessment becomes possible for estimating functions (3), (6) and (8), thanks to their proper formulation, and can be effectively exploited to develop a new nonparametric correction technique.

Consider a sample with contaminated covariates consisting of iid replicates of $\{\mathbf{W}, Y\}$. The limits of estimating functions (3), (6) and (8) can be easily determined given that they all are functionals of empirical processes. Even though Poisson, logistic, and Cox regressions deal with different outcomes, their estimating functions for $\boldsymbol{\beta}$ are surprisingly similar in that all components but $\widehat{\boldsymbol{\phi}}_{\cdot}(\cdot;\cdot)$ are invariant in the limit when $\mathbf{X}$ is replaced by $\mathbf{W}$. We can thus focus our investigation on the limiting behavior of $\widehat{\boldsymbol{\phi}}_{\cdot}(\cdot;\cdot)$. From (4), $\widehat{\boldsymbol{\phi}}_H(\mathbf{b};\mathbf{G})$ is a functional of empirical processes. Under mild regularity conditions, its limit is

$$\boldsymbol{\phi}_H(\mathbf{b};\mathbf{G}) \equiv \frac{\mathcal{E}\{H\mathbf{G}\exp(\mathbf{b}^T\mathbf{G})\}}{\mathcal{E}\{H\exp(\mathbf{b}^T\mathbf{G})\}} - \mathcal{E}(\mathbf{G}). \tag{9}$$

Useful properties of $\boldsymbol{\phi}_{\cdot}(\cdot;\cdot)$ are summarized in Appendix A. In particular, Property 1 states that the errors-in-covariates effect on component $\widehat{\boldsymbol{\phi}}_{\cdot}(\cdot;\cdot)$ in estimating functions (3), (6) and (8) is additive in the limit. Specifically,

$$\boldsymbol{\phi}_{\cdot}(\mathbf{b};\mathbf{W}) = \boldsymbol{\phi}_{\cdot}(\mathbf{b};\mathbf{X}) + \boldsymbol{\phi}_1(\mathbf{b};\boldsymbol{\varepsilon}). \tag{10}$$

Apparently, these naive estimating functions are no longer root-consistent in general. Most importantly, it is implied that they can be corrected once $\boldsymbol{\phi}_1(\mathbf{b};\boldsymbol{\varepsilon})$ is consistently estimated.

## 3. Nonparametric Estimation of $\boldsymbol{\phi}_1(\mathbf{b};\boldsymbol{\varepsilon})$

Findings in Section 2 naturally motivate correction methods and the key is to estimate $\boldsymbol{\phi}_1(\mathbf{b};\boldsymbol{\varepsilon})$. This can be readily achieved if the distribution of $\boldsymbol{\varepsilon}$ is assumed, or consistently estimated with validation dard; the former results in parametric correction (Nakamura (1990) and Stefanski (1989)). However, we are concerned with nonparametric correction and with accurate covariates unascertainable. In the following, we first reduce the estimation problem according to error clustering, then identify a structure in various realistic error-assessment data, and finally present consistent estimators for $\boldsymbol{\phi}_1(\mathbf{b};\boldsymbol{\varepsilon})$.

### 3.1. Error clustering

Estimation of $\boldsymbol{\phi}_1(\mathbf{b};\boldsymbol{\varepsilon})$ may take advantage of the independence structure in $\boldsymbol{\varepsilon}$, if it exists. Such a structure often results from independent error-contamination mechanisms for different covariates. Take cardiovascular disease research as an example: The measurement errors associated with low-density lipoprotein cholesterol and high-density lipoprotein cholesterol might be correlated, but they are likely independent of the error associated with systolic blood pressure. Suppose that $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_0^T, \ldots, \boldsymbol{\varepsilon}_K^T)^T$ with $K+1$ mutually independent clusters. Furthermore, all the errors with zero variance, if any, are grouped into $\boldsymbol{\varepsilon}_0$, corresponding to those covariates either accurately measured or contaminated by a constant. Correspondingly, write $\mathbf{b} = (\mathbf{b}_0^T, \ldots, \mathbf{b}_K^T)^T$. From Properties 2 and 3 in Appendix A, we then have $\boldsymbol{\phi}_1(\mathbf{b}; \boldsymbol{\varepsilon}) = \{\mathbf{0}^T, \boldsymbol{\phi}_1(\mathbf{b}_1; \boldsymbol{\varepsilon}_1)^T, \ldots, \boldsymbol{\phi}_1(\mathbf{b}_K; \boldsymbol{\varepsilon}_K)^T\}^T$. Thus, the estimation of $\boldsymbol{\phi}_1(\mathbf{b}; \boldsymbol{\varepsilon})$ reduces to that of $\boldsymbol{\phi}_1(\mathbf{b}_k; \boldsymbol{\varepsilon}_k)$ for $k = 1, \ldots, K$.

### 3.2. Error-assessment data

Suppose that an error-assessment data set is available for each error cluster, consisting of iid observations. We allow the data set to be internal, external, or even a combination of the two, relative to the primary set that is used to construct the naive estimating function given in Section 2. Just like their relationship with the primary set, the $K$ error-assessment sets themselves may or may not be external to each other.

Focus on the $k$th error-assessment set, i.e., specific to $\boldsymbol{\varepsilon}_k$. Denote the underlying measure by $\mathbf{Z}$, which is not accurately ascertainable but is subject to the same error-contamination mechanism as the $k$th cluster of $\mathbf{X}$ in the primary set. To understand the data set, there are two typical approaches to error assessment:

(i) One involves a reliability study. That is, possibly random and even $\mathbf{Z}$-dependent $R$ ($\geq 2$) copies of $\mathbf{Z} + \boldsymbol{\varepsilon}_k$ are available for each individual and they are independent of each other given $\mathbf{Z}$.

(ii) The other is to take a second measure through a new measurement mechanism. Most often, the intention is to validate the first measure, i.e., to accurately measure $\mathbf{Z}$. Since this is not achievable in our case, we merely require that the second measure is independent of the first one given $\mathbf{Z}$.

We recognize a data structure common to the above two types of error-assessment data and possibly others. That is, a pair of error-contaminated measures $\{\mathbf{U}, \mathbf{V}\}$ are observed from each individual in the sample, following

$$\left.\begin{aligned} \mathbf{U} &= \mathbf{Z} + \boldsymbol{\varepsilon}_k, \\ \mathbf{V} &= \mathbf{Z} + \boldsymbol{\eta}, \\ \boldsymbol{\varepsilon}_k &\perp \boldsymbol{\eta} \quad \perp \quad \mathbf{Z}. \end{aligned}\right\} \tag{11}$$

Here, $\mathbf{Z}$ is subject to error contamination by $\boldsymbol{\varepsilon}_k$ in $\mathbf{U}$, and the error contamination $\boldsymbol{\eta}$ of $\mathbf{Z}$ in $\mathbf{V}$ may or may not have the same distribution. In the case of aforementioned reliability study, any random select of two copies of $\mathbf{Z} + \boldsymbol{\varepsilon}_k$ can serve as $\{\mathbf{U}, \mathbf{V}\}$ and there are $R(R-1)$ such combinations.

### 3.3. Consistent estimation

First, consider the $k$th error cluster. Reciprocally, $\boldsymbol{\varepsilon}_k$ is contaminated by $\mathbf{Z}$ in $\mathbf{U}$ and by $-\boldsymbol{\eta}$ in $\mathbf{U} - \mathbf{V} = \boldsymbol{\varepsilon}_k - \boldsymbol{\eta}$. Since $\mathbf{Z}$ and $-\boldsymbol{\eta}$ are independent, Property 4 in Appendix A suggests two consistent estimators for $\boldsymbol{\phi}_1(\mathbf{b}_k;\boldsymbol{\varepsilon}_k)$:

$$\widehat{\boldsymbol{\varphi}}_{\rm a}(\mathbf{b}_k) = \frac{\widehat{\mathcal{E}}\mathcal{A}\{(\mathbf{U}-\mathbf{V})\exp(\mathbf{b}_k^T\mathbf{U})\}}{\widehat{\mathcal{E}}\mathcal{A}\{\exp(\mathbf{b}_k^T\mathbf{U})\}} - \widehat{\mathcal{E}}\mathcal{A}(\mathbf{U}-\mathbf{V}), \tag{12}$$

$$\widehat{\boldsymbol{\varphi}}_{\rm b}(\mathbf{b}_k) = \frac{\widehat{\mathcal{E}}\mathcal{A}[\mathbf{U}\exp\{\mathbf{b}_k^T(\mathbf{U}-\mathbf{V})\}]}{\widehat{\mathcal{E}}\mathcal{A}[\exp\{\mathbf{b}_k^T(\mathbf{U}-\mathbf{V})\}]} - \widehat{\mathcal{E}}\mathcal{A}(\mathbf{U}). \tag{13}$$

In the above expressions, $\mathcal{A}$ is the operator that averages over the $R(R-1)$ combinations of $\{\mathbf{U},\mathbf{V}\}$ within an individual in the case of reliability study, and it is otherwise unnecessary. This technique to handle replicated measures was used in Huang and Wang (1999, 2000, 2001).

Now, write $\widehat{\boldsymbol{\varphi}}_{\rm a}(\mathbf{b})=\{\mathbf{0}^T,\widehat{\boldsymbol{\varphi}}_{\rm a}(\mathbf{b}_1)^T,\ldots,\widehat{\boldsymbol{\varphi}}_{\rm a}(\mathbf{b}_K)^T\}^T$ and $\widehat{\boldsymbol{\varphi}}_{\rm b}(\mathbf{b})=\{\mathbf{0}^T,\widehat{\boldsymbol{\varphi}}_{\rm b}(\mathbf{b}_1)^T,\ldots,\widehat{\boldsymbol{\varphi}}_{\rm b}(\mathbf{b}_K)^T\}^T$. Subject to regularity conditions, every weighted average of them is consistent for $\boldsymbol{\phi}_1(\mathbf{b};\boldsymbol{\varepsilon})$. Among all the possible weights, one can derive an optimal one for the (asymptotically) efficient estimation of $\boldsymbol{\beta}$ in the class. The corrected estimating function evaluated at $\boldsymbol{\beta}$ achieves the minimum asymptotic variance with this weight. Thus, this weight depends on not only the error-assessment sets but also the primary set unless the former are external to the latter. This optimization is straightforward and will be addressed in later sections.

## 4. Nonparametric Correction: Cox Regression with External Error Assessment

We illustrate the proposed nonparametric correction technique through Cox regression with external error assessment. Simulation results are presented for the setup of external reliability data, and the procedure is applied to an AIDS study.

### 4.1. The inference procedure

The primary data set consists of iid replicates of $\{\mathbf{W},T,\Delta\}$, independent of the error-assessment data described in Section 3. For any weighted average $\widehat{\boldsymbol{\varphi}}(\mathbf{b})$ of $\widehat{\boldsymbol{\varphi}}_{\rm a}(\mathbf{b})$ and $\widehat{\boldsymbol{\varphi}}_{\rm b}(\mathbf{b})$, nonparametric correction of the naive estimating function gives:

$$\begin{aligned}\widehat{\boldsymbol{\Psi}}^{\rm C}(\mathbf{b}) = {} & \widehat{\mathcal{E}}[\{\mathbf{W}-\widehat{\mathcal{E}}(\mathbf{W})\}\Delta I(T\le\tau)] - \int_0^\tau \widehat{\boldsymbol{\phi}}_{I(T\ge t)}(\mathbf{b};\mathbf{W})\,d\widehat{\mathcal{E}}\{\Delta I(T\le t)\}\\ & +\widehat{\boldsymbol{\varphi}}(\mathbf{b})\widehat{\mathcal{E}}\{\Delta I(T\le\tau)\}.\end{aligned} \tag{14}$$

As shown in Appendix B, asymptotically a zero-crossing, $\widehat{\boldsymbol{\beta}}$ say, exists and any zero-crossing is consistent and asymptotically normal subject to mild regularity conditions.

These nonparametric-correction estimators of $\boldsymbol{\beta}$ form a class. For an asymptotically efficient one in this class, the optimal weight for the corresponding $\widehat{\boldsymbol{\varphi}}(\mathbf{b})$ is such that $\widehat{\boldsymbol{\varphi}}(\boldsymbol{\beta})$ achieves the minimum asymptotic variance among all weighted averages of $\widehat{\boldsymbol{\varphi}}_{\mathrm{a}}(\boldsymbol{\beta})$ and $\widehat{\boldsymbol{\varphi}}_{\mathrm{b}}(\boldsymbol{\beta})$, since the error assessment is external. Write the asymptotic variance of $\{\widehat{\boldsymbol{\varphi}}_{\mathrm{a}}(\boldsymbol{\beta})^T, \widehat{\boldsymbol{\varphi}}_{\mathrm{b}}(\boldsymbol{\beta})^T\}^T$ as $\boldsymbol{\Sigma}$. Elementary linear algebra then gives the optimal weight as a function of $\boldsymbol{\Sigma}$. Given that $\boldsymbol{\Sigma}$ is unknown, Appendix C presents a consistent estimate, and using it instead yields an equally efficient $\widehat{\boldsymbol{\beta}}$. We refer to this estimator as NC. To differentiate, nonparametric-correction estimators of $\boldsymbol{\beta}$ using $\widehat{\boldsymbol{\varphi}}_{\mathrm{a}}(\mathbf{b})$ and $\widehat{\boldsymbol{\varphi}}_{\mathrm{b}}(\mathbf{b})$ are labeled $\mathrm{NC_a}$ and $\mathrm{NC_b}$, respectively.

For interval estimation, the sandwich variance estimate can be constructed, with the formulas given in Appendix C.

### 4.2. Simulations with reliability data

The performance of the nonparametric-correction estimators was investigated through systematic numerical experiments with the setup of external reliability data. For comparison, the naive (NV), regression calibration (RC), and parametric-correction (PC) estimators are also constructed. The naive and regression calibration approaches use mismeasured $\mathbf{W}$ and an estimate of $\mathcal{E}(\mathbf{X}\,|\,\mathbf{W})$, respectively, in place of $\mathbf{X}$ in (8). For the estimation of $\mathcal{E}(\mathbf{X} \mid \mathbf{W})$, the formulas given in Carroll, Ruppert and Stefanski (1995, Sec. 3.4.2) are adapted to this setup. Note that we now use the external reliability data to obtain an estimate $\widehat{\mathrm{var}}(\boldsymbol{\varepsilon})$ of the error variance. The parametric-correction estimator is obtained under a normal error assumption, in which case $\boldsymbol{\phi}_1(\mathbf{b};\boldsymbol{\varepsilon}) = \mathrm{var}(\boldsymbol{\varepsilon})\mathbf{b}$. Thus, the parametrically corrected estimating function is

$$\widehat{\mathcal{E}}[\{\mathbf{W} - \widehat{\mathcal{E}}(\mathbf{W})\}\Delta I(T \le \tau)] - \int_0^\tau \widehat{\boldsymbol{\phi}}_{I(T\ge t)}(\mathbf{b};\mathbf{W})\, d\widehat{\mathcal{E}}\{\Delta I(T \le t)\} \\ + \widehat{\mathrm{var}}(\boldsymbol{\varepsilon})\mathbf{b}\widehat{\mathcal{E}}\{\Delta I(T \le \tau)\},$$

which essentially coincides with the basic one of the two asymptotically equivalent proposals in Nakamura (1992). For interval estimation, the sandwich variance estimates are formulated in the same fashion as those for the nonparametric-correction estimators.

All these estimating functions involve the time limit $\tau$. As is customary, it is set to be large enough to cover all follow-up times for a specific data set. With a correction approach, parametric and nonparametric alike, the corrected estimating function is not guaranteed to have a zero-crossing (Huang and Wang (2000) and Nakamura (1992)). In such a case, we define the estimator as the zero-crossing of the tangent plane at $\mathbf{b} = \mathbf{0}$. This is a finite-sample issue, and asymptotically a zero-crossing exists.

We first considered Cox regression with a single and error-contaminated covariate. For the primary set, we specified the unit exponential baseline survival time and standard normal true covariate with $\beta = 1$. Independent of the survival time and covariate, the censoring time was uniformly distributed on [0,7]; the censoring rate was 20%. The external reliability set was constructed with the standard normal underlying true measure and with two copies of the error-contaminated measures for each individual. The sample sizes for the primary and reliability sets were set to be the same. Table 1 reports the summary statistics from 1,000 iterations for each combination of sample sizes 400 and 800, normal and uniform error distributions, and error standard deviation $\sigma_\varepsilon$ from 0.25 to 0.75. As shown, both the naive and regression calibration estimators exhibit notable bias and poor confidence interval coverage; these problems are especially serious for the naive estimator. The parametric-correction estimator performs well when the distributional assumption of the error holds. However, it may suffer otherwise from serious root-finding failure and from bias. In contrast, the nonparametric-correction estimators have reasonably small bias and accurate confidence interval coverage. The efficiency advantage of NC over $\mathrm{NC_a}$ and $\mathrm{NC_b}$ becomes apparent with increasing error contamination. Meanwhile, we notice that the root-finding failure rate increases for both parametric and nonparametric correction approaches as the error contamination becomes more serious and the sample size decreases. In addition, the distributions of these estimators become more skewed, which might explain the increasing discrepancy between their standard deviations and standard errors. This suggests that moderate sample size might be necessary for the correction approaches in the presence of serious error contamination.

Multiple covariates were also studied. Table 2 presents a study with one covariate error-prone and the other accurately measured. In the primary set, the baseline survival time followed the unit exponential distribution, and the two underlying true covariates followed the standard bivariate normal distribution with correlation coefficient $\rho$. The censoring time was independent of them and had a uniform distribution on [0, 10]. We set $\boldsymbol{\beta} = (1, 1)^T$. Corresponding to $\rho = 0.5$, 0 and $-0.5$, the censoring rates were 21%, 18% and 15%, respectively. The external reliability set had the same structure as that for the study reported in Table 1. The primary and reliability sets were each of size 400. Various values of $\rho$ and different error distributions were investigated, and 1,000 iterations were conducted under each scenario. The relative performance of these various estimators for the error-prone covariate appears similar to that observed in the previous single-covariate model. Meanwhile, the regression coefficient estimation for the accurately-measured covariate might also be affected by the measurement error.

Table 1. Simulation summaries of estimators for the single-covariate Cox regression model.

| $\sigma_\varepsilon$ | | NV | RC | PC | $NC_a$ | $NC_b$ | NC | NV | RC | PC | $NC_a$ | $NC_b$ | NC |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | primary/reliability set size = 400 | | | | | | primary/reliability set size = 800 | | | | | |
| | | $\varepsilon \sim$ Normal | | | | | | | | | | | |
| 0.25 | F | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | B | −86 | −28 | 10 | 10 | 10 | 8 | −89 | −31 | 5 | 5 | 4 | 3 |
| | D | 70 | 75 | 84 | 86 | 85 | 85 | 49 | 53 | 59 | 59 | 59 | 59 |
| | $\widehat{\mathrm{D}}$ | 68 | 72 | 80 | 82 | 81 | 80 | 48 | 51 | 57 | 58 | 57 | 57 |
| | C | 74.6 | 91.1 | 93.5 | 93.8 | 93.5 | 93.4 | 53.3 | 89.1 | 93.9 | 93.9 | 93.7 | 93.7 |
| 0.50 | F | 0.0 | 0.0 | 0.0 | 1.1 | 0.3 | 0.1 | 0.0 | 0.0 | 0.0 | 0.5 | 0.0 | 0.0 |
| | B | −277 | −94 | 31 | 40 | 33 | 23 | −280 | −98 | 16 | 17 | 17 | 11 |
| | D | 63 | 83 | 129 | 175 | 139 | 135 | 45 | 60 | 91 | 96 | 94 | 92 |
| | $\widehat{\mathrm{D}}$ | 61 | 79 | 119 | 153 | 129 | 122 | 43 | 56 | 83 | 94 | 88 | 85 |
| | C | 1.1 | 76.2 | 93.7 | 94.0 | 95.1 | 93.8 | 0.0 | 57.5 | 94.0 | 94.7 | 94.8 | 94.1 |
| 0.75 | F | 0.0 | 0.0 | 6.0 | 10.0 | 9.4 | 3.6 | 0.0 | 0.0 | 0.9 | 4.2 | 3.2 | 0.6 |
| | B | −461 | −152 | 55 | 102 | 82 | 58 | −464 | −158 | 52 | 58 | 64 | 39 |
| | D | 55 | 100 | 395 | 577 | 417 | 316 | 39 | 71 | 171 | 287 | 252 | 192 |
| | $\widehat{\mathrm{D}}$ | 52 | 96 | 245 | 334 | 314 | 251 | 37 | 68 | 156 | 212 | 224 | 171 |
| | C | 0.0 | 61.0 | 95.8 | 93.8 | 94.4 | 93.6 | 0.0 | 36.4 | 96.9 | 95.0 | 95.3 | 94.8 |
| | | $\varepsilon \sim$ Uniform | | | | | | | | | | | |
| 0.25 | F | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | B | −88 | −31 | 7 | 5 | 6 | 5 | −92 | −35 | 1 | −1 | −1 | −1 |
| | D | 67 | 72 | 81 | 81 | 81 | 80 | 49 | 52 | 58 | 58 | 58 | 58 |
| | $\widehat{\mathrm{D}}$ | 67 | 72 | 80 | 80 | 80 | 79 | 48 | 51 | 56 | 56 | 56 | 56 |
| | C | 72.6 | 92.6 | 95.3 | 94.7 | 94.5 | 94.5 | 49.5 | 88.3 | 94.3 | 93.7 | 94.0 | 93.9 |
| 0.50 | F | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | B | −276 | −93 | 41 | 18 | 16 | 14 | −278 | −97 | 26 | 5 | 4 | 2 |
| | D | 59 | 76 | 122 | 115 | 112 | 111 | 43 | 56 | 84 | 81 | 79 | 78 |
| | $\widehat{\mathrm{D}}$ | 59 | 76 | 115 | 112 | 107 | 106 | 42 | 54 | 79 | 77 | 74 | 74 |
| | C | 0.5 | 74.3 | 95.1 | 95.1 | 94.4 | 94.0 | 0.0 | 55.2 | 94.1 | 93.7 | 93.0 | 92.7 |
| 0.75 | F | 0.0 | 0.0 | 12.0 | 1.9 | 0.0 | 0.0 | 0.0 | 0.0 | 2.8 | 0.2 | 0.0 | 0.0 |
| | B | −457 | −147 | 103 | 40 | 38 | 33 | −458 | −152 | 152 | 19 | 12 | 9 |
| | D | 49 | 87 | 589 | 182 | 171 | 168 | 36 | 63 | 180 | 140 | 109 | 108 |
| | $\widehat{\mathrm{D}}$ | 50 | 88 | 278 | 187 | 155 | 151 | 35 | 61 | 196 | 132 | 103 | 101 |
| | C | 0.0 | 56.6 | 97.5 | 94.7 | 94.8 | 94.4 | 0.0 | 30.2 | 99.1 | 94.0 | 93.2 | 93.0 |

F: root-finding failure (%); B: bias ($\times 1,000$); D: standard deviation ($\times 1,000$); $\widehat{\mathrm{D}}$: average of the standard error ($\times 1,000$); C: coverage (%) of the 95% Wald-type confidence interval.

Table 2. Simulation summaries of estimators for the double-covariate Cox regression model.

| $\rho$ | | NV | | RC | | PC | | $\mathrm{NC_a}$ | | $\mathrm{NC_b}$ | | NC | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $\varepsilon \sim$ Normal | | | | | | | | | | | |
| 0.5 | F | 0.0 | | 0.0 | | 0.0 | | 2.1 | | 0.4 | | 0.4 | |
| | B | −315 | 22 | −84 | −94 | 35 | 0 | 28 | −5 | 36 | −1 | 21 | −1 |
| | D | 66 | 78 | 94 | 82 | 145 | 92 | 159 | 97 | 158 | 92 | 145 | 91 |
| | $\widehat{\mathrm{D}}$ | 65 | 78 | 92 | 82 | 136 | 90 | 160 | 91 | 156 | 91 | 138 | 89 |
| | C | 0.5 | 94.0 | 82.9 | 77.2 | 95.4 | 95.7 | 95.2 | 93.9 | 95.9 | 95.2 | 94.8 | 95.2 |
| 0 | F | 0.0 | | 0.0 | | 0.0 | | 1.5 | | 0.3 | | 0.2 | |
| | B | −278 | −101 | −96 | −100 | 25 | 9 | 22 | 7 | 25 | 9 | 15 | 5 |
| | D | 60 | 68 | 78 | 69 | 122 | 92 | 142 | 101 | 128 | 93 | 125 | 93 |
| | $\widehat{\mathrm{D}}$ | 60 | 69 | 78 | 70 | 116 | 88 | 128 | 92 | 125 | 91 | 121 | 92 |
| | C | 1.0 | 65.6 | 74.2 | 66.2 | 95.4 | 95.5 | 95.4 | 95.3 | 96.6 | 95.8 | 95.3 | 95.8 |
| −0.5 | F | 0.0 | | 0.0 | | 0.0 | | 2.2 | | 0.6 | | 0.4 | |
| | B | −326 | −214 | −98 | −100 | 29 | 20 | 25 | 17 | 30 | 21 | 17 | 11 |
| | D | 64 | 71 | 90 | 80 | 144 | 120 | 161 | 130 | 153 | 126 | 145 | 121 |
| | $\widehat{\mathrm{D}}$ | 63 | 71 | 89 | 80 | 135 | 114 | 156 | 127 | 150 | 123 | 137 | 115 |
| | C | 0.5 | 14.9 | 77.9 | 74.9 | 95.7 | 95.2 | 95.4 | 95.8 | 95.5 | 95.8 | 94.9 | 95.4 |
| | | $\varepsilon \sim$ Uniform | | | | | | | | | | | |
| 0.5 | F | 0.0 | | 0.0 | | 0.0 | | 0.3 | | 0.0 | | 0.0 | |
| | B | −311 | 30 | −78 | −85 | 55 | 10 | 28 | 9 | 27 | 10 | 24 | 10 |
| | D | 67 | 80 | 93 | 84 | 147 | 92 | 140 | 92 | 134 | 91 | 133 | 90 |
| | $\widehat{\mathrm{D}}$ | 63 | 78 | 88 | 82 | 134 | 90 | 131 | 89 | 123 | 89 | 122 | 89 |
| | C | 0.4 | 93.4 | 81.5 | 79.1 | 94.9 | 94.5 | 95.6 | 94.8 | 94.7 | 94.9 | 94.2 | 94.9 |
| 0 | F | 0.0 | | 0.0 | | 0.0 | | 0.1 | | 0.0 | | 0.0 | |
| | B | −272 | −91 | −87 | −91 | 45 | 25 | 24 | 17 | 23 | 17 | 20 | 16 |
| | D | 60 | 69 | 78 | 69 | 123 | 89 | 118 | 88 | 115 | 87 | 114 | 87 |
| | $\widehat{\mathrm{D}}$ | 58 | 69 | 76 | 70 | 114 | 89 | 112 | 87 | 107 | 86 | 106 | 86 |
| | C | 0.6 | 71.2 | 75.2 | 72.2 | 95.6 | 94.7 | 95.6 | 94.9 | 94.3 | 95.1 | 94.2 | 95.1 |
| −0.5 | F | 0.0 | | 0.0 | | 0.0 | | 0.2 | | 0.0 | | 0.0 | |
| | B | −319 | −205 | −88 | −90 | 56 | 43 | 29 | 25 | 27 | 24 | 24 | 21 |
| | D | 61 | 68 | 85 | 75 | 140 | 115 | 134 | 111 | 128 | 107 | 127 | 106 |
| | $\widehat{\mathrm{D}}$ | 62 | 71 | 86 | 79 | 134 | 114 | 132 | 112 | 124 | 107 | 122 | 106 |
| | C | 0.0 | 18.1 | 80.6 | 76.9 | 96.5 | 96.1 | 96.6 | 96.0 | 96.0 | 96.5 | 95.8 | 96.4 |

Two columns under each setting correspond to the two covariates, with the first being error-prone and the second accurately measured. See the footnote of Table 1 for the notation.

### 4.3. Application to an AIDS trial

The development in this article is motivated by AIDS research, where CD4 count is an important biomarker measuring functionality of the immune system.

Unfortunately, CD4 count is not accurately ascertainable and is subject to substantial measurement error. Given that only primary data are available in many studies, the proposed methodology facilitates an analysis with error-assessment data obtained from external sources. We illustrate it using data from the AIDS Clinical Trials Group (ACTG) 175 trial.

The ACTG 175 trial evaluated treatments with either a single nucleoside or two nucleosides in HIV-1 infected adults whose screening CD4 counts were from 200 to 500 per cubic millimeter (Hammer et al. (1996)). The participants were randomized to one of four treatments: Zidovudine (ZDV), Zidovudine and Didanosine (ZDV+ddI), Zidovudine and Zalcitabine (ZDV+ddC), and Didanosine (ddI); the randomization was stratified according to the length of prior antiretroviral therapy. For this analysis, we are interested in the effect of the true baseline CD4 count on time to an AIDS-defining event or death in participants with prior antiretroviral therapy. In general, an observed CD4 count is contaminated by instrumental error and biological diurnal fluctuation. The true baseline ln(CD4) is conceptually defined as the average underlying measure over a a short period of time at baseline. Note that the screening CD4 count is not regarded as a baseline one.

In this trial, duplicated mismeasured baseline CD4 counts are available. By taking advantage of them, Huang and Wang (2000) addressed a similar scientific question for antiretroviral-naive participants. However, for illustration purpose, here we included only one baseline CD4 count—taken closest to the randomization—in our primary set of the antiretroviral-sophisticated participants, and used duplicated baseline CD4 counts of the antiretroviral-naive participants as external reliability data. The sample size of the primary set is 1,395, with 349, 349, 346 and 351 participants randomized to ZDV, ZDV+ddI, ZDV+ddC, and ddI treatments, respectively. The median length of follow-up was 35 months and 216 events were observed. The variance of the observed baseline ln(CD4) was estimated to be 0.125. The size of the external reliability set is 1,036 and the estimated error variance is 0.033; see Huang and Wang (2000, Fig. 1) where the data were presented. From the two data sets, the standard deviation ratio of the error and the true baseline ln(CD4) in the primary set was estimated to be 0.60.

A Cox regression model was adopted with four covariates: the true baseline ln(CD4) and three treatment indicators, with the ZDV group treated as the reference. The analysis results from various methods are presented in Table 3. As expected, the naive coefficient estimate of the true baseline ln(CD4) is substantially smaller in absolute magnitude than others. On the other hand, the regression calibration, parametric-correction estimates are close to the nonparametric-correction ones, which are almost identical themselves. Similar phenomena were observed for antiretroviral-naive participants in Huang and Wang (2000). This might be due to infrequent events, moderate error contamination, or proximity to normal errors.

Table 3. Comparison of regression coefficient estimates in the ACTG 175 data.

| | ln(CD4) | | ZDV+ddI† | | ZDV+ddC† | | ddI† | |
|---|---|---|---|---|---|---|---|---|
| | estimate | s.e. | estimate | s.e. | estimate | s.e. | estimate | s.e. |
| NV | −1.506 | 0.161 | −0.501 | 0.197 | −0.108 | 0.181 | −0.337 | 0.190 |
| RC | −2.042 | 0.247 | −0.511 | 0.198 | −0.115 | 0.181 | −0.341 | 0.191 |
| PC | −1.861 | 0.206 | −0.542 | 0.202 | −0.132 | 0.184 | −0.338 | 0.192 |
| $\mathrm{NC_a}$ | −1.932 | 0.232 | −0.551 | 0.204 | −0.138 | 0.185 | −0.338 | 0.192 |
| $\mathrm{NC_b}$ | −1.933 | 0.233 | −0.552 | 0.204 | −0.138 | 0.185 | −0.338 | 0.192 |
| NC | −1.933 | 0.233 | −0.552 | 0.204 | −0.138 | 0.185 | −0.338 | 0.192 |

†Indicators with the ZDV group being the reference.

## 5. Discussion

It is worthwhile to point out that our new technique does not replace, but rather complements that of Huang and Wang (1999, 2000, 2001). While the former most noticeably extends nonparametric correction to the setup of external error assessment, only the latter is applicable to instrumental variable estimation (Huang and Wang (1999; 2001, Sec. 4.2)). In addition, an estimating function amenable to the latter may not be so to the former. A weighted partial-score function for the proportional hazards model (Lin and Wei (1989)),

$$\int_0^\tau \Big[\widehat{\mathcal{E}}\{I(T \geq t)\exp(\mathbf{b}^T\mathbf{X})\}\, d\widehat{\mathcal{E}}\{\mathbf{X}\Delta I(T \leq t)\} \\ - \widehat{\mathcal{E}}\{I(T \geq t)\mathbf{X}\exp(\mathbf{b}^T\mathbf{X})\}\, d\widehat{\mathcal{E}}\{\Delta I(T \leq t)\}\Big],$$

is such an example, where the errors-in-covariates effect is no longer additive in the limit. Together, the two techniques facilitate nonparametric correction for a wide spectrum of error-assessment data.

We have illustrated the proposed nonparametric correction technique with external error assessment. Although its applicability to internal error assessment is apparent, several issues are worth attention because in this case the distinction between primary and error-assessment data is not as clear. To focus, consider internal reliability data and a single error cluster.

(i) Given that each individual has $R(\geq 1)$ copies of $\mathbf{W}$, the naive estimating function is no longer unique. Taking Poisson regression as an example, we suggest using

$$\widehat{\mathcal{E}}\mathcal{A}[\{\mathbf{W}-\widehat{\mathcal{E}}\mathcal{A}(\mathbf{W})\}Y]-\widehat{\mathcal{E}}(Y)\left[\frac{\widehat{\mathcal{E}}\mathcal{A}\{\mathbf{W}\exp(\mathbf{b}^T\mathbf{W})\}}{\widehat{\mathcal{E}}\mathcal{A}\{\exp(\mathbf{b}^T\mathbf{W})\}}-\widehat{\mathcal{E}}\mathcal{A}(\mathbf{W})\right],$$

where $\mathcal{A}$ averages over the $R$ copies of $\mathbf{W}$. This naive estimating function makes use of all copies of $\mathbf{W}$ and has the same limit as the one using a single copy of $\mathbf{W}$. Then, the nonparametric correction may proceed in the same fashion as in the external error-assessment case, although (straightforward) adjustments are necessary for variance estimation and for efficient estimation. Note that $R$ may be random and can even depend on $\{\mathbf{X}, Y\}$. This can be important in practice since the inclusion in the error-assessment subset may depend on the true covariates and the response.

(ii) Suppose further that the reliability study is conducted on the whole primary set, i.e., $\Pr(R \geq 2) = 1$, which is the major setup considered in Huang and Wang (1999, 2000, 2001). Interestingly, using $\widehat{\boldsymbol{\varphi}}_{\mathrm{a}}(\mathbf{b})$ along with the earlier naive estimating function for Poisson regression results in a corrected estimating function equivalent to that given in Huang and Wang (1999). However, the corrected estimating functions here for logistic and Cox regressions do not reduce to those in Huang and Wang (2000, 2001). This shows that the two techniques have similarities as well as differences.

(iii) Now, suppose that $R$ is constant and is at least 2. An alternative naive estimating function would appear equally natural, with $\mathbf{X}$ replaced by the average $\overline{\mathbf{W}}$ of the $R$ copies of $\mathbf{W}$:

$$\widehat{\mathcal{E}}[\{\overline{\mathbf{W}}-\widehat{\mathcal{E}}(\overline{\mathbf{W}})\}Y]-\widehat{\mathcal{E}}(Y)\widehat{\boldsymbol{\phi}}_1(\mathbf{b};\overline{\mathbf{W}})$$

in the case of Poisson regression, which requires a correction of $R\boldsymbol{\phi}_1(\mathbf{b};R^{-1}\boldsymbol{\varepsilon}) = \boldsymbol{\phi}_1(R^{-1}\mathbf{b};\boldsymbol{\varepsilon})$. This points to room, and therefore needs effort, for further improving the estimation of $\boldsymbol{\beta}$ in efficiency. In this regard, the result of Huang and Wang (2001, Lemma 2) on synthesizing multiple estimating functions might be useful.

Our development in this article originates from natural estimating functions in the absence of measurement error, namely the score function for Poisson regression, the weighted score functions given in Huang and Wang (2001) for logistic regression, and the partial-score function for Cox regression. One motivation for using them is that the resulting nonparametric-correction estimators are fairly efficient, at least when the magnitude of measurement error is small. Nevertheless, other estimating functions amenable to the proposed nonparametric correction technique exist. For example, (3) for Poisson regression is equivalent to $\widehat{\boldsymbol{\phi}}_Y(\mathbf{0};\mathbf{X})-\widehat{\boldsymbol{\phi}}_1(\mathbf{b};\mathbf{X})$, which is a special member in the class,

$\widehat{\boldsymbol{\phi}}_Y(\boldsymbol{\gamma};\mathbf{X}) - \widehat{\boldsymbol{\phi}}_1(\mathbf{b}+\boldsymbol{\gamma};\mathbf{X})$ for any given value of $\boldsymbol{\gamma}$. We are currently exploring approaches to taking advantage of this result for more efficient estimation of $\boldsymbol{\beta}$, especially when the measurement error is large.

As a general concern, consistent functional modeling methods do not always have satisfactory small-sample performance in the presence of substantial error contamination. Besides nonparametric correction, conditional score (Stefanski and Carroll (1985)) and parametric correction (Nakamura (1990) and Stefanski (1989)) also have this issue; see Huang and Wang (2001) among others. Specifically for parametric and nonparametric correction methods, a corrected estimating function may not always have a zero-crossing and in addition the estimator may have large outliers. Unfortunately, large-sample results may be inadequate to address such small-sample issues. Refining these correction methods is an important, albeit challenging, research topic.

**Acknowledgement**

This research was partially supported by funds from the US National Institutes of Health. Most of the work was completed when the first author was with the Fred Hutchinson Cancer Research Center. The authors thank Professor Victor DeGruttola for providing the AIDS data analyzed in Section 4.3, and Professor Ross Prentice for reviewing an earlier draft of this article. They also gratefully acknowledge that helpful comments from the reviewers have led to improvement in the presentation.

## Appendix A. Properties of $\boldsymbol{\phi}_H(\mathbf{b};\mathbf{G})$

Function $\boldsymbol{\phi}_H(\mathbf{b};\mathbf{G})$ is defined in (9). Under regularity conditions, the following properties are derived through elementary algebra:

1. For $\boldsymbol{\delta} \perp \{\mathbf{G}, H\}$, $\boldsymbol{\phi}_H(\mathbf{b};\mathbf{G}+\boldsymbol{\delta}) = \boldsymbol{\phi}_H(\mathbf{b};\mathbf{G}) + \boldsymbol{\phi}_1(\mathbf{b};\boldsymbol{\delta})$.
2. For $\mathrm{var}(\mathbf{G}) = \mathbf{0}$, $\boldsymbol{\phi}_1(\mathbf{b};\mathbf{G}) = \mathbf{0}$.
3. For $\mathbf{G}_1 \perp \mathbf{G}_2$, $\boldsymbol{\phi}_1\{(\mathbf{b}_1^T,\mathbf{b}_2^T)^T;(\mathbf{G}_1^T,\mathbf{G}_2^T)^T\} = \{\boldsymbol{\phi}_1(\mathbf{b}_1;\mathbf{G}_1)^T, \boldsymbol{\phi}_1(\mathbf{b}_2;\mathbf{G}_2)^T\}^T$.
4. For $\mathbf{G} \perp \boldsymbol{\delta}_1 \perp \boldsymbol{\delta}_2$, $\boldsymbol{\phi}_1(\mathbf{b};\mathbf{G}) = \dfrac{\mathcal{E}[(\mathbf{G}+\boldsymbol{\delta}_1)\exp\{\mathbf{b}^T(\mathbf{G}+\boldsymbol{\delta}_2)\}]}{\mathcal{E}[\exp\{\mathbf{b}^T(\mathbf{G}+\boldsymbol{\delta}_2)\}]} - \mathcal{E}(\mathbf{G}+\boldsymbol{\delta}_1)$.

## Appendix B. Asymptotic Theory Under the Set-up of Section 4.1

Consider Cox regression with external error assessment, and suppose that there is a single error cluster. Specifically, the primary data consist of $n_p$, say, iid replicates of $\{\mathbf{W}, T, \Delta\}$ and the external error assessment data consist of $n_e$, say, iid replicates of $\{\mathbf{U}, \mathbf{V}\}$. We focus on the estimating function $\widehat{\boldsymbol{\Psi}}^{\mathrm{C}}(\mathbf{b})$ given in (14) with $\widehat{\boldsymbol{\varphi}}(\mathbf{b}) = \widehat{\boldsymbol{\varphi}}_{\mathrm{a}}(\mathbf{b})$. The approach to large-sample study here is similar to that in Huang and Wang (2000), and it can be used for the developments under other set-ups considered in this article. In the study, we let $n_p \to \infty$, $n_e \to \infty$ and $n_p/n_e$ converge to a finite non-zero constant.

We rewrite the estimating function $\widehat{\mathbf{\Psi}}^{\mathrm{C}}(\mathbf{b})$ to give it an explicit functional representation:

$$\widehat{\mathbf{\Psi}}^{\mathrm{C}}(\mathbf{b}) = \widehat{\mathcal{E}}_p\{\mathbf{W}\Delta I(T \le \tau)\} - \int_0^\tau \frac{\widehat{\mathcal{E}}_p\{I(T \ge t)\mathbf{W}\exp(\mathbf{b}^T\mathbf{W})\}}{\widehat{\mathcal{E}}_p\{I(T \ge t)\exp(\mathbf{b}^T\mathbf{W})\}}\, d\widehat{\mathcal{E}}_p\{\Delta I(T \le t)\}$$
$$+ \left[\frac{\widehat{\mathcal{E}}_e\{(\mathbf{U}-\mathbf{V})\exp(\mathbf{b}^T\mathbf{U})\}}{\widehat{\mathcal{E}}_e\{\exp(\mathbf{b}^T\mathbf{U})\}} - \widehat{\mathcal{E}}_e(\mathbf{U}-\mathbf{V})\right]\widehat{\mathcal{E}}_p\{\Delta I(T \le \tau)\},$$

where $\widehat{\mathcal{E}}_p$ and $\widehat{\mathcal{E}}_e$ have been used to differentiate empirical average in the primary and error-assessment samples. Now it is clear that $\widehat{\mathbf{\Psi}}^{\mathrm{C}}(\mathbf{b})$ is a functional of the following seven empirical processes: $\widehat{\mathcal{E}}_p\{\Delta I(T \le \cdot)$, $\widehat{\mathcal{E}}_p\{\mathbf{W}\Delta I(T \le \tau)\}$, $\widehat{\mathcal{E}}_p\{I(T \ge \cdot)\exp(\mathbf{b}^T\mathbf{W})\}$, $\widehat{\mathcal{E}}_p\{I(T \ge \cdot)\mathbf{W}\exp(\mathbf{b}^T\mathbf{W})\}$, $\widehat{\mathcal{E}}_e(\mathbf{U}-\mathbf{V})$, $\widehat{\mathcal{E}}_e\{\exp(\mathbf{b}^T\mathbf{U})\}$ and $\widehat{\mathcal{E}}_e\{(\mathbf{U}-\mathbf{V})\exp(\mathbf{b}^T\mathbf{U})\}$.

Suppose that the parameter space of concern $\mathcal{B}$ is a compact set around $\boldsymbol{\beta}$. We impose the following regularity conditions on moments:

$$\begin{aligned}
&\mathcal{E}(\mathbf{X}^T\mathbf{X}) < \infty, \mathcal{E}\{\sup_{\mathbf{b}\in\mathcal{B}} \mathbf{X}^T\mathbf{X}\exp(2\mathbf{b}^T\mathbf{X})\} < \infty,\\
&\mathcal{E}(\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}) < \infty, \quad \mathcal{E}\{\sup_{\mathbf{b}\in\mathcal{B}} \boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}\exp(2\mathbf{b}^T\boldsymbol{\varepsilon})\} < \infty,\\
&\mathcal{E}(\mathbf{Z}^T\mathbf{Z}) < \infty, \ \ \mathcal{E}\{\sup_{\mathbf{b}\in\mathcal{B}} \exp(2\mathbf{b}^T\mathbf{Z})\} < \infty,\\
&\mathcal{E}(\boldsymbol{\eta}^T\boldsymbol{\eta}) < \infty.
\end{aligned}$$

First, consider the consistency of $\widehat{\boldsymbol{\beta}}$. Given that $\tau$ satisfies $\Pr(T \ge \tau) > 0$, $\widehat{\mathcal{E}}_p\{I(T \ge t)\exp(\mathbf{b}^T\mathbf{W})\}$ is bounded away from 0 for $t \in [0,\tau]$ and $\mathbf{b} \in \mathcal{B}$. Among the aforementioned seven empirical processes, those quantities that do not involve $t$ and $\mathbf{b}$ converge almost surely to limits by the standard Strong Law of Large Numbers. For the rest, the extended Strong Law of Large Numbers given in Appendix III of Andersen and Gill (1982) asserts that they converge almost surely to limits as well, uniformly in $t \in [0,\tau]$ and $\mathbf{b} \in \mathcal{B}$. Since $\widehat{\mathbf{\Psi}}^{\mathrm{C}}(\mathbf{b})$, as a functional of these seven empirical processes, is continuous with respect to the supremum norm, it converges almost surely and uniformly in $\mathbf{b} \in \mathcal{B}$ to its limit

$$\mathcal{E}\{\mathbf{X}\Delta I(T \le \tau)\} - \int_0^\tau \frac{\mathcal{E}\{I(T \ge t)\mathbf{X}\exp(\mathbf{b}^T\mathbf{X})\}}{\mathcal{E}\{I(T \ge t)\exp(\mathbf{b}^T\mathbf{X})\}}\, d\mathcal{E}\{\Delta I(T \le t)\},$$

which is monotone with a unique zero-crossing at $\boldsymbol{\beta}$. Then, it follows that $\widehat{\boldsymbol{\beta}}$ exists and converges to $\boldsymbol{\beta}$ almost surely.

Now, we show the asymptotic normality of $\widehat{\boldsymbol{\Psi}}^{\mathrm{C}}(\boldsymbol{\beta})$. Using Lemma 3 of Gill (1989) and the chain rule, one can establish that the map from the seven empirical processes to $\widehat{\boldsymbol{\Psi}}^{\mathrm{C}}(\boldsymbol{\beta})$ is compactly differentiable with respect to the supremum norm. Furthermore, the standard Central Limit Theorem and Lemma 5.1 of Tsiatis (1981) assert the asymptotic normality of these seven empirical processes under the moment conditions. Thus, by the functional delta method, we have the asymptotic normality of $\widehat{\boldsymbol{\Psi}}^{\mathrm{C}}(\boldsymbol{\beta})$.

Next, we consider the asymptotic linearity of $\widehat{\boldsymbol{\Psi}}^{\mathrm{C}}(\mathbf{b})$ around $\boldsymbol{\beta}$. Given the consistency of $\widehat{\boldsymbol{\beta}}$, standard Taylor expansion arguments reveal that it is sufficient to show that the $d\widehat{\boldsymbol{\Psi}}^{\mathrm{C}}(\mathbf{b})/d\mathbf{b}^T$ converges uniformly in $\mathbf{b} \in \mathcal{B}$. This can be achieved using the same approach as we used to establish the consistency of $\widehat{\boldsymbol{\Psi}}^{\mathrm{C}}(\mathbf{b})$.

Finally, the asymptotic normality of $\widehat{\boldsymbol{\beta}}$ is established given that of $\widehat{\boldsymbol{\Psi}}^{\mathrm{C}}(\boldsymbol{\beta})$ and the asymptotic linearity of $\widehat{\boldsymbol{\Psi}}^{\mathrm{C}}(\mathbf{b})$ around $\boldsymbol{\beta}$.

## Appendix C. Variance Estimation in Section 4.1

The influence function approach is used for variance estimation of $\widehat{\boldsymbol{\Psi}}^{\mathrm{C}}(\boldsymbol{\beta})$ (cf., Huber (1981)). Given that $\widehat{\boldsymbol{\Psi}}^{\mathrm{C}}(\boldsymbol{\beta})$ is a functional of empirical processes, the functional delta method can be used to show that, asymptotically, it is a sum of independent mean-zero contributions from individuals in the primary set and individuals in the error-assessment sets.

### Primary set

Denote the sample size by $n$ and the data from the $i$th individual by $\{\mathbf{W}_i, T_i, \Delta_i\}$, $i = 1, \ldots, n$. The contribution of the $i$th individual is estimated by

$$\begin{aligned}\boldsymbol{\xi}_i(\boldsymbol{\beta}) = n^{-1} \int_0^\tau & \left\{\mathbf{W}_i - \widehat{\mathcal{E}}(\mathbf{W}) - \widehat{\boldsymbol{\phi}}_{I(T \geq t)}(\boldsymbol{\beta}; \mathbf{W})\right\} \\ & \times \left[d\{\Delta_i I(T_i \leq t)\} - \frac{\exp(\boldsymbol{\beta}^T \mathbf{W}_i) I(T_i \geq t)}{\widehat{\mathcal{E}}\{\exp(\boldsymbol{\beta}^T \mathbf{W}) I(T \geq t)\}}\, d\widehat{\mathcal{E}}\{\Delta I(T \leq t)\}\right] \\ & + n^{-1} \widehat{\boldsymbol{\varphi}}(\boldsymbol{\beta}) \Delta_i I(T_i \leq \tau).\end{aligned}$$

### External error assessment

According to the error clustering, write $\boldsymbol{\beta} \equiv \{\boldsymbol{\beta}_0^T, \boldsymbol{\beta}_1^T, \ldots, \boldsymbol{\beta}_K^T\}^T$. Specific to the $k$th cluster, $k = 1, \ldots, K$, denote the size of the error-assessment data set by $m_k$ and the replicate of $\{\mathbf{U}, \mathbf{V}\}$ from the $l$th individual by $\{\mathbf{U}_l, \mathbf{V}_l\}$, $l = 1, \ldots, m_k$. By the functional delta method, the contributions of the $l$th individual to $\widehat{\boldsymbol{\varphi}}_{\mathrm{a}}(\boldsymbol{\beta}_k)$

and $\widehat{\boldsymbol{\varphi}}_{\mathrm{b}}(\boldsymbol{\beta}_k)$, respectively, are estimated by

$$\boldsymbol{\zeta}_{\mathrm{a}l}(\boldsymbol{\beta}_k) = m_k^{-1}\mathcal{A}\Bigg\{\Big((\mathbf{U}_l - \mathbf{V}_l) - \frac{\widehat{\mathcal{E}}\mathcal{A}\{(\mathbf{U}-\mathbf{V})\exp(\boldsymbol{\beta}_k^T\mathbf{U})\}}{\widehat{\mathcal{E}}\mathcal{A}\{\exp(\boldsymbol{\beta}_k^T\mathbf{U})\}}\Big)\frac{\exp(\boldsymbol{\beta}_k^T\mathbf{U}_l)}{\widehat{\mathcal{E}}\mathcal{A}\{\exp(\boldsymbol{\beta}_k^T\mathbf{U})\}} - \{(\mathbf{U}_l - \mathbf{V}_l) - \widehat{\mathcal{E}}\mathcal{A}(\mathbf{U}-\mathbf{V})\}\Bigg\},$$

$$\boldsymbol{\zeta}_{\mathrm{b}l}(\boldsymbol{\beta}_k) = m_k^{-1}\mathcal{A}\Bigg\{\Big(\mathbf{U}_l - \frac{\widehat{\mathcal{E}}\mathcal{A}[\mathbf{U}\exp\{\boldsymbol{\beta}_k^T(\mathbf{U}-\mathbf{V})\}]}{\widehat{\mathcal{E}}\mathcal{A}[\exp\{\boldsymbol{\beta}_k^T(\mathbf{U}-\mathbf{V})\}]}\Big)\frac{\exp\{\boldsymbol{\beta}_k^T(\mathbf{U}_l-\mathbf{V}_l)\}}{\widehat{\mathcal{E}}\mathcal{A}[\exp\{\boldsymbol{\beta}_k^T(\mathbf{U}-\mathbf{V})\}]} - \{\mathbf{U}_l - \widehat{\mathcal{E}}\mathcal{A}(\mathbf{U})\}\Bigg\}.$$

Recall that the error-assessment data sets may not be external to each other. In other words, one individual may contribute to multiple data sets. Thus, the number of individuals contributing to the $K$ error-assessment data sets, $m$, is no greater than $\sum_{k=1}^K m_k$. Accordingly, we assemble the estimated contributions of the $j$th individual, $j = 1,\ldots,m$, to $\widehat{\boldsymbol{\varphi}}_{\mathrm{a}}(\boldsymbol{\beta})$ and $\widehat{\boldsymbol{\varphi}}_{\mathrm{b}}(\boldsymbol{\beta})$, and obtain $\boldsymbol{\zeta}_{\mathrm{a}j}(\boldsymbol{\beta})$ and $\boldsymbol{\zeta}_{\mathrm{b}j}(\boldsymbol{\beta})$, respectively.

Now, to estimate $\boldsymbol{\Sigma}$, use

$$\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\beta}) = \sum_{j=1}^m \begin{pmatrix}\boldsymbol{\zeta}_{\mathrm{a}j}(\boldsymbol{\beta})\\ \boldsymbol{\zeta}_{\mathrm{b}j}(\boldsymbol{\beta})\end{pmatrix}^2,$$

where $\mathbf{v}^2 \equiv \mathbf{v}\mathbf{v}^T$ for vector $\mathbf{v}$. Since $\boldsymbol{\beta}$ is unknown, we plug in a consistent estimator, for example, $\mathrm{NC_a}$ or $\mathrm{NC_b}$, and the resulting estimate remains consistent.

For a weighted average $\widehat{\boldsymbol{\varphi}}(\boldsymbol{\beta})$ of $\widehat{\boldsymbol{\varphi}}_{\mathrm{a}}(\boldsymbol{\beta})$ and $\widehat{\boldsymbol{\varphi}}_{\mathrm{b}}(\boldsymbol{\beta})$, the estimated contribution $\boldsymbol{\zeta}_j(\boldsymbol{\beta})$ of the $j$th individual is the corresponding weighted average of $\boldsymbol{\zeta}_{\mathrm{a}j}(\boldsymbol{\beta})$ and $\boldsymbol{\zeta}_{\mathrm{b}j}(\boldsymbol{\beta})$. Note that the weight can depend on the data, as in the case of optimal $\widehat{\boldsymbol{\varphi}}(\mathbf{b})$.

### Sandwich variance estimate for $\widehat{\boldsymbol{\beta}}$

The variance of $\widehat{\boldsymbol{\Psi}}^{\mathrm{C}}(\boldsymbol{\beta})$ is consistently estimated by $\boldsymbol{\Omega}(\boldsymbol{\beta}) = \sum_{i=1}^n \boldsymbol{\xi}_i(\boldsymbol{\beta})^2 + \sum_{j=1}^m [\boldsymbol{\zeta}_j(\boldsymbol{\beta})\widehat{\mathcal{E}}\{\Delta I(T \le \tau)\}]^2$. Write $\boldsymbol{\Gamma}(\boldsymbol{\beta}) = \Big\{d\widehat{\boldsymbol{\Psi}}^{\mathrm{C}}(\mathbf{b})/d\mathbf{b}^T\ \big|_{\mathbf{b}=\boldsymbol{\beta}}\Big\}^{-1}$. Then, the sandwich variance estimate of $\widehat{\boldsymbol{\beta}}$ is $\boldsymbol{\Gamma}(\boldsymbol{\beta})\boldsymbol{\Omega}(\boldsymbol{\beta})\boldsymbol{\Gamma}(\boldsymbol{\beta})^T$. Furthermore, one plugs in $\widehat{\boldsymbol{\beta}}$ to replace the unknown $\boldsymbol{\beta}$ and the variance estimate remains consistent.

## References

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.* **10**, 1100-1120.

Buzas, J. S. (1997). Instrumental variable estimation in nonlinear measurement error models. *Comm. Statist. Ser. A* **26**, 2861-2877.

Carroll, R. J., Küchenhoff, H., Lombard, F. and Stefanski, L. A. (1996). Asymptotics for the SIMEX estimator in nonlinear measurement error models. *J. Amer. Statist. Assoc.* **91**, 242-250.

Carroll, R. J., Ruppert, D. and Stefanski, L. A. (1995). *Measurement Error in Nonlinear Models*. Chapman and Hall, London.

Carroll, R. J. and Stefanski, L. A. (1990). Approximate quasilikelihood estimation in models with surrogate predictors. *J. Amer. Statist. Assoc.* **85**, 652-663.

Cook, J. and Stefanski, L. A. (1994). A simulation extrapolation method for parametric measurement error models. *J. Amer. Statist. Assoc.* **89**, 1314-1328.

Cox, D. R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. Ser. B* **34**, 187-220.

Gill, R. D. (1989). Non- and semi-parametric maximum likelihood estimators and the von Mises method – I. *Scand. J. Statist.* **16**, 97-128.

Gleser, L. J. (1990). Improvements of the naive approach to estimation in nonlinear errors-in-variables regression models. In *Statistical Analysis of Measurement Errors Models and Application* (Edited by P. J. Brown and W. A. Fuller). American Mathematics Society, Providence.

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. *N. Engl. J. Med.* **335**, 1081-1090.

Huang, Y. and Wang, C. Y. (1999). Nonparametric correction to errors in covariates. Technical Report, Fred Hutchinson Cancer Research Center, Seattle.

Huang, Y. and Wang, C. Y. (2000). Cox regression with accurate covariates unascertainable: A nonparametric-correction approach. *J. Amer. Statist. Assoc.* **95**, 1209-1219.

Huang, Y. and Wang, C. Y. (2001). Consistent functional methods for logistic regression with errors in covariates. *J. Amer. Statist. Assoc.* **96**, 1469-1482.

Huber, P. J. (1981) *Robust Statistics*. Wiley, New York.

Lin, D. Y. and Wei, L. J. (1989). The robust inference for the Cox proportional hazards model. *J. Amer. Statist. Assoc.* **84**, 1074-1078.

Nakamura, T. (1990). Corrected score function for errors in variables models: Methodology and application to generalized linear models. *Biometrika* **77**, 127-137.

Nakamura, T. (1992). Proportional hazards model with covariates subject to measurement error. *Biometrics* **48**, 829-838.

Prentice, R. L. (1982). Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika* **69**, 331-342.

Stefanski, L. A. (1989). Unbiased estimation of a nonlinear function of a normal mean with application to measurement error models. *Comm. Statist. Ser. A* **18**, 4335-4358.

Stefanski and Carroll (1985). Conditional scores and optimal scores for generalized linear measurement-error models. *Biometrika* **74**, 703-716.

Stefanski, L. A. and Cook, J. R. (1995). Simulation-extrapolation: The measurement error jackknife. *J. Amer. Statist. Assoc.* **90**, 1247-1256.

Tsiatis, A. A. (1981). A large sample study of Cox's regression model. *Ann. Statist.* **9**, 93-108.

Department of Biostatistics, Rollins School of Public Health, Emory University, 1518 Clifton Rd. NE, Atlanta, GA 30322, U.S.A.

E-mail: yhuang5@emory.edu

Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave. N, P.O. Box 19024, Seattle, WA 98109-1024, U.S.A.

E-mail: cywang@fhcrc.org

(Received July 2003; accepted June 2005)