Published in final edited form as:

Biometrics. 2018 March ; 74(1): 118–126. doi:10.1111/biom.12741.

Cox Regression with Dependent Error in Covariates

Yijian Huang and

Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, Georgia 30322, U.S.A

Ching-Yun Wang

Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, U.S.A

Summary

Many survival studies have error-contaminated covariates due to the lack of a gold standard of measurement. Furthermore, the error distribution can depend on the true covariates but the structure may be difficult to characterize; heteroscedasticity is a common manifestation. We suggest a novel dependent measurement error model with minimal assumptions on the dependence structure, and propose a new functional modeling method for Cox regression when an instrumental variable is available. This proposal accommodates much more general error contamination than existing approaches including nonparametric correction methods of Huang and Wang (2000, *Journal of the American Statistical Association* **95**, 1209–1219; 2006, *Statistica Sinica* **16**, 861–881). The estimated regression coefficients are consistent and asymptotically normal, and a consistent variance estimate is provided for inference. Simulations demonstrate that the procedure performs well even under substantial error contamination. Illustration with a clinical study is provided.

Keywords

Functional modeling; Heteroscedastic error; Instrumental variable; Multiplicative error; Nonparametric correction; Proportional hazards model

1. Introduction

In many survival studies, some covariates may be contaminated with error due to the lack of a gold standard of measurement, e.g., CD4 count and viral load in HIV/AIDS research and nutritional intakes in cancer epidemiology. Just as with regression analyses in general (Carroll et al., 2006), accounting for the error is imperative in Cox regression since substantial bias may otherwise arise (Prentice, 1982; Hughes, 1993; Li and Ryan, 2004). Write survival time as S and censoring time as C . As a result of censoring, they are observed only through follow-up time $T = \min(S, C)$ and censoring indicator $\delta = I(S \leq C)$, where $I(\cdot)$

7. Supplementary Material

The Web Appendix referenced in Sections 1 and 6 and an R package implementing our proposals are available with this article at the *Biometrics* website on Wiley Online Library.

is the indicator function. To focus on main ideas, we shall confine our attention to time-independent covariates $\mathbf{X}^* \equiv (X, \mathbf{Z}^\top)^\top$ with scalar X being error-prone and the rest $\mathbf{Z}$ accurately measured. The proportional hazards model (Cox, 1972) postulates

$$d\Lambda(t|\mathbf{X}^*) = \exp(\boldsymbol{\beta}^\top \mathbf{X}^*) d\Lambda_0(t), \quad S \perp\!\!\!\perp C|\mathbf{X}^*, \quad (1)$$

where $\Lambda(\cdot|\mathbf{X}^*)$ is the cumulative hazard function of S given $\mathbf{X}^*$, $\Lambda_0(\cdot)$ is an unspecified baseline cumulative hazard function, $\boldsymbol{\beta}$ is an unknown regression coefficient, and $\perp\!\!\!\perp$ denotes statistical independence. While X is not directly observable, its error-contaminated version W instead is observed, as well as so-called replication data or instrumental data (cf. Carroll et al., 2006, Section 2.3) but not validation data due to the absence of gold standard.

A number of estimation methods have been developed. Regression calibration (Prentice, 1982; Wang et al., 1997; Xie et al., 2001) is widely adopted to produce approximate estimation. Another approximate but inconsistent method is the first-order bias-correcting estimator of Li and Ryan (2004). When the measurement error distribution is known or consistently estimated, Hu et al. (1998) proposed likelihood-based approaches under, however, restrictive assumptions on the censoring mechanism and even the distribution of $\mathbf{X}^*$. Functional modeling methods including corrected score (Nakamura, 1992) and conditional score (Tsiatis and Davidian, 2001) avoid these assumptions and are generally regarded as more robust. Huang and Wang (2000) proposed a nonparametric correction method that further does away with distributional assumptions on the measurement error, when replicated mismeasured covariates are available. Hu and Lin (2004) relaxed the data requirement to an internal reliability subset, but at the price of symmetric error distribution. Huang and Wang (2006) later developed new nonparametric correction methods to accommodate a broad variety of internal and external error-assessment data. Most recently, Song and Wang (2014) adapted and extended the nonparametric correction of Huang and Wang (2000) to the estimation with an instrumental variable. Note that, throughout, W is considered as continuously distributed; see the Web Appendix for related discussion.

In spite of their generalities, those nonparametric correction methods nonetheless have a limitation to require independence between the measurement error and the true covariates. This is not always realistic in practice, as often manifested in terms of error heteroscedasticity. To further complicate the matter, the error dependence on the true covariates is typically difficult to characterize. In this article, we develop a novel consistent functional modeling method to tackle such dependent error, in the circumstance that an instrumental variable U is available. We propose a general model for dependent measurement error and instrumental variable in Section 2. This new dependent measurement error model relaxes the classical measurement error model, and imposes minimal assumptions on the dependence structure. Meanwhile, our assumption on the instrumental variable is also much weaker than that adopted in the literature. In Section 3, we propose a consistent estimation procedure via estimating functions. Simulations are reported in Section 4, and an illustration given in Section 5. Final remarks are provided in Section 6. Technical details are deferred to the Appendix.

2. General model for dependent measurement error and instrumental variable

Write equality in distribution as $\sim$. We propose a dependent measurement error model:

$$W = X + \eta + \xi, \quad \eta \perp \{T, \Delta, \mathbf{X}^\circ, \xi\}, \quad \xi \perp \{T, \Delta\} | \mathbf{X}^\circ, \quad \xi \sim -\xi | \mathbf{X}^\circ, \quad (2)$$

where the measurement error $\varepsilon \equiv \eta + \xi$ has two independent components, homogeneous η and heterogeneous ξ . The homogeneous component η is completely independent and free of any distributional assumption; its expectation may not even be 0. On the other hand, the heterogeneous component ξ is allowed to depend on the true covariates, particularly X . However, certain distributional assumptions on ξ are necessary for the identifiability of β in the proportional hazards model (1); as a trivial example, the non-identifiability arises when ξ is a linear but unknown function of $\mathbf{X}^\circ$. We impose zero-symmetry on ξ , which is a reasonable and commonly adopted assumption for measurement error (e.g., Hu and Lin, 2004). Nevertheless, the measurement error ε as a whole is not required to be symmetric in distribution. When ξ vanishes and η has 0 mean, the above error model reduces to the classical measurement error model.

Remark 1

Throughout we take the measurement error as additive. However, some special cases under the measurement error model (2) may have alternative formulations, e.g.,

$$W = X\zeta + \eta, \quad \eta \perp \{T, \Delta, \mathbf{X}^\circ, \zeta\}, \quad \zeta \perp \{T, \Delta\} | \mathbf{X}^\circ, \quad \zeta - 1 \sim -(\zeta - 1) | \mathbf{X}^\circ,$$

with non-negative multiplicative error ζ and additive error η . This error contamination can be regarded as purely multiplicative when $\eta = 0$.

In our set-up, instrumental variable U is observed satisfying the following condition:

$$U \perp \{T, \Delta, W\} | \mathbf{X}^\circ. \quad (3)$$

As compared with W , U also contains information about X but requires less structure. A replicate of W , as considered in Huang and Wang (2000), is a special case of U . But many other variables correlated with X may serve as well, so that our development is more generally applicable. The illustration given in Section 5 will show such an example.

Remark 2

Under the dependent measurement error model (2), the instrumental variable condition (3) allows for dependence of the instrumental variable U on the measurement error ε , specifically the heterogeneous component ξ . This relaxes a typical assumption in the

literature that an instrumental variable is uncorrelated with the measurement error (cf. Carroll et al., 2006, Section 6.1).

While formulated in the context of the proportional hazards model as the primary model, our model for dependent measurement error and instrumental variable can clearly be extended for other primary models including the generalized linear models.

3. Consistent estimation

The data consist of n independent replicates of $\{T_i, W_i, U_i, \mathbf{Z}_i\}$: $\{T_i, W_i, U_i, \mathbf{Z}_i\}$, $i = 1, \dots, n$. Write counting process $N(t) = \mathbb{I}(T \leq t)$ and at-risk process $Y(t) = \mathbb{I}(T > t)$. According to counting-process martingale theory, $N(t) - \int_0^t Y(s) \exp(\beta^\top \mathbf{X}^\circ) d\Lambda_0(s)$ is a martingale with respect to filtration $\mathcal{F}_t = \sigma\{N(s), Y(s+): 0 \leq s \leq t; \mathbf{X}^\circ; U\}$, under the proportional hazards model (1) and the instrumental variable condition (3). Thus,

$$\left\{ \exp(-\beta_1 X/2) \left\{ N(t) - \int_0^t Y(s) \exp(\beta^\top \mathbf{X}^\circ) d\Lambda_0(s) \right\} \mid \mathbf{X}^\circ, U \right\} = 0 \quad (t \geq 0), \quad (4)$$

where $\mathbb{E}$ denotes expectation. Then, under the measurement error model (2),

$$\left\{ \exp(-\beta_1 W/2) N(t) - \int_0^t Y(s) \exp(\beta_1 W/2 + \beta_2^\top \mathbf{Z}) d\Omega_0(s) \mid \mathbf{X}^\circ, U \right\} = 0 \quad (t \geq 0), \quad (5)$$

where $(\beta_1, \beta_2^\top)^\top \equiv \beta$ and $\Omega_0(t) = \Lambda_0(t) \mathbb{E}\{\exp(-\beta_1 \eta/2)\} / \mathbb{E}\{\exp(\beta_1 \eta/2)\}$. When X in (4) is replaced with W as in (5), heterogeneous error component ξ does not affect the equality due to its symmetry in distribution whereas the effect of homogeneous component η is counteracted through $\Omega_0(\cdot)$. In parallel with notation $\mathbf{X}^\circ \equiv (X, \mathbf{Z}^\top)^\top$, write $\mathbf{U}^\circ \equiv (U, \mathbf{Z}^\top)^\top$. The preceding identity then suggests a class of estimating equations for $\{\beta, \Omega_0(\cdot)\}$, indexed by function $\phi(\cdot)$:

$$\begin{aligned} \mathbb{E}_n \left[\mathbf{U}^\circ \exp\{b_1 \phi(\mathbf{U}^\circ)/2\} \left\{ \exp(-b_1 W/2) \Delta - \exp(b_1 W/2 + \mathbf{b}_2^\top \mathbf{Z}) \Omega(T) \right\} \right] &= \mathbf{0}, \\ \mathbb{E}_n \left[\exp\{b_1 \phi(\mathbf{U}^\circ)/2\} \left\{ \exp(-b_1 W/2) N(t) - \int_0^t Y(s) \exp(b_1 W/2 + \mathbf{b}_2^\top \mathbf{Z}) d\Omega(s) \right\} \right] &= 0 \quad (t \geq 0), \end{aligned}$$

where $\mathbb{E}_n$ denotes sample average, e.g., $\mathbb{E}_n(U) = n^{-1} \sum_{i=1}^n U_i$. With b_1 and $\mathbf{b}_2$ fixed, the second equation admits a unique solution for $\Omega(\cdot)$, which is a step function with jumps only at observed failure times. Plugging this solution into the first equation then yields a profile estimating equation for β ,

$$\begin{aligned} & \mathbb{E}_n \int_0^\infty \exp[-b_1 \{W - \phi(\mathbf{U}^\circ)\}/2] \\ & \times \left[\mathbf{U}^\circ - \frac{\mathbb{E}_n \{Y(t) \mathbf{U}^\circ \exp[b_1 \{W + \phi(\mathbf{U}^\circ)\}/2 + \mathbf{b}_2^\top \mathbf{Z}]\}}{\mathbb{E}_n \{Y(t) \exp[b_1 \{W + \phi(\mathbf{U}^\circ)\}/2 + \mathbf{b}_2^\top \mathbf{Z}]\}} \right] dN(t) = \mathbf{0}. \end{aligned} \quad (6)$$

Remark 3

In the special case of $\xi = 0$, the estimating function in (6) may be regarded as a corrected version from the following counterpart in the absence of measurement error,

$$\mathbb{E}_n \int_0^\infty \exp[-b_1 \{X - \phi(\mathbf{U}^\circ)\} / 2] \left[\mathbf{U}^\circ - \frac{\mathbb{E}_n \{Y(t) \mathbf{U}^\circ \exp [b_1 \{X + \phi(\mathbf{U}^\circ)\} / 2 + \mathbf{b}_2^\top \mathbf{Z}]\}}{\mathbb{E}_n \{Y(t) \exp [b_1 \{X + \phi(\mathbf{U}^\circ)\} / 2 + \mathbf{b}_2^\top \mathbf{Z}]\}} \right] dN(t).$$

These two functions differ by a factor of $\mathbb{E}\{\exp(-b_1 \eta/2)\}$ in limit and thus have the same limiting solutions. However, in the presence of the heterogeneous error component, these two estimating functions no longer have a simple relationship. Thus, our proposal is different from typical correction method.

Remark 4

In the special case of $W = U = X$, the estimating function in (6) does not necessarily reduce to any one commonly adopted in the absence of measurement error such as the weighted partial score (e.g., Lin, 1991), except when $\phi(\mathbf{U}^\circ) = X + \text{constant}$. This also distinguishes our proposal from corrected score (Nakamura, 1992), conditional score (Tsiatis and Davidian, 2001), and nonparametric correction (Huang and Wang, 2000, 2006; Song and Wang, 2014) among others.

Remark 5

The proportional hazards model (1) is parameterized by not only β but also $\Lambda_0(\cdot)$. However, $\Lambda_0(\cdot)$ is not fully identifiable but only up to $\Omega_0(\cdot)$. This identifiability issue can be resolved with additional assumption on η , e.g., zero-symmetry. In this case, $\Lambda_0(\cdot) = \Omega_0(\cdot)$ and one may follow Huang and Wang (2000, Section 2.3) to develop consistent estimation for $\Lambda_0(\cdot)$.

Theorem 1—*Suppose that the proportional hazards model (1), the measurement error model (2), the instrumental variable condition (3), and the regularity conditions in the Appendix hold. Then, with probability tending to 1, there exists a solution $\hat{\beta}$ to estimating equation (6) that is consistent for β . Furthermore, $n^{1/2}(\hat{\beta} - \beta)$ is asymptotically normal with mean $\mathbf{0}$ and a covariance matrix that can be consistently estimated by a sandwich variance estimate given in the Appendix.*

Function $\phi(\cdot)$ in the estimating equation (6) has so far been considered as fixed. A simple choice takes

$$\text{Proposal 1: } \phi(\cdot) = 0, \quad (7)$$

as an example. However, such a formulation may be relaxed to allow for data adaptation. Parameterize $\phi(\cdot)$ with a finite-dimensional index ρ , i.e., $\phi(\cdot) \equiv \phi(\cdot; \rho)$, and then adopt data-adaptive $\phi(\cdot; \hat{\rho})$ in (6) with $\hat{\rho}$ estimated from the data. For instance,

Proposal 2: $\phi(\cdot; \hat{\rho}) = \text{fitted linear function of regressing } W \text{ over } U^\circ$, (8)

where $\hat{\rho}$ is the least-squares estimate. As such, the estimating function in (6) reduces to the partial score when $W = U = X$. Thus, adopting (8) may result in good estimation efficiency at least when the measurement error is small and the instrumental variable is highly correlated with X . The following result establishes that a data-adaptive $\phi(\cdot; \hat{\rho})$ can be treated the same as fixed $\phi(\cdot; \rho)$ under mild conditions, which clearly hold for choice (8).

Theorem 2—*Adopt the conditions in Theorem 1, and further assume that (i) $\hat{\rho} - \rho = o_p(n^{-1/4})$ and (ii) the second derivative of $\phi(u^\circ; \rho)$ with respect to ρ exists and is bounded for ρ in a neighborhood of ρ and $u^\circ \in \text{supp}(U^\circ)$, where $\text{supp}(\cdot)$ denotes support. Then, with probability tending to 1, the estimating equation (6) using $\phi(\cdot; \hat{\rho})$ has a consistent solution, which is first-order asymptotically equivalent to that of the estimating function using fixed $\phi(\cdot; \rho)$. Furthermore, the sandwich variance estimate in Theorem 1 remains consistent when $\phi(\cdot; \rho)$ is replaced with $\phi(\cdot; \hat{\rho})$.*

Theorems 1 and 2 claim the existence of a consistent solution. Of course, global consistency would be more desirable. This is a general issue with estimating functions involving instrumental variable (e.g., Huang and Wang, 2001, Section 4.2). Our problem is further complicated by the dependent measurement error; see Remark 3. Nevertheless, we have obtained a few encouraging results.

Theorem 3—*Consider estimating equation (6) with $\phi(\cdot) = 0$, and suppose that the conditions in Theorem 1 hold. Within any compact set containing β , all solutions are consistent for β and asymptotic equivalent to each other under either of the following scenarios:*

- i. $\mathbb{E}(U | X^\circ = x^\circ)$ is a linear function of x° with a non-zero coefficient for the leading element;
- ii. X is the only covariate and $\mathbb{E}(U | X = x)$ is a strictly monotone function of x .

Scenario (i) covers many realistic situations including the case that U is an independent replicate of W , as considered in Huang and Wang (2000). Nevertheless, the linearity condition may be further relaxed, at least in the single covariate case as given in scenario (ii).

Remark 6

While this result is specific to the case of $\phi(\cdot) = 0$, the corresponding estimator can then be used to identify a consistent solution to an estimating equation with another $\phi(\cdot)$. The solution closest to a consistent estimator is necessarily consistent.

4. Simulations

We have conducted extensive simulations to assess performance of the proposed estimators corresponding to (7) and (8). For comparison, the naive method by replacing X with W in

standard Cox regression and an adapted nonparametric correction method (Song and Wang, 2014, Section 3) were included in these studies. The estimating function for the latter is given by

$$\mathbb{E}_n \int_0^\infty \left[\mathbf{U}^\circ - \frac{\mathbb{E}_n \{Y(t) \mathbf{U}^\circ \exp(\mathbf{b}^\top \mathbf{W}^\circ)\}}{\mathbb{E}_n \{Y(t) \exp(\mathbf{b}^\top \mathbf{W}^\circ)\}} \right] dN(t) = \mathbf{0}, \quad (9)$$

where $\mathbf{W}^\circ \equiv (W, \mathbf{Z}^\top)^\top$. This estimator is consistent in the case of homogeneous measurement error, but may not be so generally under the dependent measurement error model (2). For most other existing covariate measurement error methods, including regression calibration, it is unclear how to adapt them to our data structure and thus they were not considered.

We employed Newton's method for root-finding of an estimating function, starting from zero and enforcing an improvement at each iteration by halving the Newton step repeatedly until a reduced Euclidean norm of the estimating function. A failure was declared when this algorithm did not result in a zero-crossing. To guard against potential inconsistent solutions for the proposed estimating functions, corresponding to (7) and (8), further root-finding was carried out to alternate between the two using each other's solution as the initial value until these solutions no longer changed. The premise is that consistent solutions of the two estimating functions are closer to each other than inconsistent ones, if any, and Newton's method leads to a solution close to the initial value. Nevertheless, solution updating was rarely needed in our numerical studies. To assess normality adequacy of a scalar estimator, we introduce a skewness measure based on evenness of the distribution above and below the median,

$$\frac{Q(0.975) - Q(0.5)}{Q(0.5) - Q(0.025)}, \quad (10)$$

where $Q(\cdot)$ denotes the quantile function of the estimator.

We first considered a model with a single and error-prone covariate. The baseline survival time followed the standard exponential distribution, the true covariate X had the standard normal distribution, and $\beta = 1$. The censoring time was covariate-dependent, taking a uniform distribution between 0 and 4.3 if X is negative or between 0 and 8.6 otherwise; the censoring rate was approximately 25%. Four scenarios of covariate error contamination were considered. The first two involved a heterogeneous component ξ that was normally distributed given X with the conditional variance proportional to $\Phi(X)$, making the measurement error heteroscedastic; $\Phi(\cdot)$ is the cumulative distribution function of the standard normal. However, the two scenarios differed in the homogeneous component η , one being a constant but the other taking the asymmetric Beta(2,1) distribution scaled to have the same variance as ξ . The last two scenarios had homogeneous error, with either normal or scaled Beta(2,1) distribution. Moderate and substantial error contamination, with error

standard deviation $\sigma_{\eta+\xi} = 1/2$ and 1, was investigated. Instrumental variable U followed the standard normal distribution, and jointly (X, U) was bivariate normal with a correlation coefficient of 0.8. The sample size was taken to be 300 or 600.

Table 1 reports the results from 1000 simulation replicates for each setting. Not surprisingly, the naive estimator sustained quite large bias, towards null, in all settings. The adapted nonparametric-correction estimator was biased under heteroscedastic error. With moderate error contamination, the bias was not outstanding but the Wald confidence interval had nonetheless noticeable under-coverage. However, the bias became prominent with substantial error contamination. Furthermore, in a few settings, under both heteroscedastic and homogeneous errors alike, the estimator was highly right-skewed with the skewness measure exceeding 2. In contrast, our proposals worked well in general. The root-finding failure, if any, was negligible, the bias was small, the standard error tracked standard deviation closely, and the 95% Wald confidence interval had coverage close to the nominal level. The distribution exhibited only moderate right skewness even with substantial measurement error. Under homogeneous errors, they were more efficient with normal error but less so with scaled Beta(2,1) error than the adapted nonparametric-correction estimator. Finally, between the two proposals, they were mostly comparable to each other but the one corresponding to (7) was a bit more efficient under the substantial heteroscedastic errors.

We also studied a model with two covariates and $\beta = (1, 1)^T$ under heteroscedastic error contamination with constant η and normally distributed ξ given X . The two covariates, X and Z , followed the standard bivariate normal with a correlation coefficient of either 0 or 0.5, Z and instrumental variable U are independent given X , and the design was otherwise the same as the first scenario of the preceding single-covariate model. The censoring rates were about 27% and 30% with X and Z being independent and correlated, respectively. The results as given in Table 2, from 1000 simulation replicates for each setting, show a pattern largely similar to those in Table 1 for the coefficient estimation of X . The bias in the estimated coefficient of Z was also evident in the naive estimation, although to a lesser degree.

5. Illustration with an AIDS study

We analyzed the ACTG 175 trial of the AIDS Clinical Trials Group, which evaluated treatments of one or two nucleosides in HIV-infected adults with a screening CD4 count between 200 and 500 and no history of AIDS-defining illness at baseline (Hammer et al., 1996). Among the 923 study participants with at least one year prior exposure to antiretroviral therapy, 900 had a baseline CD4 count measurement, other than the screening one, prior to the start of study treatment and within 5 days of randomization. These 900 patients, as included in the analysis data, were randomized to four treatment groups: 225 to zidovudine, 223 to zidovudine and didanosine, 220 to zidovudine and zalcitabine, and 232 to didanosine. This analysis focused on effect of true baseline CD4 count, among others, on time to AIDS or death in these antiretroviral-sophisticated patients. The median follow-up was 35 months and 158 endpoints were observed.

As well known, CD4 count has no gold standard of measurement. The true baseline $\log(\text{CD4})$ was defined as the technical error-free $\log(\text{CD4})$ averaged over a short period of time at baseline. The scatterplot of the first baseline versus screening CD4 counts, shown in Figure 1, clearly demonstrates the measurement error; the first baseline CD4 count was the one closest to the randomization in the case of multiple measurements available. To understand the measurement error in the baseline $\log(\text{CD4})$, we explored the subset of 277 patients who had a second measurement from a different blood sample. This gave an estimated variance ratio of 0.53 between the error and the true baseline $\log(\text{CD4})$. Furthermore, the sample variances of the difference between the two baseline $\log(\text{CD4})$ were 0.088, 0.089, and 0.068 for the sub-samples with screening CD4 in intervals [200, 300), [300, 400), and [400, 500], respectively. This suggested error heteroscedasticity in the baseline $\log(\text{CD4})$, albeit perhaps moderate in magnitude.

We considered a proportional hazards model with 4 covariates: the true baseline $\log(\text{CD4})$ and three indicators for the four treatments. Clearly, the screening $\log(\text{CD4})$ may not serve as W , since its distribution is truncated by the screening interval $[\log 200, \log 500]$ which complicates the associated measurement error. Nevertheless, conditionally on the true baseline $\log(\text{CD4})$, the measurement error in the screening $\log(\text{CD4})$ would still be independent of other variables. We therefore used the screening $\log(\text{CD4})$ as U , and the first baseline $\log(\text{CD4})$ as W . Such a screening measure is readily available in applications, and its adoption as instrumental variable is novel to our knowledge and also seems quite natural. The second baseline CD4 count measures as available in the sub-sample were not utilized in this illustration as such data are not always available in other applications. Table 3 shows the estimators based on the naive, adapted nonparametric correction, and two proposed methods. For the coefficient of $\log(\text{CD4})$, the naive and adapted nonparametric-correction estimators were both smaller in magnitude, particularly the former, than the proposed ones. Meanwhile, these approaches yielded reasonably similar coefficient estimators, in terms of difference relative to standard error, for the treatment indicators. These results largely echo the pattern observed in our simulations.

6. Concluding remarks

As compared to existing consistent functional modeling methods, this proposal further relaxes the error contamination mechanism to allow for largely unspecified dependence on the true covariates. This development has great practical significance since measurement error heteroscedasticity arises in many applications. Meanwhile, surprisingly the proposed estimation procedure also behaved better numerically than the adapted nonparametric correction method under substantial error contamination. Pathological finite-sample behavior is a common concern among functional modeling methods for nonlinear regression; see Huang (2014) among others. In view of equations (6) and (9), presumably such pathological behavior is largely due to the exponential terms as affected by the measurement error in W . However, the coefficient of W in equation (6) has an extra factor of $1/2$ or $-1/2$, effectively halving the measurement error magnitude and thus mitigating the adverse impact.

The proposal can be generalized for time-dependent covariates as well as multiple error-prone covariates; see the Web Appendix. Meanwhile, a few issues or problems warrant further investigations. First, it is not completely clear how to fully exploit the proposed class of estimating functions for improved estimation in terms of efficiency and numerical stability. This is a general issue with M -estimation. With a small number of estimating functions, one may utilize the generalized method of moments to combine them for a more efficient estimator asymptotically. Yet, the finite-sample performance may suffer. It becomes an even more difficult problem when the number of estimating functions is large. Second, it is worthwhile to look into the specialized setting that replicates of the mismeasured covariate are available. In this circumstance, any two replicates may serve as W and U in estimating equation (6). However, an approach, as used in the nonparametric correction of Huang and Wang (2000), to average over all possible such pairs for (W, U) within each empirical average $\mathbb{E}_n$ may lead to efficiency improvement in the estimation. Finally, in some applications, the instrumental variable may not be observed for every individual. The proposal might be applied to the sub-sample with complete data but could be quite inefficient. As a general strategy, one may augment the estimation by using methods for missing data (e.g., Wang, 2012).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Support by grants from the US National Institutes of Health HL113451 and AI050409 for Huang, and CA53996, GM100573, HL121347, HL130483, and MH105857 for Wang, is gratefully acknowledged. Part of this research was presented at the Banff International Research Station (BIRS) Workshop “Newest developments and urgent issues in measurement error and latent variable problems,” Banff, Canada in August 2016. The authors thank Professors Victor DeGruttola and Michael Hughes for permission to use the ACTG 175 trial data, and Professor Ross Prentice and the reviewers for their helpful comments which have led to an improved presentation.

References

- Carroll, RJ., Ruppert, D., Stefanski, LA., Crainiceanu, CM. Measurement Error in Nonlinear Models: A Modern Perspective. Boca Raton: Chapman and Hall; 2006.
- Cox DR. Regression models and life tables (with Discussion). Journal of the Royal Statistical Society B. 1972; 34:187–220.
- Gill RD. Non- and semi-parametric maximum likelihood estimators and the von Mises method — I. Scandinavian Journal of Statistics. 1989; 16:97–128.
- Hammer SM, Katzenstein DA, Hughes MD, Gundacker H, Schooley RT, Haubrich RH, et al. A trial comparing nucleoside monotherapy with combination therapy in HIV-infected adults with CD4 cell counts from 200 to 500 per cubic millimeter. The New England Journal of Medicine. 1996; 335:1081–1090. [PubMed: 8813038]
- Hu C, Lin DY. Semiparametric failure time regression with replicates of mismeasured covariates. Journal of the American Statistical Association. 2004; 99:105–118.
- Hu P, Tsiatis AA, Davidian M. Estimating the parameters in the Cox model when covariate variables are measured with error. Biometrics. 1998; 54:1407–1419. [PubMed: 9883541]
- Huang Y. Corrected score with sizable covariate measurement error: Pathology and remedy. Statistica Sinica. 2014; 24:357–374. [PubMed: 24489451]
- Huang Y, Wang CY. Cox regression with accurate covariates unascertainable: A nonparametric-correction approach. Journal of the American Statistical Association. 2000; 95:1209–1219.

- Huang Y, Wang CY. Consistent functional methods for logistic regression with errors in covariates. *Journal of the American Statistical Association*. 2001; 96:1469–1482.
- Huang Y, Wang CY. Errors-in-covariates effect on estimating functions: Additivity in limit and nonparametric correction. *Statistica Sinica*. 2006; 16:861–881.
- Hughes MD. Regression dilution in the proportional hazards model. *Biometrics*. 1993; 49:1056–1066. [PubMed: 8117900]
- Kosorok, MR. *Introduction to Empirical Processes and Semiparametric Inference*. New York: Springer; 2008.
- Li Y, Ryan L. Survival analysis with heterogeneous covariate measurement error. *Journal of the American Statistical Association*. 2004; 99:724–735.
- Lin DY. Goodness-of-fit analysis for the Cox regression model based on a class of parameter estimators. *Journal of the American Statistical Association*. 1991; 86:725–728.
- Nakamura T. Proportional hazards model with covariates subject to measurement error. *Biometrics*. 1992; 48:829–838. [PubMed: 1420844]
- Prentice RL. Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika*. 1982; 69:331–342.
- Song X, Wang CY. Proportional hazards model with covariate measurement error and instrumental variables. *Journal of the American Statistical Association*. 2014; 109:1636–1646. [PubMed: 25663724]
- Tsiatis AA, Davidian M. A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika*. 2001; 88:447–458.
- Wang CY. Robust best linear estimation for regression analysis using surrogate and instrumental variables. *Biostatistics*. 2012; 13:326–340. [PubMed: 22285992]
- Wang CY, Hsu L, Feng ZD, Prentice RL. Regression calibration in failure time regression. *Biometrics*. 1997; 53:131–145. [PubMed: 9147589]
- Xie SX, Wang CY, Prentice RL. A risk set calibration method for failure time regression by using a covariate reliability sample. *Journal of the Royal Statistical Society B*. 2001; 63:855–870.

Appendix

An asymptotic study with proofs of Theorems 1, 2, and 3

Write $\mathbf{b} \equiv (b_1, \mathbf{b}_2^\top)^\top$ and introduce four random quantities:

$$A(t, \mathbf{b}) = \exp[-b_1\{W - \phi(\mathbf{U}^\circ)\}/2]N(t), \quad \mathbf{B}(\mathbf{b}) = \mathbf{U}^\circ \exp[-b_1\{W - \phi(\mathbf{U}^\circ)\}/2]\Delta, \\ G(t, \mathbf{b}) = Y(t) \exp[b_1\{W + \phi(\mathbf{U}^\circ)\}/2 + \mathbf{b}_2^\top \mathbf{Z}], \quad \mathbf{H}(t, \mathbf{b}) = \mathbf{U}^\circ G(t, \mathbf{b}).$$

Denote the left-hand side of (6) by $\Psi(\mathbf{b})$, which can be represented as a functional of four empirical processes:

$$\Psi(\mathbf{b}) = \mathbb{E}_n\{\mathbf{B}(\mathbf{b})\} - \int_0^\infty \frac{\mathbb{E}_n\{\mathbf{H}(t, \mathbf{b})\}}{\mathbb{E}_n\{G(t, \mathbf{b})\}} d\mathbb{E}_n\{A(t, \mathbf{b})\}. \quad (\text{A.1})$$

We impose the following regularity conditions:

Condition 1: The upper support point of T , $\tau = \sup\{t: \text{pr}(T > t) > 0\}$, is finite.

Furthermore, $\text{pr}(S > \tau) > 0$ and $\text{pr}(C = \tau) > 0$.

Condition 2: Variables W , U , and $\mathbf{Z}$ are all bounded.

Condition 3: Function $\phi(\mathbf{u}^\circ)$ is bounded for $\mathbf{u}^\circ \in \text{supp}(\mathbf{U}^\circ)$.

Condition 4: The derivative of

$$\Psi_\infty(\mathbf{b}) = \{\mathbf{B}(\mathbf{b})\} - \int_0^\tau \frac{\{\mathbf{H}(t, \mathbf{b})\}}{\{G(t, \mathbf{b})\}} d\{A(t, \mathbf{b})\}$$

at $\boldsymbol{\beta}$, denoted by $\mathbf{D}$, is nonsingular.

Condition 1 is commonly adopted for Cox regression to avoid lengthy technical tail treatment.

A.1. Proof of Theorem 1

In light of the functional representation (A.1), we shall exploit empirical process theory; see Huang and Wang (2000) and Kosorok (2008, Section 4.2.1) for a similar approach in related contexts. Condition 1 effectively limits the time scale to finite interval $[0, \tau]$. Let $\mathcal{B}$ be an arbitrary compact neighborhood of $\boldsymbol{\beta}$. Under Conditions 2 and 3, the classes of functions, $\{A(t, \mathbf{b}) : t \in [0, \tau], \mathbf{b} \in \mathbb{E}\}$, $\{\mathbf{B}(\mathbf{b}) : \mathbf{b} \in \mathbb{E}\}$, $\{G(t, \mathbf{b}) : t \in [0, \tau], \mathbf{b} \in \mathbb{E}\}$, and $\{\mathbf{H}(t, \mathbf{b}) : t \in [0, \tau], \mathbf{b} \in \mathbb{E}\}$ are all Donsker; see, e.g., Kosorok (2008, Section 4.2.1).

Since every Donsker class is a Glivenko–Cantelli class, the empirical processes, $\mathbb{E}_n\{A(t, \mathbf{b})\}$, $\mathbb{E}_n\{\mathbf{B}(\mathbf{b})\}$, $\mathbb{E}_n\{G(t, \mathbf{b})\}$, and $\mathbb{E}_n\{\mathbf{H}(t, \mathbf{b})\}$, converge in probability to their respective expectations, uniformly in $t \in [0, \tau]$, if applicable, and $\mathbf{b} \in \mathbb{E}$. By Condition 1, $\mathbb{E}\{G(t, \mathbf{b})\}$ is bounded away from 0 uniformly in $t \in [0, \tau]$ and $\mathbf{b} \in \mathbb{E}$. Therefore, $\Psi(\mathbf{b})$ converges in probability to $\Psi_\infty(\mathbf{b})$, uniformly in $\mathbf{b} \in \mathbb{E}$. From identity (5), it can be shown $\Psi_\infty(\boldsymbol{\beta}) = 0$. By a coordinate-wise Taylor expansion, there exists $\mathbf{D}(\mathbf{b})$ such that

$$\Psi_\infty(\mathbf{b}) = \Psi_\infty(\boldsymbol{\beta}) + \mathbf{D}(\mathbf{b})(\mathbf{b} - \boldsymbol{\beta}) = \mathbf{D}(\mathbf{b})(\mathbf{b} - \boldsymbol{\beta})$$

and $\mathbf{D}(\mathbf{b})$ converges to $\mathbf{D}$ as $\mathbf{b}$ approaches $\boldsymbol{\beta}$. Note that $\mathbf{D}$ is nonsingular by Condition 4. Therefore, for $\mathbf{b}$ in a sufficiently small neighborhood of $\boldsymbol{\beta}$, $\mathbf{D}(\mathbf{b})$ is nonsingular and thus $\Psi_\infty(\mathbf{b})$ has a unique solution $\boldsymbol{\beta}$. It then follows that, with probability tending to 1, there exists a solution $\hat{\boldsymbol{\beta}}$ to estimating equation (6) that is consistent for $\boldsymbol{\beta}$.

By a coordinate-wise Taylor expansion,

$$0 = \Psi(\hat{\boldsymbol{\beta}}) = \Psi(\boldsymbol{\beta}) + \Psi'(\boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + o_p(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$$

since all the second-order partial derivatives of $\Psi(\mathbf{b})$ are bounded by Conditions 2 and 3. The techniques adopted previously can be used to show that $\Psi'(\boldsymbol{\beta})$ converges in probability to $\mathbf{D}$, which is nonsingular by Condition 4. Thus,

$$\hat{\beta} - \beta = -D^{-1} \Psi(\beta) \{1 + o_p(1)\}. \quad (A.2)$$

Since their associated classes of functions are Donsker, $\mathbb{E}_n\{A(\cdot, \beta)\}$, $\mathbb{E}_n\{B(\beta)\}$, $\mathbb{E}_n\{G(\cdot, \beta)\}$, and $\mathbb{E}_n\{H(\cdot, \beta)\}$ jointly are asymptotically normal. Using Gill (1989, Lemma 3) and the chain rule, one can show that $\Psi(\beta)$ is a compactly differentiable functional of $\mathbb{E}_n\{A(\cdot, \beta)\}$, $\mathbb{E}_n\{B(\beta)\}$, $\mathbb{E}_n\{G(\cdot, \beta)\}$, and $\mathbb{E}_n\{H(\cdot, \beta)\}$. Therefore, $\Psi(\beta)$ is asymptotically normal by the functional delta method. In light of the asymptotic linearity established earlier, the asymptotic normality of $\hat{\beta}$ follows from Slutsky's theorem.

The influence curve of $\Psi(\beta)$ can be estimated by

$$\omega(\beta) = B(\beta) - \int_0^\tau \frac{\mathbb{E}_n\{H(t, \beta)\}}{\mathbb{E}_n\{G(t, \beta)\}} dA(t, \beta) - \int_0^\tau \left\{ H(t, \beta) - \frac{\mathbb{E}_n\{H(t, \beta)\}}{\mathbb{E}_n\{G(t, \beta)\}} G(t, \beta) \right\} \frac{d\mathbb{E}_n\{A(t, \beta)\}}{\mathbb{E}_n\{G(t, \beta)\}}.$$

Then, the asymptotic variance of $n^{1/2}\Psi(\beta)$ is estimated by $\hat{V} = \mathbb{E}_n\{\omega(\hat{\beta})\omega(\hat{\beta})^\top\}$. On the other hand, denote the derivative of $\Psi(b)$ at $\hat{\beta}$ by $\hat{D}$. It can be shown that $\hat{V}$ is a consistent variance estimate and $\hat{D}$ is consistent for D . Therefore, the sandwich variance estimate $\hat{D}^{-1}\hat{V}(\hat{D}^{-1})^\top$ is consistent for the asymptotic variance of $n^{1/2}(\hat{\beta} - \beta)$.

A.2. Proof of Theorem 2

With data-adaptive $\phi(\cdot; \hat{\rho})$ in (6), write $\Psi(b; \hat{\rho})$ and $\tilde{\beta}$ to differentiate them from their counterparts $\Psi(b) \equiv \Psi(b; \rho)$ and $\hat{\beta}$ in the case of fixed $\phi(\cdot) \equiv \phi(\cdot; \rho)$. We shall argue that the perturbation due to $\hat{\rho}$ is negligible in the proof of Theorem 1. Under assumptions (i) and (ii), $\Psi(b; \hat{\rho}) - \Psi(b; \rho) = o_p(1)$ uniformly in $b \in \mathcal{B}$ by Taylor expansion. The consistency result on a solution to $\Psi(b; \hat{\rho})$ then follows. In parallel with (A.2), one obtains that

$$\tilde{\beta} - \beta = -D^{-1} \Psi(\beta; \hat{\rho}) \{1 + o_p(1)\}.$$

On the other hand, similar to the asymptotic normality proof of $\Psi(\beta)$, one can show that $\Psi(\beta; \hat{\rho}) / |_{\hat{\rho}=\rho} = O_p(n^{-1/2})$ under assumption (ii). Therefore, a Taylor expansion gives

$$\Psi(\beta; \hat{\rho}) = \Psi(\beta; \rho) + O_p(n^{-1/2})(\hat{\rho} - \rho) + O\{(\hat{\rho} - \rho)^2\} = \Psi(\beta; \rho) + o_p(n^{-1/2})$$

under assumptions (i) and (ii), since the second derivative of $\Psi(\beta; \hat{\rho})$ with respect to $\hat{\rho}$ is bounded. It then follows that $\tilde{\beta}$ and $\hat{\beta}$ are $o_p(n^{-1/2})$ apart and thus asymptotically equivalent. Finally, similar techniques can be used to show that the perturbation to the sandwich variance estimate is also negligible.

A.3. Proof of Theorem 3

It suffices to show $\Psi_{\infty}(\mathbf{b})$ admits a unique solution in any compact set containing $\boldsymbol{\beta}$. Under the proportional hazards model (1), the measurement error model (2), and the instrumental variable condition (3),

$$\begin{aligned}\Psi_{\infty}(\mathbf{b}) &= \{\exp(-b_1\eta/2)\} \times \left\{ \int_0^{\tau} v(\mathbf{X}^{\circ}, b_1) \left[\mathbf{U}^{\circ} - \frac{\{Y(t)\mathbf{U}^{\circ}v(\mathbf{X}^{\circ}, b_1)\exp(\mathbf{b}^{\top}\mathbf{X}^{\circ})\}}{\{Y(t)v(\mathbf{X}^{\circ}, b_1)\exp(\mathbf{b}^{\top}\mathbf{X}^{\circ})\}} \right] dN(t) \right\} \\ &= \{\exp(-b_1\eta/2)\} \int_0^{\tau} \frac{\psi(t, \mathbf{b}) d\Lambda_0(t)}{\{Y(t)v(\mathbf{X}^{\circ}, b_1)\exp(\mathbf{b}^{\top}\mathbf{X}^{\circ})\}},\end{aligned}$$

where $v(\mathbf{X}^{\circ}, b_1) = \exp(-b_1X/2)\mathbb{E}\{\exp(b_1\xi/2) | \mathbf{X}^{\circ}\}$ and

$$\begin{aligned}\psi(t, \mathbf{b}) &= \{Y(t)\mathbf{U}^{\circ}v(\mathbf{X}^{\circ}, b_1)\exp(\boldsymbol{\beta}^{\top}\mathbf{X}^{\circ})\} \{Y(t)v(\mathbf{X}^{\circ}, b_1)\exp(\mathbf{b}^{\top}\mathbf{X}^{\circ})\} \\ &\quad - \{Y(t)\mathbf{U}^{\circ}v(\mathbf{X}^{\circ}, b_1)\exp(\mathbf{b}^{\top}\mathbf{X}^{\circ})\} \{Y(t)v(\mathbf{X}^{\circ}, b_1)\exp(\boldsymbol{\beta}^{\top}\mathbf{X}^{\circ})\};\end{aligned}$$

equality (4) is exploited to obtain the second equality. Subsequently,

$$\begin{aligned}\psi(t, \mathbf{b}) &= [Y_1(t)Y_2(t)v(\mathbf{X}_1^{\circ}, b_1)v(\mathbf{X}_2^{\circ}, b_1)\mathbf{U}_1^{\circ}\{\exp(\boldsymbol{\beta}^{\top}\mathbf{X}_1^{\circ} + \mathbf{b}^{\top}\mathbf{X}_2^{\circ}) - \exp(\mathbf{b}^{\top}\mathbf{X}_1^{\circ} + \boldsymbol{\beta}^{\top}\mathbf{X}_2^{\circ})\}] \\ &= \frac{1}{2} [Y_1(t)Y_2(t)v(\mathbf{X}_1^{\circ}, b_1)v(\mathbf{X}_2^{\circ}, b_1)(\mathbf{U}_1^{\circ} - \mathbf{U}_2^{\circ})\{\exp(\boldsymbol{\beta}^{\top}\mathbf{X}_1^{\circ} + \mathbf{b}^{\top}\mathbf{X}_2^{\circ}) - \exp(\mathbf{b}^{\top}\mathbf{X}_1^{\circ} + \boldsymbol{\beta}^{\top}\mathbf{X}_2^{\circ})\}] \\ &= \frac{1}{2} \{Y_1(t)Y_2(t)v(\mathbf{X}_1^{\circ}, b_1)v(\mathbf{X}_2^{\circ}, b_1)\exp\{\mathbf{b}^{\top}(\mathbf{X}_1^{\circ} + \mathbf{X}_2^{\circ})\}[(\mathbf{U}_1^{\circ}|\mathbf{X}_1^{\circ}) - (\mathbf{U}_2^{\circ}|\mathbf{X}_2^{\circ})]\{\exp\{(\boldsymbol{\beta} - \mathbf{b})^{\top}\mathbf{X}_1^{\circ}\} - \exp\{(\boldsymbol{\beta} - \mathbf{b})^{\top}\mathbf{X}_2^{\circ}\}\}\},\end{aligned}$$

where $\{Y_i(\cdot), \mathbf{X}_i^{\circ}, \mathbf{U}_i^{\circ}\}$, $i = 1, 2$, are two independent replicates of $\{Y(\cdot), \mathbf{X}^{\circ}, \mathbf{U}^{\circ}\}$.

If $\mathbb{E}(U | \mathbf{X}^{\circ}) = X$, or equivalently $\mathbb{E}(\mathbf{U}^{\circ} | \mathbf{X}^{\circ}) = \mathbf{X}^{\circ}$, it is then clear that $(\boldsymbol{\beta} - \mathbf{b})^{\top} \boldsymbol{\psi}(t, \mathbf{b}) = 0$ and thus $(\boldsymbol{\beta} - \mathbf{b})^{\top} \Psi_{\infty}(\mathbf{b}) = 0$, where the equality holds only if $\mathbf{b} = \boldsymbol{\beta}$ for any finite $\mathbf{b}$. Therefore, $\Psi_{\infty}(\mathbf{b})$ under scenario (i) admits a unique solution since it has the same solution.

Under scenario (ii), without loss of generality, suppose that $\mathbb{E}(U/X = x)$ is strictly increasing in x . Then, almost surely $(\boldsymbol{\beta} - \mathbf{b})\{\mathbb{E}(U_1/X_1) - \mathbb{E}(U_2/X_2)\}[\exp\{(\boldsymbol{\beta} - \mathbf{b})X_1\} - \exp\{(\boldsymbol{\beta} - \mathbf{b})X_2\}] > 0$ unless $\mathbf{b} = \boldsymbol{\beta}$. Therefore, $\Psi_{\infty}(\mathbf{b})$ admits a unique solution.

Figure 1.
Screening and baseline CD4 count measurements from 900 patients in the ACTG

Table 1

Simulation summary statistics of the estimators in the single-covariate model.

size	F	B	D	E	C	S	F	B	D	E	C	S	
$\sigma_{\eta,\xi}=1/2$													
Heteroscedastic error: $\eta=0, \xi \mid X \sim \text{normal}$													
300	NV	0.0	-306	71	68	1.9	1.14	0.0	-599	53	50	0.0	1.10
	NC	0.1	-46	127	126	90.6	1.29	1.2	-87	1187	288	65.2	2.08
	P1	0.0	12	134	139	96.9	1.17	0.0	25	167	171	96.1	1.37
	P2	0.0	14	135	136	95.5	1.31	1.4	39	197	210	95.8	1.64
600	NV	0.0	-310	51	49	0.0	1.10	0.0	-601	37	36	0.0	1.07
	NC	0.0	-55	92	88	86.1	1.21	0.2	-163	117	104	50.2	1.64
	P1	0.0	7	97	98	94.9	1.06	0.0	15	120	118	94.9	1.19
	P2	0.0	8	99	96	95.3	1.19	0.3	26	147	141	95.5	1.72
Heteroscedastic error: $\eta \sim \text{scaled Beta}(2,1), \xi \mid X \sim \text{normal}$													
300	NV	0.0	-282	71	70	3.3	1.23	0.0	-587	53	51	0.0	1.21
	NC	0.0	-12	134	130	94.6	1.44	1.1	-45	553	265	86.1	2.00
	P1	0.0	17	141	141	95.5	1.17	0.0	31	173	175	95.8	1.48
	P2	0.0	18	135	136	95.3	1.29	0.8	42	188	195	95.9	1.68
600	NV	0.0	-287	50	50	0.0	1.02	0.0	-589	37	37	0.0	1.04
	NC	0.0	-25	92	90	91.8	1.25	0.2	-88	118	108	77.5	1.48
	P1	0.0	6	99	99	94.3	1.10	0.0	13	121	120	94.3	1.23
	P2	0.0	7	95	95	94.8	1.10	0.0	21	133	129	94.9	1.37
Homogeneous error: $\eta \sim \text{normal}, \xi=0$													
300	NV	0.0	-271	74	72	4.7	1.16	0.0	-593	56	53	0.0	1.09
	NC	0.0	16	140	139	95.2	1.29	3.3	66	392	262	94.2	2.13
	P1	0.0	12	139	142	96.1	1.22	0.2	25	179	179	96.1	1.46
	P2	0.0	12	132	133	95.0	1.24	0.0	32	183	183	95.8	1.51
600	NV	0.0	-274	52	51	0.1	1.05	0.0	-595	39	38	0.0	1.04
	NC	0.0	7	100	96	94.7	1.22	1.5	45	286	184	94.2	1.93
	P1	0.0	7	100	100	95.3	1.06	0.0	15	126	123	94.7	1.21

size	F	B	D	E	C	S	F	B	D	E	C	S	
	$\sigma_{\eta,\xi}=1/2$						$\sigma_{\eta,\xi}=1$						
P2	0.0	6	95	94	95.2	1.19	0.1	17	127	122	95.3	1.37	
Homogeneous error: $\eta \sim$ scaled Beta(2,1), $\xi=0$													
300	NV	0.0	-248	71	71	8.2	1.02	0.0	-561	54	54	0.0	1.00
	NC	0.0	24	141	132	94.5	1.23	0.0	43	188	167	93.9	1.70
	P1	0.0	24	151	143	94.4	1.44	0.0	41	195	180	94.6	1.78
	P2	0.0	24	141	135	94.7	1.37	0.1	46	197	183	95.9	1.70
600	NV	0.0	-252	50	50	0.0	1.12	0.0	-563	38	38	0.0	1.11
	NC	0.0	12	95	93	94.8	1.07	0.0	21	117	113	95.0	1.24
	P1	0.0	11	103	100	95.5	1.19	0.0	18	127	122	95.5	1.30
	P2	0.0	11	95	94	94.9	1.16	0.0	20	123	121	95.6	1.38

NV: naive method; NC: adapted nonparametric correction based on (9); P1 and P2: proposals corresponding to (7) and (8), respectively.
F: root-finding failure (%); B: bias ($\times 10^3$); D: standard deviation ($\times 10^3$); E: mean standard error ($\times 10^3$);
C: coverage (%) of 95% Wald confidence interval; S: skewness defined in (10).
B, D, E, C, and S were based on replicates with successful root-finding only.

Table 2

Simulation summary statistics of the estimators in the double-covariate model under heteroscedastic error contamination.

size	F	B	D	E	C	S	F	B	D	E	C	S	
$\sigma_{\eta,\xi} = 1/2$													
X and Z : independent													
300	NV	0.0	-294	72	69	2.3	1.13	0.0	-590	54	51	0.0	1.14
			-81	87	83	80.8	1.15		-169	85	81	43.1	1.09
	NC	0.1	-27	150	154	91.6	1.29	1.4	-116	221	163	72.5	1.71
			25	119	123	94.2	1.35		57	178	153	94.5	1.49
	P1	0.0	25	140	142	95.4	1.17	0.2	48	193	184	95.7	1.52
			20	111	106	94.1	1.13		34	141	133	95.7	1.30
600	P2	0.0	30	145	140	94.6	1.22	3.2	55	199	215	95.0	1.53
			21	111	104	93.3	1.26		37	148	147	95.2	1.40
	NV	0.0	-301	49	50	0.0	1.06	0.0	-595	37	37	0.0	1.08
			-88	59	59	66.4	0.99		-176	58	57	14.1	0.96
	NC	0.0	-44	90	89	88.4	1.17	0.5	-152	112	101	55.7	1.41
			11	76	74	95.0	1.23		28	112	101	94.4	1.37
300	P1	0.0	13	98	99	96.3	1.22	0.0	23	122	120	95.5	1.35
			10	76	74	94.9	1.04		16	92	89	94.2	1.18
	P2	0.0	16	98	97	94.8	1.17	0.4	38	149	148	95.8	1.81
			10	75	73	94.6	1.16		21	102	102	95.4	1.25
	X and Z : correlated with coefficient 0.5												
	300	NV	0.0	-341	78	75	1.1	1.03	0.0	-646	56	54	0.0
			44	100	94	92.4	1.25		84	101	94	84.5	1.17
NC		0.2	-23	179	164	92.7	1.52	2.1	-101	291	218	74.7	2.23
			20	122	118	93.7	1.13		53	223	175	92.6	1.02
P1		0.0	26	165	164	95.7	1.20	0.5	49	218	219	96.5	1.62
			14	117	113	93.7	1.03		24	148	141	94.2	1.02
600	P2	0.0	31	173	165	96.2	1.43	4.4	58	250	280	95.5	1.77
			18	116	110	93.0	1.14		33	159	159	94.0	1.12
	NV	0.0	-348	54	54	0.0	1.07	0.0	-651	39	39	0.0	1.10

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

size	$\sigma_{\eta,\xi} = 1/2$						$\sigma_{\eta,\xi} = 1$					
	F	B	D	E	C	S	F	B	D	E	C	S
NC	0.0	-42	107	106	89.7	1.27	0.6	-145	144	142	62.7	1.67
		10	81	79	94.4	1.13		23	120	114	92.4	1.08
	P1	0.0	13	110	115	96.2	1.22	0.0	25	141	144	96.4
P2												1.46
		7	80	79	94.2	1.01		12	98	97	94.4	1.03
	0.0	17	112	114	95.4	1.17	0.9	41	171	177	95.7	1.60
		8	79	78	94.5	1.14		17	107	107	94.3	1.14

See the footnote of Table 1.

The two rows for each setting correspond to the regression coefficients of X and Z .

Table 3

Comparison of regression coefficient estimators in the ACTG 175 data.

	log(CD4)		ZDV+ddl		ZDV+ddC		ddl	
	Est	SE	Est	SE	Est	SE	Est	SE
NV	-1.579	0.220	-0.515	0.235	-0.200	0.211	-0.423	0.218
NC	-2.440	0.365	-0.626	0.257	-0.276	0.233	-0.441	0.224
P1	-2.762	0.457	-0.431	0.259	-0.182	0.231	-0.358	0.237
P2	-2.795	0.449	-0.452	0.255	-0.101	0.231	-0.426	0.238

NV: naive method; NC: adapted nonparametric correction based on (9); P1 and P2: proposals corresponding to (7) and (8), respectively.

ZDV: zidovudine; ddl: didanosine; ddC: zalcitabine.