Survival Analysis with Heterogeneous Covariate Measurement Error

Yi Li* Louise Ryan†

*Harvard School of Public Health and Dana-Farber Cancer Institute, yili@jimmy.harvard.edu

†Harvard School of Public Health and Dana-Farber Cancer Institute, lryan@hsph.harvard.edu

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

<http://biostats.bepress.com/harvardbiostat/paper10>

Copyright ©2003 by the authors.

Survival Analysis with Heterogeneous Covariate Measurement Error

Yi Li and Louise Ryan *

Abstract

This paper is motivated by a time-to-event analysis where the covariate of interest was measured at the wrong time. We show that the problem can be formulated as a special case of survival analysis with heterogeneous covariate measurement error, and develop a general analytic framework. We study the asymptotic behavior of the naive partial likelihood estimates and analytically demonstrate that, under the heterogeneous measurement error structure and certain independence assumptions, these naive estimates will shrink toward zero, and that the degree of attenuation increases as the measurement error increases. We also give some counter examples for reverse-attenuation when the independence conditions are violated. We use our analytical results to derive a simple bias correcting estimator, which performs well in simulations for small and moderate amount of measurement error. Our framework can be used to provide insight into the behavior of the commonly used partial likelihood score test for testing no association between a failure outcome and an exposure, for example in the presence of measurement error or mistiming error. In particular, we derive the asymptotic distribution of the naive partial likelihood score test under a series of local alternatives, and discuss the asymptotic relative efficiency. As a result, a simple sample size formula to account for the contamination of covariates is obtained.

KEY WORDS: Cox models; Mistimed Covariates; Asymptotic Bias Analysis; Corrected Estimator; Asymptotic Relative Efficiency; Schoenfeld's Sample Size Formula.

RUNNING TITLE: Heterogeneous Covariate Measurement Error.

*Yi Li is Assistant Professor, Department of Biostatistics, Harvard School of Public Health and Dana-Farber Cancer Institute, Boston, MA 02115, and Louise Ryan is Professor, Department of Biostatistics, Department of Biostatistics, Harvard School of Public Health and Dana-Farber Cancer Institute, Boston, MA 02115. Yi Li's email address is yili@jimmy.harvard.edu and mailing address is: Department of Biostatistical Sciences, Dana-Farber Cancer Institute, 44 Binney Street, Boston, MA 02115. This work was supported in part by U. S. National Cancer Institute grants R01 CA95747 (Li) and R01 CA48061 (Ryan). The authors thank the Editor, an Associate Editor and three referees for helpful comments.

Collection of Biostatistics
Research Archive

1 INTRODUCTION

One serious problem in survival analysis is that major time-varying covariates of interest are often mistimed. For example, in clinical trials, some baseline biometric measurements, e.g. platelet count and serum creatinine, may not be available at time of entry but may be assessed during the early part of treatment. The ongoing Home Allergen Study (Gold et al., 1999), which motivates this work, was designed to assess environmental effects, such as bacterial endotoxin exposure at birth, on immunological function, allergy, and asthma in infants and young children. The scientific hypothesis is that immunological response to endotoxin exposure in infancy helps to “prime” the infant’s ability to respond to environmental triggers, and, hence, exposure to bacterial endotoxins early in life may confer protection against the development of allergy and asthma in later childhood. However, most noticeably in the design of this study, except for some newly born infants, the endotoxin exposure levels were not assessed at birth, and were often substituted with measurements made later in the same household. Issues of mistimed covariates have been addressed in a heuristic way by Keiding (1992) and in the context of a pharmacokinetic study by Higgins et al. (1997), but a detailed study has not been made in the context of survival analysis.

Of course, failure time regression subject to covariate measurement errors or missing covariates has aroused much interest in the past two decades. Prentice (1982) has shown the impact of measurement error by deriving the induced hazard function in the presence of covariate measurement error, and advocated a regression calibration method to draw inference; Zhou and Pepe (1995) and Zhou and Wang (2000) have discussed the use of the calibration approach when some covariates are missing, and Xie et al. (2001) and Wang et al. (2001) have applied a risk set calibration procedure in a measurement error setting. Nakamura (1992), Huang and Wang (2000), Tsiatis et al. (2001) and Hu and Lin (2002) have proposed corrected partial likelihood score approaches to correcting biases caused by measurement error. On the other hand, nonparametric maximum likelihood approaches have been adopted by Zhong et al. (1996), Hu et al. (1998) and Chen and Little (1999) and maximum partial likelihood estimators have been suggested by Paik and Tsai (1997). None of these authors, however, have derived expressions describing the asymptotic behavior of the standard Cox proportional hazards model when the covariates are contaminated. Furthermore, none of them have considered the setting of heterogeneous measurement error, namely, when the measurement error

variance changes across individuals. In a different context, the issue of heterogeneous measurement error has been addressed by Carroll and Stefanski (1990) for generalized linear models, but not for censored data. Our first objective, motivated by the work of Hughes (1993) on regression dilution in the naive partial likelihood estimates under a specific measurement error model, is to provide additional insight into the asymptotic behavior of these naive estimates in a broader context. Specifically, we recast the original mistiming problem in a more general framework of measurement error models, and discuss the asymptotic properties of the naive maximum partial likelihood estimates. We prove that under certain independence conditions the naive estimates will shrink toward zero, and the degree of attenuation increases as the measurement error becomes more severe. We also give some counter examples for reverse-attenuation when the independence conditions are violated. Moreover, based on the asymptotic biases, we derive a simple bias correcting estimator, which can be obtained with output from Cox regression analyses in standard software. Simulations indicated that this simple estimator performs well in simulations for small and moderate measurement errors. The second part of this article focuses on the behavior of the commonly used partial likelihood score test for hypothesis testing in the presence of general covariate measurement error. In particular, we derive the asymptotic distribution of the naive partial likelihood score test under a series of local alternatives, and compute its asymptotic relative efficiency. Our results yield an appealing sample size formula, useful for designing an observational study, to compensate for the efficiency loss due to covariate contamination in hypothesis testing.

2 Models and Notation

2.1 Mistimed Covariates Model

Assume that the survival times of m independent individuals are subject to right censoring, and that censoring is noninformative. Let $T_i = \min(\tilde{T}_i, C_i)$ be the observed survival for subject i ($i = 1, \dots, m$), where $\tilde{T}_i$ is the true survival time, and C_i is the potential censoring time. The individual-specific time origin is the date of a beginning event, for instance, birth, which effectively excludes the possibility of left censoring. Let $\delta_i = I(\tilde{T}_i \leq C_i)$ be the noncensoring indicator, which takes value 1 if a failure was observed and 0 otherwise. For subject i , let $\mathbf{X}_i(\cdot)$ be an l_1 -dimensional covariate process, which is actually continuous in time, but may be observed only at a few time points. In

our motivating home allergen study, for example, $\mathbf{X}_i(t)$ is the household Endotoxin exposure for subject i at time t . Suppose for each subject we are interested in relating the outcome T_i to the covariate process $\mathbf{X}_i(\cdot)$ measured at time 0, controlled for the effects of $\mathbf{Z}_i$, an $l_2 \times 1$ accurately measured covariate vector. That is, the outcomes $(T_i, \delta_i), i = 1, \dots, n$, are independent, with the hazard function being linked to the covariates through the following model

$$\lim_{dt \rightarrow 0} (dt)^{-1} P(t \leq \tilde{T}_i \leq t + dt | \tilde{T}_i \geq t, \mathbf{X}_i(\cdot), \mathbf{Z}_i) = \lambda\{t, \mathbf{X}_i(0), \mathbf{Z}_i\}. \quad (1)$$

By noninformative censoring we mean that C_i is independent of $\tilde{T}_i$ conditional on $\mathbf{X}_i(0)$ and $\mathbf{Z}_i$. A common choice of the hazard function is the proportional hazards model (Cox, 1972):

$$\lambda\{t, \mathbf{X}_i(0), \mathbf{Z}_i\} = \lambda_0(t) \exp\{\beta'_x \mathbf{X}_i(0) + \beta'_z \mathbf{Z}_i\}, \quad (2)$$

where β_x and β_z are fixed effects and $\lambda_0(t)$ is an unspecified baseline hazard function.

Model (1) indicates that, conditional on $\mathbf{X}_i(0)$ and $\mathbf{Z}_i$, the measurements of $\mathbf{X}_i(\cdot)$ at other time points provide no extra information regarding the outcome, an analogue to the classical assumption of non-differential measurement error (Carroll et al., 1995). The statistical challenge is that $\mathbf{X}_i(\cdot)$ is not measured at time 0, but is available at a later time, say, $\mathcal{T}_i$ (often pre-arranged in practice), which is assumed to be independent of the outcome and the concerned covariates. It is, however, customary in data analysis to fit a proportional hazards model (2) by ignoring the mistiming error and directly replacing the unobserved $\mathbf{X}_i(0)$ with the observed $\mathbf{X}_i(\mathcal{T}_i)$. Hence, it will be imperative to analyze the resulting biases.

The mistimed covariate model is completed by specifying the covariate process $\mathbf{X}_i(\cdot)$. We assume that $\mathbf{X}_i(t)$ are i.i.d stochastic processes

$$\mathbf{X}_i(t) = \mathbf{X}_i(0) + \sigma(t)\mathbf{B}_i(t) \quad (3)$$

where $\sigma(t)$ controls the magnitude of perturbation and the $\mathbf{B}_i(t)$ are independent mean zero stochastic processes with variance-covariance matrix $\mathbf{D}(t)$, which may depend on time t . But for simplicity (and identifiability), we assume that $\mathbf{D}(t)$ is independent of t and hence write it as $\mathbf{D}$. In a setting of mistimed covariates, we typically assume that the scale variance function $\sigma(t)$ is a non-decreasing function of t , exemplifying that the departure from the true covariate tends to amplify with time. Choices of $\sigma(t)$, for example, include, $\sigma(t) \equiv \text{constant}$ or $\sigma(t) = t$ or $\sigma(t) = \exp(t) - 1$, which will be

used in our numerical studies; Figure 1 illustrates the heterogeneous measurement error against the time lag, $\mathcal{T}$, under various $\sigma(\cdot)$. Often a moment-based method can be used to consistently estimate $\sigma(t)$ and $\mathbf{D}$ when the process $\mathbf{X}_i(\cdot)$ can be observed at multiple time points (see e.g. Carroll et al., 1993; Carroll et al., 1995). In a structural model, the true covariates $\mathbf{X}_i(0)$ are considered to be independently generated from a parameterized distribution, for example, a multivariate normal distribution, while in a functional model, they are considered as fixed but unobserved constants.

The naive estimator is the estimator under model (2) which ignores the mistiming error by directly replacing $\mathbf{X}_i \stackrel{\text{def}}{=} \mathbf{X}_i(0)$ by $\mathbf{W}_i \stackrel{\text{def}}{=} \mathbf{X}_i(\mathcal{T}_i)$. Thus, the naively specified hazard function is

$$\lambda_{i,\text{naive}}(t|\mathbf{W}_i, \mathbf{Z}_i) = \lambda_{0,\text{naive}}(t)e^{\mathbf{W}_i' \boldsymbol{\beta}_{x,\text{naive}} + \mathbf{Z}_i' \boldsymbol{\beta}_{z,\text{naive}}}. \quad (4)$$

We would expect that ignoring the mistiming error and directly utilizing the common tools for survival analysis, e.g. Cox partial likelihood score approach (Cox, 1972) will produce biased results. We investigate these analytically in Section 4. First, however, we embed our mistimed covariates model into a broader framework in the next section.

2.2 Heterogeneous Measurement Error Model

Model (3) differs from the classical measurement error model by involving a heterogeneous measurement error variance that depends on other available information, in our motivating example, the assessment time of covariates. This motivates a more general measurement error structure for the unobserved covariates:

$$\mathbf{W}_i = \mathbf{X}_i + \mathbf{u}_i \quad (5)$$

where, conditional on a subject-specific nonnegative random variable τ_i , the measurement error $\mathbf{u}_i$ are independently normally distributed with mean 0 and variance-covariance matrix $\tau_i \mathbf{D}$, and are independent of X_i, T_i, C_i . Here $\mathbf{D}$ is a nonrandom positive definite matrix (which may depend on observed quantities) and τ_i reflects the magnitude of measurement error. We further assume that τ_i are independent random variables with mean σ^2 , and a finite moment generating function $M_\tau(v, \sigma^2)$ (in a neighborhood of 0), where

$$M_\tau(v, \sigma^2) = E(e^{v\tau_i}; \sigma^2). \quad (6)$$

Finally, we assume that $\mathbf{X}_i$ has variance-covariance $\boldsymbol{\Sigma}_x$, but its distribution function, for the time being, is left unspecified. This new class of measurement error models, allowing the measurement

error variance to vary across subjects, is general and encompasses the classical additive measurement error structure (Fuller, 1987 and Carroll et al., 1995) and the heteroscedastic measurement error structure (Carroll and Stefanski, 1990). For example, it reduces to the classical additive model with $\tau_i = \sigma^2$ almost surely, the heteroscedastic measurement error with $\tau_i = f(\mathbf{X}_i)$ for some given function $f(\cdot)$, and the mistimed covariate model (3) with $\tau_i = \sigma(\mathcal{T}_i)$, where $\mathcal{T}_i$ is the actual time when the error-prone covariate was assessed. The measurement error models with nonconstant numbers of replicated samples (see e.g. Hu et al., 1998 and Xie et al., 2001) also fall into (5).

We turn now to an assessment of the asymptotic behavior of the naive estimates based on the standard Cox partial likelihood approach in the presence of heterogeneous covariate measurement errors. We proceed by introducing some standard counting process notation. Let $N_i = I(T_i \leq t, \delta_i = 1), i = 1, \dots, m$ be a right continuous process for individual i which documents the number of observed failures, 0 or 1, in an interval $[0, t]$, and $Y_i(t) = I(T_i \geq t)$, a predictable process indicating whether a subject is still at risk at time t . We assume $(N_i, Y_i, \mathbf{X}_i, \mathbf{Z}_i, \mathbf{W}_i, \tau_i)$ are i.i.d copies of $(N, Y, \mathbf{X}, \mathbf{Z}, \mathbf{W}, \tau)$.

On a probability space, say, $(\Omega, \mathcal{F}, P)$, denote by $\mathcal{F}_t = \sigma\{N_i(u), Y_i(u+), 0 \leq u < t, \mathbf{X}_i, \mathbf{Z}_i, \mathbf{W}_i, \tau_i, i = 1, \dots, m\}$ the increasing filtrations which contain the information about survival and covariates up to time t . Suppose, with respect to $\mathcal{F}_t$, N_i has an intensity function $Y_i(t)\lambda(t; \mathbf{X}_i, \mathbf{Z}_i)$, which may not necessarily be of the proportional form (2). Hence,

$$M_i(t) = N_i(t) - \int_0^t Y_i(u)\lambda(u; \mathbf{X}_i, \mathbf{Z}_i)du$$

are $\mathcal{F}_t$ -adapted independent local square integrable martingales with the variation processes being $\langle M_i, M_j \rangle(t) = 0$, if $i \neq j$, and $\langle M_i, M_i \rangle(t) = \int_0^t Y_i(u)\lambda(u; \mathbf{X}_i, \mathbf{Z}_i)du$ (see, e.g. Fleming and Harrington, 1991).

Now consider the situation where the intensity function for N_i is misspecified by (4). Denote by $\mathbf{W}_{i*} = (\mathbf{W}_i, \mathbf{Z}_i)$, $\beta = (\beta_x, \beta_z)$, and introduce $\mathbf{S}^{(j)}(t) = m^{-1} \sum_i \mathbf{W}_{i*}^{\otimes j} Y_i(t)\lambda(t; \mathbf{X}_i, \mathbf{Z}_i)$, $\mathbf{s}^{(j)}(t) = E\{\mathbf{S}^{(j)}(t)\}$, $\mathbf{S}^{(j)}(\beta, t) = m^{-1} \sum_i \mathbf{W}_{i*}^{\otimes j} Y_i(t) \exp(\beta' \mathbf{W}_{i*})$, $\mathbf{s}^{(j)}(\beta, t) = E\{\mathbf{S}^{(j)}(\beta, t)\}$ where $j = 0, 1, 2$, and the expectations are taken with respect to the true distributions of $N, Y, \mathbf{X}, \mathbf{W}, \mathbf{Z}$ based on models (8) and (5), and for a vector $\mathbf{a}$, $\mathbf{a}^{\otimes 0} = 1$, $\mathbf{a}^{\otimes 1} = \mathbf{a}$, $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}'$.

Under the naive model (4), the regression coefficients β would typically be estimated by maxi-

mizing the log partial likelihood function (Cox, 1972):

$$\ell(\beta) = \sum_{i=1}^m \int_0^{T_o} \beta' \mathbf{w}_{i*} dN_i - \int_0^{T_o} \log S^{(0)}(\beta, t) \sum_{i=1}^m dN_i(t),$$

or, equivalently, by solving the partial likelihood score equation,

$$U(\beta) = \sum_{i=1}^m \int_0^{T_o} \left\{ \mathbf{w}_{i*} - \frac{\mathbf{S}^{(1)}(\beta, t)}{S^{(0)}(\beta, t)} \right\} dN_i(t) = 0,$$

where T_o is a pre-specified constant such that it is within the support of the observed failure time, that is, $P\{\tilde{T}_i < C_i \wedge T_o\} > 0$. In practice, T_o is often the observed maximal study duration for each individual.

The next theorem shows that the sequence of naive estimates $\hat{\beta}_{naive}$ converges in probability to β^* , the solution to

$$\mathbf{h}(\beta) = \int_0^{T_o} \mathbf{s}^{(1)}(t) dt - \int_0^{T_o} \frac{\mathbf{s}^{(1)}(\beta, t)}{s^{(0)}(\beta, t)} s^{(0)}(t) dt. \quad (7)$$

Theorem 1 *Under regularity conditions R.1 and R.2 listed in Appendix A.1, the maximum partial likelihood estimate $\hat{\beta}_{naive}$ is a consistent estimator of β^* .*

Indeed, the asymptotic distribution of $\hat{\beta}_{naive}$ can also be investigated along the line of Lin and Wei (1989), and a related theorem is documented in Appendix A.2.

3 Proportional Hazards Models

Given the established asymptotic property of the naive estimates, we can evaluate the distortion of the covariate effects under a variety of model misspecifications. In the following we focus on the setting where one correctly specifies the proportional hazards model, but fails to account for mistiming errors or covariate measurement errors. Specifically, we apply Theorem 1 to study the asymptotic biases in the naive estimators when the mistiming error is ignored, and also numerically examine how the degree of censoring influences these biases.

3.1 Asymptotic Bias

Suppose the true hazard follows a proportional hazards model,

$$\lambda(t, \mathbf{X}, \mathbf{Z}) = \lambda_0(t) \exp\{\beta_{1x}^{(0)} X_1 + \dots + \beta_{l_1x}^{(0)} X_{l_1} + \beta_{1z}^{(0)} Z_1 + \dots + \beta_{l_2z}^{(0)} Z_{l_2}\}. \quad (8)$$

where $\mathbf{X} = (X_1, \dots, X_{l_1})$ and $\mathbf{Z} = (Z_1, \dots, Z_{l_2})$.

Denote by $\mathbf{X}_* = (\mathbf{X}, \mathbf{Z})$ and $\beta_0 = \{\beta_x^{(0)}, \beta_z^{(0)}\} = \{\beta_{1x}^{(0)}, \dots, \beta_{l_1x}^{(0)}, \beta_{1z}^{(0)}, \dots, \beta_{l_2z}^{(0)}\}$. Let $G(t, \mathbf{X}_*)$, $g(t, \mathbf{X}_*)$, and $C(t, \mathbf{X}_*)$ be the survival function for the survival time, the density function for the survival time, and the survival function for the censoring time respectively, where $G(t, \mathbf{X}_*) = e^{-\Lambda_0(t) \exp(\beta_0' \mathbf{X}_*)}$, $g(t, \mathbf{X}_*) = -\frac{\partial}{\partial t} G(t, \mathbf{X}_*)$ and $C(t, \mathbf{X}_*)$ is left unspecified. Assume that $(N_i, Y_i, \mathbf{W}_i, \mathbf{X}_i, \mathbf{Z}_i)$ are independent across subjects. Then, as implied by Theorem 1, the asymptotic limit of the naive estimate $\hat{\beta}_{naive}$, denoted by

$$\beta^*(\sigma^2) = \{\beta_{1x}^*(\sigma^2), \dots, \beta_{l_1x}^*(\sigma^2), \beta_{1z}^*(\sigma^2), \dots, \beta_{l_2z}^*(\sigma^2)\},$$

is the solution to (7), or, more specifically, to

$$0 = \int_0^{T_o} E\{gC\mathbf{X}_*\} - E(gC) \left\{ \frac{E(GCe^{\beta' \mathbf{X}_* \mathbf{X}_*})}{E(GCe^{\beta' \mathbf{X}_*})} + \mathbf{F} \mathbf{D} \beta_x p\left(\frac{1}{2} \beta_x' \mathbf{D} \beta_x, \sigma^2\right) \right\} dt \quad (9)$$

where $p(v, \sigma^2) \stackrel{def}{=} \frac{\partial}{\partial v} \log M_\tau(v, \sigma^2)$, $\mathbf{F} = (\mathbf{I}_{l_1}, \mathbf{0}_{l_1 \times (l_1 + l_2)})'$, $\mathbf{I}_{l_1}$ is an $l_1 \times l_1$ identity matrix and $\mathbf{0}_{l_1 \times (l_1 + l_2)}$ is an $l_1 \times (l_1 + l_2)$ matrix with all its entries being 0. Here we write $g = g(t, \mathbf{X}_*)$ and $C = C(t, \mathbf{X}_*)$ for notational simplicity. That (7) equals to (9) in the case of proportional hazards model follows from the non-differentiability assumption of measurement error and the double expectation theorem.

With regularity condition (R.2) and by the implicit function theorem, the solution to (9) exists, and is a smooth function of σ^2 , reflecting the average magnitude of measurement errors $\mathbf{u}_i$. In particular, when $\sigma^2 = 0$ ($\tau_i = 0$ almost surely), $\beta^*(0) = \beta_0$ is the solution to (9). When $\sigma^2 \rightarrow \infty$, $\beta_x^*(\sigma^2) \rightarrow 0$, and the contaminated covariate $\mathbf{W}$ provides no information about the outcome. In this case, we essentially deal with the Cox proportional hazards model with omitted covariates, and the asymptotic limit $\beta_z^*(\infty)$ is the same as that identified by Struthers and Kalbfleisch (1986), Bretagnolle and Huber-Carol (1988) and Schmoor and Schumacher (1997).

Under certain conditions, a more in-depth investigation leads to the following theorem, which establishes that $\beta^*(\sigma^2)$ is indeed a monotone function of σ^2 , the average of measurement error variance in the setting of heterogeneous measurement error [c.f. (5)]. That is, the attenuation in the naive estimates becomes more severe as σ^2 increases; see Appendix A.3 for a proof.

Theorem 2 *In addition to regularity conditions (R.1) and (R.2) in Appendix A.1, assume conditions (C.0)-(C.5) listed in Appendix A.3.*

- If $\beta_x^{(0)} = \mathbf{0}$, then $\beta_x^*(\sigma^2) \equiv 0$; for each $j = 1, \dots, l_1$, if $\beta_{jx}^{(0)} > 0$, then $\beta_{jx}^*(\sigma^2) > 0$ and $\beta_{jx}^*(\sigma^2)/\partial\sigma^2 < 0$; if $\beta_{jx}^{(0)} < 0$, then $\beta_{jx}^*(\sigma^2) < 0$ and $\beta_{jx}^*(\sigma^2)/\partial\sigma^2 > 0$.
- If $\beta_x^{(0)} = \mathbf{0}$, then $\beta_{jz}^*(\sigma^2) = \beta_{jz}^{(0)}$ for each $j = 1, \dots, l_2$; assume there exists a $1 \leq k \leq l_1$ such that $\beta_{kx}^{(0)} \neq 0$, if $\beta_{jz}^{(0)} > 0$, then $\beta_{jz}^*(\sigma^2) > 0$ and $\beta_{jz}^*(\sigma^2)/\partial\sigma^2 < 0$, and if $\beta_{jz}^{(0)} < 0$, then $\beta_{jz}^*(\sigma^2) < 0$ and $\beta_{jz}^*(\sigma^2)/\partial\sigma^2 > 0$.

Remark 1: While Prentice (1982) and Hughes (1993) demonstrated regression dilution in the naive partial likelihood estimates under a specific measurement error model, this theorem analytically elucidates the attenuation phenomenon under a more general measurement error model. That is, under some assumptions of independence, as long as the observed covariates correlate positively with the unobserved true covariates, the presence of measurement error will only attenuate the covariate effect by the naive partial likelihood estimation, while the direction of impact will be preserved.

In the usual linear regression, the naive estimates for the effects of correctly measured covariates are consistent if they are independent of the contaminated covariates. In contrast, this theorem indicates that this is not the case in the survival analysis. Under some independence conditions, the naive estimate of the effect for a correctly measured covariate, albeit independent of the contaminated covariate, is attenuated as well, and the degree of such attenuation increases with the magnitude of measure error. Nonetheless, the direction of the covariate effect is preserved. Numerical studies under the survival model (8) and the mistimed covariate model (3) were performed to illustrate the theoretical results, where the parameter values used were as follows: $l_1 = l_2 = 1$; baseline hazard $\lambda_0 = 1$, the true regression coefficients $\beta_x^{(0)} = 1, \beta_z^{(0)} = 1$; the true X and Z were independent and were assumed to follow the standard normal distribution; the conditional variance of measurement error $\tau_i = \sigma_0^2 \sigma(\mathcal{T}_i)$ in (5); the censoring time was assumed to be independent of X and survival time and follows an exponential distribution, that is, $C(t) = \exp(-at)$; the maximal study duration for each individual $T_o = 2$. We assume that the measurement time $\mathcal{T}_i$ for the error-prone covariate follows a random uniform $U[0, 1]$. Denote by $\rho = \sigma_x/(\sigma_x^2 + \sigma^2)^{1/2}$ the marginal correlation between the observed value W and the true underlying covariate X . By varying σ_0^2 , we let ρ range from 1 to 0, with $\rho = 1$ corresponding to no mistiming error and $\rho = 0$ indicating the worst situation where the observed covariate W contain no information about the true underlying covariate X . We chose a to be 0.13, 0.47, 1.7, 5.25 to obtain censoring proportions roughly equal to 10%, 30%, 60%, and

80%. Figure 2 illustrates the asymptotic relative bias, defined by the difference between the estimate and the true value divided by the true value, in the naive estimates $\hat{\beta}_{x,naive}$ and $\hat{\beta}_{z,naive}$, under measurement error variance function $\sigma(t) = t$, and with different censoring proportions. It shows that the biases increase as the correlation between the true and the observed covariates becomes weaker, which coincides with our theoretical results. In addition, we notice that the biases in β_x^* and β_z^* increase when more observations are censored. Similar patterns repeat when $\sigma(t) = 1$ or $\exp(t) - 1$.

Remark 2: Though it may be a common conception that the measurement error would cause attenuation in the estimation of regression coefficients, reverse-attenuation examples can be found if the independence assumptions on the covariates and the measurement error are violated. To see this, we denote by Σ_* the variance-covariance matrix for the true covariates $\mathbf{X}_*$ and $\mathbf{D}_*$ a $(l_1 + l_2) \times (l_1 + l_2)$ matrix whose first $l_1 \times l_1$ block is $\sigma^2 \mathbf{D}$ and the rest zero. Then under the rare event assumption (i.e the event occurs with a negligible probability), using a linearization technique (see Carroll et al. 1995), the asymptotic limit of the naive estimate will be approximated by

$$\beta^* = \Sigma_*(\Sigma_* + \mathbf{D}_*)^{-1}\beta_0. \quad (10)$$

If Σ_* or $\mathbf{D}_*$ is not diagonal (i.e. the independence assumptions are not satisfied), we can construct the scenarios where some individual components of β^* have up-ward biases or zero-crossings. For example, consider a bi-covariate model (8) with $l_1 = 2$ and $l_2 = 0$. and assume $\mathbf{X} = (X_1, X_2) \sim MVN(0, \Sigma_*)$ and measurement error $\mathbf{u} \sim MVN(0, \mathbf{D}_*)$. Set the true parameter values as $\beta_1^{(0)} = 1, \beta_2^{(0)} = 3.6$. Using (10), we compute the asymptotic limits of the naive estimates by considering different structures of Σ_* and $\mathbf{D}_*$.

(i) (correlated covariates with independent error) Let $\Sigma_* = (\Sigma_{ij})_{2 \times 2}$, where $\Sigma_{11} = \Sigma_{22} = 1$ and $\Sigma_{12} = \Sigma_{21} = 0.9$, and $\mathbf{D}_* = \text{diag}(0.4, 0.4)$. Then $\beta_1^* = 1.64(> \beta_1^{(0)})$ and $\beta_2^* = 2.16$.

(ii) (independent covariates with correlated error) Let $\Sigma_* = \mathbf{I}_2$, and $\mathbf{D}_* = (D_{ij})_{2 \times 2}$ where $D_{11} = D_{22} = 1$ and $D_{12} = D_{21} = -0.35$. Then $\beta_1^* = 1.44(> \beta_1^{(0)})$ and $\beta_2^* = 2.93$.

(iii) (zero crossing) Let $\Sigma_{11} = \Sigma_{22} = 1$ and $\Sigma_{12} = \Sigma_{21} = -0.9$, and $\mathbf{D}_* = \text{diag}(0.4, 0.4)$. Then $\beta_1^* = -0.61(< 0)$ and $\beta_2^* = 1.53$.

3.2 A First Order Bias Correcting Estimator

Immediately, when σ^2 is small, an approximation to the true β_0 can be obtained via inverting a Taylor's expansion. Specifically, expanding $\beta^*(\sigma^2)$ around $\sigma^2 = 0$ gives

$$\beta_0 = \beta^*(\sigma^2) - \sigma^2 \frac{\partial}{\partial \sigma^2} \beta^*(\sigma^2)|_{\sigma^2=0} + o(\sigma^2) \quad (11)$$

where

$$\frac{\partial}{\partial \sigma^2} \beta^*(\sigma^2)|_{\sigma^2=0} = - \left\{ \int_0^{T_0} E(gC) dt \right\} \mathbf{V}^{-1}(\beta_0, 0) \mathbf{F} \mathbf{D} \beta_x^{(0)}, \quad (12)$$

where

$$\begin{aligned} \mathbf{V}(\beta, \sigma^2) = & \int_0^{T_0} \frac{E(gC)}{\{E(GC e^{\beta \mathbf{X}_*})\}^2} (E[GC e^{\beta \mathbf{X}_*} \{\mathbf{X}_* (\mathbf{X}_*)' + p_1(\frac{1}{2} \beta'_x \mathbf{D} \beta_x, \sigma^2) (\mathbf{F} \mathbf{D} \beta_x)^{\otimes 2} \\ & + p(\frac{1}{2} \beta'_x \mathbf{D} \beta_x, \sigma^2) \mathbf{F} \mathbf{D} \mathbf{F}\}] E(GC e^{\beta \mathbf{X}_*}) - \{E(GC \mathbf{X}_* e^{\beta \mathbf{X}_*})\}^{\otimes 2}) dt. \end{aligned} \quad (13)$$

Omitting the higher order terms in σ^2 from (11) and considering (12), we derive in Appendix A.5 a first order bias correcting estimator,

$$\tilde{\beta} = (\mathbf{I} - \sigma^2 \bar{N} \hat{\mathbf{V}}^{-1} \mathbf{F} \mathbf{D} \mathbf{F}')^{-1} \hat{\beta}_{naive} \quad (14)$$

where $\mathbf{I}$ is an $(l_1 + l_2) \times (l_1 + l_2)$ identity matrix, and $m\bar{N} = \sum_{i=1}^m N_i(T_0)$, the number of events observed during study, $\hat{\beta}_{naive}$ is the naive estimate, $\hat{\mathbf{V}}$ is an approximation to $\mathbf{V}(\beta_0, 0)$, given by

$$\hat{\mathbf{V}} = \int_0^{T_0} \left[\frac{\mathbf{S}^{(2)}(\hat{\beta}_{naive}, t)}{S^{(0)}(\hat{\beta}_{naive}, t)} - \left\{ \frac{\mathbf{S}^{(1)}(\hat{\beta}_{naive}, t)}{S^{(0)}(\hat{\beta}_{naive}, t)} \right\}^{\otimes 2} \right] \frac{1}{m} \sum_{i=1}^m dN_i(t),$$

and $S^{(j)}(\beta, t) = m^{-1} \sum_i \mathbf{W}_{i*}^{\otimes j} Y_i(t) \exp(\beta \mathbf{W}_{i*})$, $j = 0, 1, 2$. Here, σ^2 is assumed known or can be consistently estimated from validation data sets or reliability samples. Notice that $(m\hat{\mathbf{V}})^{-1}$ is the naive variance-covariance estimate of $\hat{\beta}_{naive}$, which is available from the common statistical software, e.g. SAS PROC PHREG and S-Plus COXPH. Hence, the corrected estimates can be conveniently obtained from the output of the naive fitting, bringing considerably less computation. As indicated in Appendix A.5, the corrected estimator is applicable in much more general situations without the restrictive assumptions on the covariates and the measurement error required for the theoretical bias analyses. Note that (14) reduces to the naive estimates when $\sigma^2 = 0$, and that the form of (14) resembles the corrected estimator in linear regressions (Carroll et al., 1995, ch. 2) and in survival analysis with constant measurement error variance (Kong, 1999).

An application of the delta method yields an approximate ‘sandwich’ variance estimator for (14),

$$\widehat{\text{Var}}(\tilde{\beta}) = (\mathbf{I} - \sigma^2 \bar{N} \hat{\mathbf{V}}^{-1} \mathbf{FDF}')^{-1} \widehat{\text{Var}}(\hat{\beta}_{naive}) (\mathbf{I} - \sigma^2 \bar{N} \hat{\mathbf{V}}^{-1} \mathbf{FDF}')^{-T}. \quad (15)$$

Here $\widehat{\text{Var}}(\hat{\beta}_{naive})$ is given in (29) in Appendix A.5, and is available by invoking the option of robust variance in Splus COXPH and SAS PHREG (version 8.1) (or SAS PHLEV macro, downloadable from www.mayo.edu/hsr/biostat.html), which agrees well with the naive variance obtained from the Martingale residuals for a relatively small σ^2 .

However, (15) may lead to underestimation, especially when σ^2 is large, as it fails to account for the variation of $\bar{N}$ and $\hat{\mathbf{V}}$ in the formulation. A simple but effective alternative is to apply a Bootstrap procedure (Efron, 1979). Specifically, we resample m subjects, with replacement, from $(T_i, \delta_i, \mathbf{W}_{i*})|_{i=1}^m$ to obtain a new data set $\{T_{(i)}, \delta_{(i)}, \mathbf{W}_{(i*)}\}_{i=1}^m$. Given this new dataset, we use (14) to compute the corrected estimates. Such a procedure can be repeated K times to obtain a sequence of estimates, $\tilde{\beta}^{(k)}, k = 1, \dots, K$. The bootstrap variance estimates can hence be calculated using the sample variances

$$\text{var}_{\text{boot}}(\tilde{\beta}) = \frac{1}{K-1} \sum_{k=1}^K \{\tilde{\beta}^{(k)} - \bar{\beta}_{boot}\} \{\tilde{\beta}^{(k)} - \bar{\beta}_{boot}\}',$$

where $\bar{\beta}_{boot} = \frac{1}{K} \sum_{k=1}^K \tilde{\beta}^{(k)}$. In practice, it is adequate to choose a moderate number of resamplings, K , say, in the range 25 to 100 (Lange, 1998, p.301). We chose $K = 30$ in simulations.

3.3 Simulations and Two Worked Examples

Simulations were performed to examine the finite sample performance of the first order corrected estimator (14). With similar configurations as in the foregoing section, we varied the sample size from 80 to 500 and the average measurement error variance from 0.05 to 0.90. The parameter in the censoring distribution was taken to be 0.47, corresponding roughly to a 30% censoring proportion. A total of 1000 simulations were conducted for each configuration. Figure 3 depicts the relative biases in the corrected estimates, compared with those in the naive estimates. It appears that with a small sample size, the corrected estimator performs well for small and moderate measurement errors (e.g. $0 \leq \sigma^2 \leq 0.5$), while, with a moderate or a large sample size (e.g. $m = 200$ or $m = 500$), the corrected estimator performed well even for large measurement errors (e.g. $\sigma^2 = 0.9$). To examine the performance of the estimator with different magnitude of the underlying covariate effect and

variance function $\sigma(\cdot)$, we varied $\beta_x^{(0)}$ from 0.25 to 1.0 and chose $\sigma(t) = 1$ or $\exp(t) - 1$ and found similar patterns. Table 1 displays the averages of the naive estimates and the corrected estimates, along with their empirical variances and the average bootstrap variances (with $\sigma(t) = t$). It appears that the bootstrap variances are in agreement with the empirical counterparts for small sample sizes with small and moderate measurement errors, and for large sample sizes even with relatively large measurement errors.

We illustrate the practical usage of (14) with two published studies. The first is Framingham study, a subset of which was analysed by Xie et al. (2001). Of particular interest was the effect of long term average systolic blood pressure (SBP) on developing coronary heart disease for middle aged men. A total of 85 failures of 423 subjects were observed in a span of 10 years of follow up. The true covariate X_i is the long-term average SBP, which was subject to error due to variations of SBP from time to time. Multiple measurements of SBP were made for each individual, which were assumed to follow

$$W_{ij} = X_i + u_{ij}, \quad (16)$$

for $i = 1, \dots, n_i$. Here $u_{ij} \stackrel{i.i.d}{\sim} N(0, \sigma_u^2)$, and are independent of X_i . The within-subject average of measurements, e.g. $\bar{W}_i = \frac{1}{n_i} \sum_i W_{ij}$, was regarded as the surrogate for X_i . Hence, for each individual the measurement error variance is σ_u^2/n_i . In this data subset $n_i \equiv 2$, and σ_u^2 was identified to be 0.039, indicating for each individual the average measurement error variance is 0.019. The reported naive estimate for the Cox model was 1.441 with naive variance 0.089. Immediately, using (14) and (15) with $l = 0$, we acquired that the first order corrected estimate is 1.69 (SE=0.350) which matches well with the ordinary regression calibration estimate 1.74 (SE=0.366) and the risk set regression calibration estimate 1.76 (SE=0.372) reported by Xie et al. (2001).

Another example is the ACTG 116b/117 study (Kahn et al., 1992) on the relationship between the number of CD4 counts and the disease progression free survival. Of 912 patients accrued, a total of 334 events, including disease progression or death, were observed during study. The number of CD4 replicates varied across patients: 56, 310, 541, and 5 patients had 1,2,3,4 replicate measurements respectively. Under model (16), Hu et al. (1998) calculated $\sigma_u^2 = 0.1505$, and reported the naive estimate of β_x is -0.6015 with naive variance 0.00319. The average measurement error variance were calculated by $\sigma^2 = (56 \times \sigma_u^2 + 310 \times \sigma_u^2/2 + 541 \times \sigma_u^2/3 + 5 \times \sigma_u^2/4)/912 = 0.0646$. Hence, applying formulae (14) and (15) yielded the corrected estimate -0.6497 (SE=0.061), which is in

agreement with the regression calibration estimate -0.6307 (SE=0.059), fully parametric estimate -0.6597 (SE=0.062), fully nonparametric estimate -0.6216 (SE=0.057) and semiparametric estimate -0.6641 (SE=0.063) reported by Hu et. al (1998).

4 Hypothesis Testing in the Presence of General Covariate Measurement Error

A relatively unexplored subject in survival analysis is on hypothesis testing in the presence of mistiming error or general measurement error. For simplicity of exposition, we postulate that the true hazard adheres to a single covariate proportional hazards model,

$$\lambda(t, X_i) = \lambda_0(t) \exp(\beta X_i).$$

For testing the null hypothesis $\beta = 0$, a naive partial likelihood score test would practically be utilized by ignoring mistiming error or measurement error and computing

$$m^{-1/2}U_m = m^{-1/2} \sum_{i=1}^m \int_0^{T_o} \{W_i - \bar{W}(t)\} dN_i(t), \quad (17)$$

where $\bar{W}(t) = \frac{\sum_{i=1}^m Y_i(t)W_i}{\sum_{i=1}^m Y_i(t)}$. Under the null hypothesis, the naive model (4) coincides with the true model. Therefore, test (17) is asymptotically a mean zero normally distributed random variable. With a properly calculated variance (given in Theorem 4), test (17) is valid, and retains the nominal level under the null hypothesis. But, as shown in the following, the induced survival model under the alternative hypothesis does not preserve proportionality, and, as a result, the loss of efficiency in test (17) shall be expected (Lagakos, 1988). Hence, it will be of substantial interest to analytically characterize the efficiency loss due to covariate measurement errors.

4.1 Asymptotic Relative Efficiency and Its Applications

To facilitate the following discussion, suppose we may invert (5) so that

$$X_i = \alpha_i^* + \rho_i^* W_i + \epsilon_i \quad (18)$$

where $\rho_i^* = \frac{\sigma_x^2}{\sigma_x^2 + \tau_i}$, $\sigma_x^2 = \text{Var}(X)$, α_i^* is a constant possibly depending on τ_i , and ϵ_i is a mean 0 random variable and is independent of W_i , conditional on τ_i . Indeed (18) holds if we assume that the unobserved covariate X_i follows a normal distribution.

Similarly we may calculate the hazard function for each counting process N_i based on the observed covariates W_i and measurement error variance τ_i ,

$$\lambda(t|W_i, \tau_i) = \lambda_0(t)e^{\beta\alpha_i^* + \beta\rho_i^*W_i + \psi(\beta, W_i, \tau_i, t)} \quad (19)$$

where $\psi(\beta, W_i, \tau_i, t) = \log E\left(e^{\beta\epsilon_i}|\tilde{T}_i \geq t, W_i, \tau_i\right)$.

Note that $\psi(\beta, W_i, \tau_i, t) \equiv 0$ when $\tau_i = 0$ or $\beta = 0$, corresponding to the case of no measurement error and the null model respectively; and because of measurement error, the induced hazard function fails to preserve the proportionality. Hence the simple log rank test (17) may be inefficient relative to that in the absence of measurement error. We explore this more formally by deriving the asymptotic relative efficiency (ARE), a useful device in comparing tests, of the simple log rank test versus its counterpart in the absence of measurement error. To proceed, we consider a sequence of local alternatives that converges to the null hypothesis at the appropriate rate as sample size increases to infinity (Fleming and Harrington, 1991). Under these alternatives, the log rank statistic has, asymptotically, a finite mean and variance.

For notational convenience, denote by

$$\phi(\beta, W_i, \tau_i, t) = \beta\alpha_i^* + \beta\rho_i^*W_i + \psi(\beta, W_i, \tau_i, t), \quad (20)$$

$$s_\beta^{(j)}(t) = E\{Y_i(t)W_i^j\phi_\beta(\beta, W_i, \tau_i, t)\lambda_0(t)\}, \quad (21)$$

$$s^{(j)}(t) = E\{Y_i(t)W_i^j\lambda_0(t)\}, \quad (22)$$

where the expectation is taken with respect to N, Y, W, τ for $j = 0, 1, 2$, and $\phi_\beta(\cdot)$ is the partial derivative of $\phi(\cdot)$ with respect to β . Assuming sufficient regularity conditions [see, e.g. Andersen et al. (1982)], and a sequence of local alternatives, $m^{-1/2}\eta$ for some fixed η ($-\infty < \eta < \infty$), we have the following theorem, whose proof can be found in Appendix A.6, about the asymptotic behavior of the log rank test (17).

Theorem 3 *With the conditions listed in Andersen and Gill (1982) and under the hypothesis $\beta_m = m^{-1/2}\eta$, $m^{-1/2}U_m$ converges in distribution to a normal distribution with mean*

$$\int_0^{T_o} \eta \frac{s^{(1)}(t)s_\beta^{(0)}(t) - s^{(0)}(t)s_\beta^{(1)}(t)}{s^{(0)}(t)} dt$$

and variance

$$\int_0^{T_o} \left[\frac{s^{(2)}(t)}{s^{(0)}(t)} - \left\{ \frac{s^{(1)}(t)}{s^{(0)}(t)} \right\}^2 \right] s^{(0)}(t) dt,$$

where $s_{\beta}^{(j)}(t)$ and $s^{(j)}(t)$ are defined in (21) and (22), and all the expectations are taken under the null hypothesis $\beta = 0$, and the derivatives are evaluated at $\beta = 0$.

Therefore, under the null hypothesis, that is, $\eta = 0$, the naive score test is unbiased, and under the local alternatives, $\eta \neq 0$, the asymptotic distribution of the naive score test is still normal with the same variance as in the null case, but the asymptotic mean is shifted. Thus, we can compute the asymptotic efficacy of (17), which is defined by its noncentrality,

$$\frac{\eta^2 \left(\int_0^{T_o} \frac{s^{(1)}(t)s_{\beta}^{(0)} - s^{(0)}(t)s_{\beta}^{(1)}}{s^{(0)}(t)} dt \right)^2}{\int_0^{T_o} \left[\frac{s^{(2)}(t)}{s^{(0)}(t)} - \left\{ \frac{s^{(1)}(t)}{s^{(0)}(t)} \right\}^2 \right] s^{(0)}(t) dt}.$$

It is reasonable in practice to assume that the censoring time is independent of X_i and W_i , which results in a more simplified efficacy. Under the null hypothesis $\beta = 0$, since the observed survival time T is independent of W , then for $j = 0, 1, 2$,

$$\begin{aligned} s_{\beta}^{(j)}(t) &= \pi(t)\lambda_0(t)E\{W_i^j \phi_{\beta}(\beta, W_i, \tau_i, t)\} \\ s^{(j)}(t) &= \pi(t)\lambda_0(t)E(W_i^j), \end{aligned}$$

where $\pi(t) = E\{Y_i(t)\} = P(T_i \geq t)$. Direct calculations give that

$$\frac{\partial}{\partial \beta} \psi(\beta, W, \tau, t)|_{\beta=0} = 0, \quad (23)$$

and, hence, $\phi_{\beta}(\beta, W, \tau, t)|_{\beta=0} = \alpha^* + \rho^* W_i$, where $\rho^* = \sigma_x^2/(\sigma_x^2 + \tau)$. It follows that the simplified efficacy is

$$eff_{naive} = \frac{\eta^2 [E(W)E(\rho^* W) - E(\rho^* W^2)]^2}{\text{Var}(W)} \int_0^{T_o} \pi(t)\lambda_0(t) dt. \quad (24)$$

Write $P_{T_o} = \int_0^{T_o} \pi(t)\lambda_0(t) dt$. With basic martingale theory, it follows that $P_{T_o} = E\{N(T_o)\}$ under the null hypothesis, indicating P_{T_o} is the probability of observing a failure by time T_o . Moreover, using double expectation gives that $\text{Var}(W) = \sigma_x^2 + \sigma^2$, $E(\rho^* W) = E\{E(\rho^* W|\tau)\} = E(\rho^*)E(X)$, and $E(\rho^* W^2) = E\{E(\rho^* W^2|\tau)\} = E\{\rho^*(\sigma_x^2 + \tau)\} + E(\rho^*)\{E(X)\}^2 = \sigma_x^2 + E(\rho^*)\{E(X)\}^2$. Hence, (24) can be rewritten as

$$eff_{naive} = \frac{\eta^2 \sigma_x^4}{\sigma_x^2 + \sigma^2} P_{T_o}. \quad (25)$$

Therefore, in the absence of measurement error ($\sigma^2 = 0$), the maximum efficacy of test (17) is achieved at $\eta^2 \sigma_x^2 P_{T_o}$.

By calculating the ratio of efficacies, it follows that the asymptotic relative efficiency (ARE) comparing the log rank test in the presence of measurement error to that in the absence of measurement error is

$$\frac{\sigma_x^2}{\sigma_x^2 + \sigma^2}, \quad (26)$$

which is a monotonically decreasing function of σ^2/σ_x^2 , the relative variability of measurement error with respect to the true underlying covariate. A relative efficiency less than 1 indicates that more observations are needed for the test in the presence of measurement error. Noticeably in the context of mistimed covariate problem, this result does not depend on the specific forms of the variance function $\sigma(t)$ in (3) and the distribution of the measurement time $\mathcal{T}_i$, but only on σ^2 , the average error variance.

The above calculations have an immediate application in study design. Specifically, for a fixed alternative $\beta = \beta_a$ which is in the range of $O(m^{-1/2})$, by Theorem 4, the distribution of the simple log rank test under the true model and under the null hypothesis is approximately given by

$$N \left\{ m^{1/2} \beta_a \sigma_x^2 P_{T_o}, (\sigma_x^2 + \sigma^2) P_{T_o} \right\}.$$

It then follows that the sample size needed to detect $H_a : \beta = \beta_a \neq 0$ versus $H_0 : \beta = 0$ for the naive test with power δ and one-sided type I error level of ϵ is

$$\frac{(\mathcal{Z}_{1-\epsilon} + \mathcal{Z}_\delta)^2}{\beta_a^2 \text{eff}_{naive}},$$

or, equivalently, the number of events required is

$$(\mathcal{Z}_{1-\epsilon} + \mathcal{Z}_\delta)^2 \frac{\sigma_x^2 + \sigma^2}{\beta_a^2 \sigma_x^4} \quad (27)$$

where $\mathcal{Z}_q$ is the $100 \times q$ percentile of a standard normal distribution. Note that (27) is essentially an extension of Schoenfeld's sample size formula (Schoenfeld, 1983) which applies in the absence of measurement error. Hence, if a conventional partial likelihood score test is opted for, the sample size should be inflated by σ^2/σ_x^2 to compensate for the efficiency loss due to measurement error.

4.2 Optimality of Log Rank Tests with General Covariate Measurement Error

A natural question would be whether the efficiency of test (17) can be improved within a more general class of weighted log rank tests:

$$m^{-1/2} \tilde{U}_m = m^{-1/2} \sum_{i=1}^m \int_0^{T_o} r_m(t) \{W_i - \bar{W}(t)\} dN_i(t), \quad (28)$$

by choosing a proper weight function $r_m(t)$. Here $r_m(t)$ is a bounded predictable process, converging uniformly in probability to a bounded nonrandom function $r(t)$ over $[0, T_o]$. For example, $r_m(t) = \{\hat{S}(t-)\}^\rho \{1 - \hat{S}(t-)\}^\gamma$, where $\hat{S}(\cdot)$ is the Kaplan-Meier survival estimate, corresponds to the $G^{\rho, \gamma}$ class test [see, e.g. Fleming and Harrington (1991, ch.7)]. Yet, as revealed by the following theorem (proved in Appendix A.7), with a noninformative censoring mechanism, test (17) is in fact the optimal test within the general class (28).

Theorem 4 *Assume that for each i the censoring time C_i is independent of the observed covariate W_i . Test (17) achieves the maximum efficacy within the general log-rank test class (28).*

Hence, tests beyond the classical log-rank test family (28) need to be considered for improving efficiency. Alternative choices, for example, include generalized log rank tests where a subject-specific weight function is assigned to each individual:

$$m^{-1/2}U_m = m^{-1/2} \sum_{i=1}^m \int_0^{T_o} r_{m,i}(t) \{W_i - \tilde{W}_m(t)\} dN_i(t),$$

where $\tilde{W}_m(t) = \frac{\sum_{i=1}^m Y_i(t) r_{m,i}(t) W_i}{\sum_{i=1}^m Y_i(t) r_{m,i}(t)}$. A heuristic strategy would be to put more weight on subjects whose covariates are measured with more precision, though this warrants a detailed investigation. In a different context, Kong and Slud (1997) and DiRienzo and Lagakos (2001) considered improving test (28) for two treatment comparisons with a misspecified proportional hazard model.

5 Discussion

In this article, we have discussed survival analysis with mistimed covariates within a more general heterogenous measurement error framework. To understand the consequence of mistiming error on parameter estimation, we have focused on the asymptotic behavior of the naive maximum partial likelihood estimates, and have shown that, under the heterogeneous measurement error structure and certain independence assumptions, these naive estimates will shrink toward zero, and that the degree of attenuation increases as the measurement error increases. We also give some counter examples for reverse-attenuation when these independence conditions are violated. To our knowledge, this result is also new in the context of survival analysis and is against the common conception that measurement error will always lead to attenuation.

Moreover, as a by-product of our asymptotic bias analyses, we have obtained a simple bias correcting estimator, which performs well for small sample sizes with small and moderate measurement errors. As the proposed estimator capitalizes on the existing standard statistical software, the computational burden is considerably less than that for the other bias correcting approaches, for example, calibration regression or corrected partial likelihood score. However, as our estimator was derived under the assumption of small and (conditionally) normal measurement error, one may need to develop more sophisticated methods, for example, likelihood based methods, for large amount of measurement error or for non-normal measurement error.

We have also considered the asymptotic behavior of the commonly used partial likelihood score test for testing no association between a failure outcome and an exposure in the presence of measurement error or mistiming error. In particular, we have derived the asymptotic distribution of the naive partial likelihood score test under a series of local alternatives, and calculated the asymptotic relative efficiency. As a result, we have obtained a sample size formula, easily implementable by practitioners when designing an observational study to compensate for the efficiency loss due to covariate measurement errors.

Appendix: Technical Details

A.1 Regularity Conditions and a Proof for Theorem 1

We assume the following regularity conditions.

(R.1) There exists a neighborhood $\mathcal{B}$ of β^* such that

$$\sup_{t \in [0, T_o], \beta \in \mathcal{B}} |\mathbf{S}^{(j)}(\beta, t) - \mathbf{s}^{(j)}(\beta, t)| \xrightarrow{p} 0.$$

Assume $\mathbf{s}^{(j)}(\beta, t)$, $j = 0, 1$, are bounded on $\mathcal{B} \times [0, T_o]$, and $s^{(0)}(\beta, t)$ is bounded away from 0 on $\mathcal{B} \times [0, T_o]$.

(R.2) The negative second-order derivative of $\mathbf{h}(\beta)$,

$$\mathbf{A}(\beta) = \int_0^{T_o} \left[\frac{\mathbf{s}^{(2)}(\beta, t)}{s^{(0)}(\beta, t)} - \left\{ \frac{\mathbf{s}^{(1)}(\beta, t)}{s^{(0)}(\beta, t)} \right\}^{\otimes 2} \right] s^{(0)}(t) dt$$

is positive definite at β^* .

Using the techniques of Lemma 3.1 of Andersen and Gill (1982), $\ell(\beta)$ can be shown asymptotically equivalent to

$$H(\beta) = \int_0^{T_o} \beta' \mathbf{s}^{(1)}(t) dt - \int_0^{T_o} \log\{s^{(0)}(\beta, t)\} s^{(0)}(t) dt.$$

Hence, by Theorem 2.1 of Struthers and Kalbfleisch (1986), we immediately have this consistency result.

A.2 Asymptotic Normality of Naive Estimates

Theorem 5

$$m^{1/2}(\hat{\beta}_{naive} - \beta^*) \xrightarrow{d} N\{0, \mathbf{A}^{-1}(\beta^*) \mathbf{B}(\beta^*) \mathbf{A}^{-1}(\beta^*)\}$$

where

$$\mathbf{B}(\beta^*) = E \left[\int_0^{T_o} \left\{ \mathbf{W}_{i*} - \frac{\mathbf{s}^{(1)}(\beta^*, t)}{s^{(0)}(\beta^*, t)} \right\} dN_i(t) - \int_0^{T_o} \frac{Y_i(t) \exp(\beta^* \mathbf{W}_{i*})}{s^{(0)}(\beta^*, t)} \left\{ \mathbf{W}_{i*} - \frac{\mathbf{s}^{(1)}(\beta^*, t)}{s^{(0)}(\beta^*, t)} \right\} s^{(0)}(t) dt \right]^2.$$

Moreover, $\mathbf{A}(\beta^*)$ and $\mathbf{B}(\beta^*)$ can be consistently estimated by $\hat{\mathbf{V}}(\hat{\beta}_{naive})$ and $\hat{\mathbf{B}}(\hat{\beta}_{naive})$, respectively. Here,

$$\begin{aligned}\hat{\mathbf{V}}(\beta) &= \frac{1}{m} \sum_{i=1}^m \int_0^{T_o} \left[\frac{\mathbf{S}^{(2)}(\beta, t)}{S^{(0)}(\beta, t)} - \left\{ \frac{\mathbf{S}^{(1)}(\beta, t)}{S^{(0)}(\beta, t)} \right\}^{\otimes 2} \right] dN_i(t), \\ \mathbf{G}_i(\beta) &= \int_0^{T_o} \left\{ \mathbf{W}_{i*} - \frac{\mathbf{S}^{(1)}(\beta, t)}{S^{(0)}(\beta, t)} \right\} dN_i(t) - \sum_{j=1}^m \int_0^{T_o} \frac{Y_j(t) \exp(\beta' \mathbf{W}_{j*})}{m S^{(0)}(\beta, t)} \left\{ \mathbf{W}_{j*} - \frac{\mathbf{S}^{(1)}(\beta, t)}{S^{(0)}(\beta, t)} \right\} dN_j(t), \\ \text{and } \hat{\mathbf{B}}(\beta) &= m^{-1} \sum_{i=1}^m \mathbf{G}_i(\beta)^{\otimes 2}.\end{aligned}$$

Thus, a consistent variance estimator of $\hat{\beta}_{naive}$ is

$$\widehat{\text{Var}}(\hat{\beta}_{naive}) = \frac{1}{m} \hat{\mathbf{V}}^{-1}(\hat{\beta}_{naive}) \hat{\mathbf{B}}(\hat{\beta}_{naive}) \hat{\mathbf{V}}^{-1}(\hat{\beta}_{naive}). \quad (29)$$

Algebraically, this type of variance estimators is equivalent to the approximate robust jackknife variance, and related statistical software is available (see Therneau and Grambsch, 2000). The above theorem suggests that even in the presence of covariate measurement errors, it is possible to construct confidence intervals based on the biased estimates using asymptotic normality and estimation of the asymptotic variance.

A.3 Additional Regularity Conditions for Bias Analysis (Theorem 3) and Lemmas

We postulate the same regularity conditions as in Bretagnolle and Huber-Carol (1988) for Theorem 3 (except for (C.0)).

- (C.0) Assume the measurement error variance-covariance matrix $\mathbf{D}$ is a diagonal matrix and the scale measurement error variance τ has a moment generating function $M_\tau(v, \sigma^2)$ such that $\frac{\partial^2}{\partial v \partial \sigma^2} \log M_\tau(v, \sigma^2) > 0$ when $v, \sigma^2 \geq 0$.
- (C.1) The additional covariate vector $\mathbf{Z} = (Z_1, \dots, Z_l)$ are ‘pertinent’ in model (8) in that the true effect of each component, say, $\beta_{jz}^{(0)}, j = 1, \dots, l$, is non-zero.
- (C.2) $\mathbf{Z}$ has time independent components with a finite moment generating function and all the components of $\mathbf{X}_* = (\mathbf{X}, \mathbf{Z})$ are independent and are independent of measurement error u .
- (C.3) there is no proper linear subspace including $\mathbf{X}$ and $\mathbf{Z}$ almost surely.

(C.4) $P(C \wedge T_o < D | \mathbf{X}, \mathbf{Z}) < 1$ almost surely to guarantee observability of covariates.

(C.5) the survival function of the censoring does not depend on $\mathbf{X}$ and $\mathbf{Z}$.

Remark: As indicated by the following Lemma, (C.0) is not restrictive and is satisfied by the nonnegative random variables whose distributions are in the exponential distribution family, examples including Bernoulli: $M_\tau(v, \sigma^2) = (1 - \sigma^2) + \sigma^2 e^v$; Poisson: $e^{\sigma^2 \{ \exp(v) - 1 \}}$; Chi-square: $(1 - 2v)^{-1/\sigma^2/2}, v < 1/2$; and exponential: $(1 - \sigma^2 v)^{-1}, v < 1/\sigma^2$. The assumptions is also satisfied for the random uniform, e.g. $\tau \sim U[0, 2\sigma^2]$, whose moment generating function is $M_\tau(v, \sigma^2) = (e^{2v\sigma^2} - 1)/2v\sigma^2$. (C.1) is a technical assumption to guarantee the information matrix is of DP+ type, which will be defined later. (C.2) can be satisfied as long as the covariates $\mathbf{Z}$ can be controlled. (C.3) excludes collinearity among covariates. (C.4) excludes the cases where failures can only be observed among a subset of covariate values. (C.5) is also a technical assumption, which can be tested, via the Kaplan-Meier estimate, by whether the distribution of C depends on the observable covariates W and $\mathbf{Z}$. As indicated by Bretagnolle and Huber-Carol (1988), it is possible to construct a counter example for zero-crossing estimation, i.e. a positive covariate effect, through the biased estimate, results in a negative one, when the censoring distribution depends on the studied covariates, $\mathbf{X}$ and $\mathbf{Z}$.

Lemma 1 *Let τ be any arbitrary nonnegative random variable whose distribution is in the exponential family and $M_\tau(v, \mu) = E(e^{v\tau})$ its moment generating function, where $\mu = E(\tau)$. Then*

$$\frac{\partial^2}{\partial v \partial \mu} \log M_\tau(v, \mu) > 0.$$

Proof: The probability density function (if τ is continuous) or the probability mass function (if τ is discrete) in the exponential family can be parameterized by

$$f(\tau; \theta, \phi) = \exp \left\{ \frac{\tau\theta - b(\theta)}{a(\phi)} + c(\tau, \phi) \right\}$$

where $\mu = E(\tau) = b'(\theta)$ and $\text{Var}(\tau) = a(\phi)b''(\theta)$ (Nelder and McCullagh, 1989).

Direct calculations give that

$$\frac{\partial^2}{\partial v \partial \mu} M_\tau(v, \mu) = \frac{E(\tau e^{v\tau} S_\mu) E(e^{v\tau}) - E(\tau e^{v\tau}) E(e^{v\tau} S_\mu)}{\{E(e^{v\tau})\}^2} \quad (30)$$

where $S_\mu = \frac{\partial}{\partial \mu} \log f(\tau; \theta, \phi)$ and the integrals involved in the expectations are taken with respect to the Lebesgue measure if τ is continuous or with respect to the counting measure if τ is discrete. By the chain rule

$$S_\mu = \frac{\partial}{\partial \theta} \log f(\tau; \theta, \phi) \frac{d\theta}{d\mu} = \frac{\tau - \mu}{\text{Var}(\tau)}.$$

Hence, substituting in S_μ yields

$$\frac{E(\tau^i e^{v\tau} S_\mu)}{E(\tau^i e^{v\tau})} = \left\{ \frac{E(\tau^{i+1} e^{v\tau})}{E(\tau^i e^{v\tau})} - \mu \right\} \frac{1}{\text{Var}(\tau)} \quad (31)$$

for $i = 0, 1$. Additionally, an application of the Schwarz inequality gives that

$$E(e^{v\tau})E(\tau^2 e^{v\tau}) > \{E(\tau e^{v\tau})\}^2. \quad (32)$$

Combining (31) and (32), we have that

$$\frac{E(\tau e^{v\tau} S_\mu)}{E(\tau e^{v\tau})} > \frac{E(e^{v\tau} S_\mu)}{E(e^{v\tau})},$$

which implies the numerator in (30) is positive and finishes the proof. $\square$

We state below a technical lemma, which will be useful in the proof of Theorem 3.

Lemma 2 *A symmetric square matrix M is called DP+ type if it is positive definite and its off-diagonal elements are all negative. Then each element of the inverse of a DP+ type matrix is positive.*

Proof: See Bretagnolle and Huber-Carol (1988).

A.4 Proof of Theorem 3

Proof: If $\beta_x^{(0)} = \mathbf{0}$, the naive model (4) correctly specifies the hazard function. Hence for any $\sigma^2 \geq 0$, the limit of the naive estimates $\beta_x^*(\sigma^2) = 0$ and $\beta_{jz}^*(\sigma^2) = \beta_{jz}^{(0)}$ for $1 \leq j \leq l$.

Without loss of generality, we consider the situation when $\beta_{1x}^{(0)} \neq 0$. We may assume the other components of β_0 are non-negative. Otherwise, for instance, if there exists a j such that $\beta_{jz}^{(0)} < 0$, one can always reverse the sign of the corresponding covariate Z_j .

Denote the right hand side of (9) by $\mathbf{S}(\beta)$, where $\beta = (\beta'_x, \beta'_z) = (\beta_{1x}, \dots, \beta_{l_1x}, \beta_{1z}, \dots, \beta_{l_2z})$. Denote by $\mathbf{X}_{-1} = (X_2, \dots, X_{l_1})'$ and $\beta_{-1,x} = (\beta_{2x}, \dots, \beta_{l_1x})'$ and consider $S_{1x}(0, \beta_{-1,x}, \beta_z)$, the first component of $\mathbf{S}(\beta)$ when $\beta_{1x} = 0$,

$$S_x(0, \beta_{-1,x}, \beta_z) = \int_0^{T_0} E[\{X_1 - E(X_1 e^{\beta'_{-1,x} \mathbf{X}_{-1} + \beta'_z \mathbf{Z}} CG) / E(e^{\beta'_{-1,x} \mathbf{X}_{-1} + \beta'_z \mathbf{Z}} CG)\} gC] dt \quad (33)$$

We adopt a symmetrization technique [see, e.g. Bretagnolle and Huber-Carol (1988)] to show that $S_x(0, \beta_{-1,x}, \beta_z) > 0$. We first show for each t , $E(Xe^{\beta'_{-1,x}\mathbf{X}_{-1} + \beta'_z\mathbf{Z}}CG) - E(e^{\beta'_{-1,x}\mathbf{X}_{-1} + \beta'_z\mathbf{Z}}CG)E(X) \leq 0$. Let $X''_1, \mathbf{X}''_{-1}, \mathbf{Z}''$ be independent copy of $X_1, \mathbf{X}_{-1}$ and $\mathbf{Z}$, and C'' and G'' the same function as C and G but taken at $\mathbf{X}'', \mathbf{Z}''$. By condition (C.5), C does not depend on $\mathbf{X}$ and $\mathbf{Z}$, which implies $C = C''$. Hence,

$$\begin{aligned} & E(X_1 e^{\beta'_{-1,x}\mathbf{X}_{-1} + \beta'_z\mathbf{Z}}CG) - E(e^{\beta'_{-1,x}\mathbf{X}_{-1} + \beta'_z\mathbf{Z}}CG)E(X_1) \\ &= \frac{1}{2}E\{(X_1 - X''_1)(e^{\beta'_{-1,x}\mathbf{X}_{-1} + \beta'_z\mathbf{Z}}G - e^{\beta'_{-1,x}\mathbf{X}''_{-1} + \beta'_z\mathbf{Z}''}G'')C\}. \end{aligned} \quad (34)$$

We only prove the situation when $\beta_{1x}^{(0)} > 0$. The proof will apply exactly for the situation when $\beta_{1x}^{(0)} < 0$.

When $\beta_{1x}^{(0)} > 0$, $e^{\beta'_{-1,x}\mathbf{X}_{-1} + \beta'_z\mathbf{Z}}G$ is decreasing in X_1 , so is its expectation conditional on X_1 . Hence, the conditional integrand on the right side of (34) is non-positive and, therefore, (34) is non-positive.

Now we want to prove

$$\psi = \int_0^{T_o} E[\{X_1 - E(X_1)\}gC]dt = - \int_0^{T_o} E[\{X_1 - E(X_1)\}CdG(t)] > 0. \quad (35)$$

In fact, integrating by parts and noticing that $C(0)G(0) \equiv 1$, we have

$$\psi = \int_0^{T_o} E[\{X - E(X)\}GdC(t)].$$

Again using the symmetrization technique, for each t , we have

$$2E[\{X - E(X)\}GdC(t)] = E[(X_1 - X''_1)(G'' - G)dC(t)] \geq 0.$$

This follows because G is decreasing in X_1 and so is its conditional expectation on X_1 . Also $C(t)$ is decreasing in t , Hence, the integrand in (35) is non-negative. By continuity, we only need to show the integrand in (35) is strictly positive at some point t . In fact, using the same symmetrization technique and conditions (C.3) and (C.4), one can show the integrand in (35) is strictly positive at $t = 0$. Hence, we have that $S_x(0, \beta_{-1,x}, \beta_z) > 0$ for any $\beta_{-1,x}$ and β_z .

Now write

$$Q(a, t) = \frac{E\{X_1 e^{aX_1 + \beta'_{-1,x}\mathbf{X}_{-1} + \beta'_z\mathbf{Z}}CG\}}{E\{e^{aX_1 + \beta'_{-1,x}\mathbf{X}_{-1} + \beta'_z\mathbf{Z}}CG\}}.$$

For each $t \geq 0$, one can show by using the Cauchy- Schwarz inequality,

$$\frac{\partial}{\partial a} Q(a, t) > 0.$$

Some algebra gives

$$\partial S_{1x}(\beta_{1x}, \beta_{-1,x}, \beta_z) / \partial \beta_{1x} = - \int_0^{T_o} E(gC) \left\{ \frac{\partial}{\partial \beta_{1x}} Q(\beta_{1x}, t) + q(\beta_{1x}, \sigma^2) \right\} dt$$

where $q(\beta_{1x}, \sigma^2) = p_1(\frac{1}{2}\beta'_x \mathbf{D}\beta_x, \sigma^2) D_{11}^2 \beta_{1x}^2 + p(\frac{1}{2}\beta'_x \mathbf{D}\beta_x, \sigma^2)$, $p_1(v, \sigma^2) = \frac{\partial}{\partial v} p(v, \sigma^2) = \frac{\partial^2}{\partial v^2} \log M_\tau(v, \sigma^2)$.

Since τ is a nonnegative random variable, $p(v, \sigma^2) \geq 0$ and by the convexity of $\log M_\tau(v, \sigma^2)$ with respect to v , $p_1(v, \sigma^2) \geq 0$. Therefore, $q(\beta_{1x}, \sigma^2) > 0$. Hence, $\partial S_{1x}(\beta_x, \beta_{-1,x}, \beta_z) / \partial \beta_{1x} < 0$, indicating $S_{1x}(\beta_{1x}, \beta_{-1,x}, \beta_z)$ is a decreasing function of β_{1x} for any fixed $\beta_{-1,x}, \beta_z$. As $S(0, \beta_{-1,x}, \beta_z) > 0$, the solution to $S_x(\beta, \beta_{-1,x}, \beta_z) = 0$ for a fixed β_z will be strictly positive. In particular, $\beta_{1x}^*(\sigma^2) > 0$. With exactly the same argument, one can show $\beta_{jz}^*(\sigma^2) > 0$ for all $1 \leq j \leq l$.

We now prove the monotonicity with respect to σ^2 . Differentiating (9) with respect to σ^2 on the both sides and collecting terms give

$$\frac{\partial}{\partial \sigma^2} \beta^*(\sigma^2) = - \left[p_2 \left\{ \frac{1}{2} (\beta_x^* \mathbf{D} \beta_x^*, \sigma^2) \right\} \int_0^{T_o} E(gC) dt \right] \mathbf{V}^{-1}(\beta^*, \sigma^2) \mathbf{F} \mathbf{D} \beta_x^* \quad (36)$$

where we abbreviate $\beta^*(\sigma^2)$ by $\beta^* = (\beta_x^*, \beta_z^*)$ and

$$p_2 \left\{ \frac{1}{2} \beta_x^* \mathbf{D} \beta_x^*, \sigma^2 \right\} \stackrel{def}{=} \frac{\partial}{\partial \sigma^2} p \left\{ \frac{1}{2} \beta_x^* \mathbf{D} \beta_x^*, \sigma^2 \right\} = \frac{\partial^2}{\partial v \partial \sigma^2} \log M_\tau \left(\frac{1}{2} \beta_x^* \mathbf{D} \beta_x^*, \sigma^2 \right) > 0$$

by the assumption and $\mathbf{V}(\beta^*, \sigma^2)$ is defined in (13). Note that (36) holds even without the independence assumptions on the components of the covariate vector and the measurement error.

Hence, for any $\mathbf{x} = (x_1, \dots, x_{l_1+l_2})' \in R^{l_1+l_2}$ such that $\mathbf{x} \neq \mathbf{0}$,

$$\begin{aligned} & \mathbf{x}' \mathbf{V}(\beta^*, \sigma^2) \mathbf{x} \\ &= E(GC e^{\beta^* \mathbf{X}_*} [(\mathbf{x}' \mathbf{X}_*)^2 + p_1(\frac{1}{2} \beta_x^* \mathbf{D} \beta_x^*, \sigma^2) \|\mathbf{x}' \mathbf{F} \mathbf{D} \beta_x^*\|_2^2 + p(\frac{1}{2} \beta_x^* \mathbf{D} \beta_x^*, \sigma^2) \|\mathbf{x}' \mathbf{F} \mathbf{D}^{1/2}\|_2^2]) E(GC e^{\beta^* \mathbf{X}_*}) \\ & \quad - \{E(GC e^{\beta^* \mathbf{X}_*} \mathbf{x}' \mathbf{X}_*)\}^2 \\ &> E(GC e^{\beta^* \mathbf{X}_*} (\mathbf{x}' \mathbf{X}_*)^2) E(GC e^{\beta^* \mathbf{X}_*}) - \{E(GC e^{\beta^* \mathbf{X}_*} \mathbf{x}' \mathbf{X}_*)\}^2 \geq 0, \end{aligned}$$

where the last inequality is by the Cauchy-Schwarz inequality. Here for a column vector $\mathbf{a}$, $\|\mathbf{a}\|_2^2 = \mathbf{a}' \mathbf{a}$. Hence $\mathbf{V}(\beta^*, \sigma^2)$ is positive. Note, the off-diagonal element of $\mathbf{V}(\beta^*, \sigma^2)$ is

$$V_{s_1, s_2} = E(GC e^{\beta^* \mathbf{X}_*} X_{s_1}^* X_{s_2}^*) E(GC e^{\beta^* \mathbf{X}_*}) - E(GC e^{\beta^* \mathbf{X}_*} X_{s_1}^*) E(GC e^{\beta^* \mathbf{X}_*} X_{s_2}^*)$$

where $1 \leq s_1 \neq s_2 \leq l+1$, $X_{s_1}^*$ and $X_{s_2}^*$ are the s_1 -th and s_2 -th elements of $\mathbf{X}_*$ respectively. Again, using the symmetrization technique and the assumption that $\beta_x^{(0)} > 0$ and $\beta_{jz}^{(0)} > 0$, $1 \leq j \leq l$, one can show $V_{s_1, s_2} < 0$ for $1 \leq s_1 \neq s_2 \leq l+1$. Hence, $\mathbf{V}(\beta^*, \sigma^2)$ is of DP+ type. By lemma 2, $\mathbf{V}^{-1}(\beta^*, \sigma^2)$ is a matrix with all its elements being positive and the theorem follows immediately. $\square$

A.5 Bias Correcting Estimator

When $\sigma^2 = 0$ is small, expanding $\beta^*(\sigma^2)$ around $\sigma^2 = 0$ yields

$$\beta^*(\sigma^2) = \beta_0 + \sigma^2 \frac{\partial}{\partial \sigma^2} \beta^*(\sigma^2)|_{\sigma^2=0} + o(\sigma^2)$$

where $\frac{\partial}{\partial \sigma^2} \beta^*(\sigma^2)|_{\sigma^2=0}$ is given by (36) evaluated at $\sigma = 0$. Direct calculation gives $p_2(v, 0) \equiv 1$ for $v \geq 0$. Hence, (12), and furthermore, (14) hold.

By martingale theory and the law of large numbers, we have

$$\bar{N} \stackrel{def}{=} \frac{1}{m} \sum_i^m \int_0^{T_o} dN_i(t) \rightarrow \int_0^{T_o} E(gC) dt$$

in probability. Using the fact that $p_1(v, 0) \equiv 0$ and $p(v, 0) \equiv 0$ for $v \geq 0$, hence, $q\{\frac{1}{2}\beta_x^{(0)} \mathbf{D}\beta_x^{(0)}, 0\} = 0$, one may rewrite $\mathbf{V}(\beta_0, 0)$ defined by (13) as

$$\mathbf{V}_0 = \int_0^{T_o} \left[\frac{E(GC e^{\beta_0 \mathbf{X}_*} \mathbf{X}_*^* (\mathbf{X}_*)')}{E(GC e^{\beta_0 \mathbf{X}_*})} - \left\{ \frac{E(GC \mathbf{X}_* e^{\beta_0 \mathbf{X}_*})}{E(GC e^{\beta_0 \mathbf{X}_*})} \right\}^{\otimes 2} \right] E(gC) dt.$$

Additionally, denote by $S^{(j)}(\beta, t) = m^{-1} \sum_i \mathbf{W}_{i*}^j Y_i(t) \exp(\beta \mathbf{W}_{i*})$, $j = 0, 1, 2$, and $M_\tau = M_\tau\{\frac{1}{2}\beta_x^{(0)} \mathbf{D}\beta_x^{(0)}, \sigma^2\}$, $M_\tau^{(1)} = \frac{\partial}{\partial \beta_x} M_\tau(\frac{1}{2}\beta_x' \mathbf{D}\beta_x, \sigma^2)|_{\beta_x = \beta_x^{(0)}}$ and $M_\tau^{(2)} = \frac{\partial^2}{\partial \beta_x \partial \beta_x'} M_\tau(\frac{1}{2}\beta_x' \mathbf{D}\beta_x, \sigma^2)|_{\beta_x = \beta_x^{(0)}}$. Then, under model (5), applying the double expectation theorem, one has

$$\begin{aligned} s^{(0)}(\beta_0, t) &= E(Y e^{\beta_0 \mathbf{W}^*}) = E(GC e^{\beta_0 \mathbf{X}_*}) M_\tau \\ s^{(1)}(\beta_0, t) &= E(Y \mathbf{W}^* e^{\beta_0 \mathbf{W}^*}) = E(GC \mathbf{X}_* e^{\beta_0 \mathbf{X}_*}) M_\tau + E(GC e^{\beta_0 \mathbf{X}_*}) \mathbf{F} M_\tau^{(1)} \\ s^{(2)}(\beta_0, t) &= E\{Y \mathbf{W}^* (\mathbf{W}^*)' e^{\beta_0 \mathbf{W}^*}\} = E\{GC \mathbf{X}_* (\mathbf{X}_*)' e^{\beta_0 \mathbf{X}_*}\} M_\tau + M_\tau^{(1)} E(GC (\mathbf{X}_*)' e^{\beta_0 \mathbf{X}_*}) \otimes \mathbf{F} \\ &\quad + E(GC e^{\beta_0 \mathbf{X}_*}) \mathbf{F} M_\tau^{(2)} \mathbf{F}', \end{aligned}$$

where for two square matrices $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{A} \otimes \mathbf{B} = \mathbf{A}' \mathbf{B}' + \mathbf{B} \mathbf{A}$.

Therefore, with regularity condition (R.1), one obtains that, on a finite interval $[0, T_0]$,

$$\begin{aligned}\tilde{S}^{(0)}(\beta_0, t) &\stackrel{def}{=} M_\tau^{-1} S^{(0)}(\beta_0, t) \\ &\rightarrow M_\tau^{-1} s^{(0)}(\beta_0, t) = E(GC e^{\beta_0 \mathbf{X}_*})\end{aligned}$$

$$\begin{aligned}\tilde{\mathbf{S}}^{(1)}(\beta_0, t) &\stackrel{def}{=} M_\tau^{-1} \mathbf{S}^{(1)}(\beta_0, t) - M_\tau^{-2} S^{(0)}(\beta_0, t) \mathbf{F} M_\tau^{(1)} \\ &\rightarrow M_\tau^{-1} \mathbf{s}^{(1)}(\beta_0, t) - M_\tau^{-2} s^{(0)}(\beta_0, t) \mathbf{F} M_\tau^{(1)} = E(GC X^* e^{\beta_0 \mathbf{X}_*})\end{aligned}$$

and

$$\begin{aligned}\tilde{\mathbf{S}}^{(2)}(\beta_0, t) &\stackrel{def}{=} M_\tau^{-1} \mathbf{S}^{(2)}(\beta_0, t) - M_\tau^{-2} M_\tau^{(1)} (\mathbf{S}^{(1)}(\beta_0, t))' \otimes \mathbf{F} \\ &\quad + S^{(0)}(\beta_0, t) \mathbf{F} [2M_\tau^{-3} M_\tau^{(1)} \{M_\tau^{(1)}\}' - M_\tau^{-2} M_\tau^{(2)}] \mathbf{F}' \\ &\rightarrow M_\tau^{-1} \mathbf{s}^{(2)}(\beta_0, t) - M_\tau^{-2} M_\tau^{(1)} (\mathbf{s}^{(1)}(\beta_0, t))' \otimes \mathbf{F} \\ &\quad + s^{(0)}(\beta_0, t) \mathbf{F} [2M_\tau^{-3} M_\tau^{(1)} \{M_\tau^{(1)}\}' - M_\tau^{-2} M_\tau^{(2)}] \mathbf{F}' \\ &= E\{GC e^{\beta_0 \mathbf{X}_*} \mathbf{X}_* (\mathbf{X}_*)'\}\end{aligned}$$

in probability uniformly in a finite interval $[0, T_0]$. Hence, with similar arguments in the proof of Lemma 3.1 in Andersen and Gill (1982), one can show

$$\hat{\mathbf{V}}(\beta_0) \stackrel{def}{=} \int_0^{T_0} \left[\frac{\tilde{\mathbf{S}}^{(2)}(\beta_0, t)}{\tilde{S}^{(0)}(\beta_0, t)} - \left\{ \frac{\tilde{\mathbf{S}}^{(1)}(\beta_0, t)}{\tilde{S}^{(0)}(\beta_0, t)} \right\}^{\otimes 2} \right] \frac{1}{m} \sum_{i=1}^m dN_i(t) \rightarrow \mathbf{V}_0$$

in probability. As $\hat{\mathbf{V}}(\beta)$ is continuous in β , when σ^2 is relatively small, $\hat{\mathbf{V}}(\beta_0)$ can be well approximated by $\hat{\mathbf{V}}(\hat{\beta}_{naive})$, where $\hat{\beta}_{naive}$ is the naive partial likelihood estimate. Thus, a first order approximation to (11) is

$$\tilde{\beta} = \left\{ \mathbf{I} - \sigma^2 \bar{N} \hat{\mathbf{V}}^{-1}(\hat{\beta}_{naive}) \mathbf{F} \mathbf{D} \mathbf{F}' \right\}^{-1} \hat{\beta}_{naive},$$

where $\mathbf{I}$ is an $(l_1 + l_2) \times (l_1 + l_2)$ identity matrix. Note that the above formula is indeed applicable in much more general situations as the derivation of (36) holds without the restrictive assumptions on the covariates and the measurement error required for the theoretical bias analyses.

When σ^2 is small, $M_\tau = 1 + O(\sigma^2)$, $M_\tau^{(1)} = O(\sigma^2)$, $M_\tau^{(2)} = O(\sigma^2)$. Therefore, $\hat{\mathbf{V}}(\hat{\beta}_{naive})$ can be further approximated by

$$\hat{\mathbf{V}} = \int_0^{T_0} \left[\frac{\mathbf{S}^{(2)}(\hat{\beta}_{naive}, t)}{S^{(0)}(\hat{\beta}_{naive}, t)} - \left\{ \frac{\mathbf{S}^{(1)}(\hat{\beta}_{naive}, t)}{S^{(0)}(\hat{\beta}_{naive}, t)} \right\}^{\otimes 2} \right] \frac{1}{m} \sum_{i=1}^m dN_i(t),$$

which results in (14).

A.6 Proof to Theorem 4

Proof: Define

$$A(t) = \frac{\sum_i^m Y_i(t)W_i}{\sum_i^m Y_i(t)}, \quad \tilde{A}(t) = \frac{\sum_i^m Y_i(t)W_i e^{\phi(\beta, W_i, \tau_i, t)}}{\sum_i^m Y_i(t) e^{\phi(\beta, W_i, \tau_i, t)}}.$$

Then

$$m^{-1/2}U = m^{-1/2} \sum_i \int_0^{T_o} \{W_i - \tilde{A}(t)\} dN_i(t) + m^{-1/2} \sum_i \int_0^{T_o} \{\tilde{A}(t) - A(t)\} dN_i(t). \quad (37)$$

Some algebra shows that the first term on the right side of (37) is equal to

$$m^{-1/2} \sum_i \int_0^{T_o} \{W_i - \tilde{A}(t)\} d\tilde{M}_i(t)$$

where $\tilde{M}_i(t) = N_i(t) - \int_0^t Y_i(u) \lambda_0(u) \exp\{\phi(\beta, W_i, \tau_i, u)\} du$ is the martingale with respect to the filtration $\mathcal{F}_t = \sigma\{N_i(u), Y_i(u+), W_i, \tau_i, i = 1, \dots, m, 0 \leq u < t\}$. Hence, by Theorem 4.2 of Andersen and Gill (1982), under the local alternative $\beta_m = m^{-1/2}\eta \rightarrow 0$, it converges weakly to a normal distribution with mean 0 and variance $\int_0^{T_o} \left[\frac{s^{(2)}(t)}{s^{(0)}(t)} - \left\{ \frac{s^{(1)}(t)}{s^{(0)}(t)} \right\}^2 \right] s^{(0)}(t) dt$, where the expectations involved are taken under $\beta = 0$.

Similarly, by a Taylor expansion about $\beta = 0$ and some empirical processes argument one may show

$$m^{1/2} \{\tilde{A}(t) - A(t)\} \rightarrow \eta \frac{s^{(1)}(t)s_{\beta}^{(0)}(t) - s^{(0)}(t)s_{\beta}^{(1)}(t)}{\{s^{(0)}(t)\}^2}$$

in probability uniformly on $[0, T_o]$, where the expectations involved are taken under $\beta = 0$ and the derivatives are evaluated at $\beta = 0$. Hence, it follows with the argument of Lemma 3.1 in Andersen and Gill (1982), the second term on the right side of (37) converges in probability to

$$\int_0^{T_o} \eta \frac{s^{(1)}(t)s_{\beta}^{(0)}(t) - s^{(0)}(t)s_{\beta}^{(1)}(t)}{s^{(0)}(t)} dt.$$

Then applying Slutsky's theorem yields the desired result immediately. $\square$

A.7 Proof to Theorem 5

Proof: Similar to the proof of Theorem 4, under the local alternative $\beta_m = m^{-1/2}\eta \rightarrow 0$, the general test statistic $m^{-1/2}\tilde{U}_m$ converges weakly to a normal distribution with mean

$$\int_0^{T_o} \eta r(t) \frac{s^{(1)}(t)s_{\beta}^{(0)}(t) - s^{(0)}(t)s_{\beta}^{(1)}(t)}{s^{(0)}(t)} dt.$$

and variance

$$\int_0^{T_o} r^2(t) \left[\frac{s^{(2)}(t)}{s^{(0)}(t)} - \left\{ \frac{s^{(1)}(t)}{s^{(0)}(t)} \right\}^2 \right] s^{(0)}(t) dt,$$

where the expectations involved are taken under $\beta = 0$ and the derivatives involved are evaluated at $\beta = 0$.

Under $\beta = 0$ and with the assumption that the censoring time C is independent of W , one obtains that

$$\begin{aligned} s^{(1)}(t) &\stackrel{def}{=} E\{Y(t)W\} = E\{Y(t)\}\mu, \\ s^{(2)}(t) &\stackrel{def}{=} E\{Y(t)W^2\} = E\{Y(t)\}\{\sigma_x^2 + \sigma^2 + \mu^2\}, \\ s_{\beta}^{(0)}(t) &\stackrel{def}{=} E\{Y(t)\phi_{\beta}\} = E\{Y(t)\}\{E(\rho^*)\mu + E(\alpha^*)\}, \\ s_{\beta}^{(1)}(t) &\stackrel{def}{=} E\{Y(t)W\phi_{\beta}\} = E\{Y(t)\}\{\sigma_x^2 + E(\rho^*)\mu^2 + E(W\alpha^*)\}. \end{aligned}$$

Thus, some algebra gives the efficacy of test (28)

$$\frac{\eta^2 \sigma_x^4}{(\sigma_x^2 + \sigma^2)^2} \frac{\left\{ \int_0^{T_o} r(t) s^{(0)}(t) dt \right\}^2}{\int_0^{T_o} r^2(t) s^{(0)}(t) dt}. \quad (38)$$

For fixed $\eta \neq 0$, $\sigma_x^2 > 0$ and σ^2 , by the Schwarz inequality, efficacy (38) achieves maximum when $r(t)$ is a nonzero constant function. That is, test (17) achieves maximal efficacy within the class of (28).

REFERENCES

- Andersen, P. K. and Gill, R. D. (1982), "Cox's regression model for counting processes: A large sample study (Com: p1121-1124)," *The Annals of Statistics*, 10, 1100-1120.
- Bretagnolle, J. and Huber-Carol, C. (1988), "Effects of omitting covariates in Cox's model for survival data," *Scandinavian Journal of Statistics*, 15, 125-138.
- Carroll, R. and Stefanski, L. (1990), " APPROXIMATE QUASI-LIKELIHOOD ESTIMATION IN MODELS WITH SURROGATE PREDICTORS," *J AM STAT ASSOC*, 85, 652-663.
- Carroll, R., Ruppert, D. and Stefanski, L.A. (1995), *Measurement Error in Nonlinear Models*, London: Chapman and Hall.
- Chen, H. Y. and Little, R. J. A. (1999), "Proportional hazards regression with missing covariates," *Journal of the American Statistical Association*, 94, 896-908.
- DiRienzo, A.G. and Lagakos, S.W. (2001), "Bias correction for score tests arising from mis-specified proportional hazards regression models," *Biometrika*, 88, 421-34.
- Clayton, D.G. and Cuzick, J. (1985), "Multivariate generalizations of the proportional hazards model (with Discussion)." *J. R. Statist. Soc. A*, 148, 82-117.
- Cox, D. R. (1972), "Regression models and life-tables (with discussion)," *Journal of the Royal Statistical Society, Series B, Methodological*, 34, 187-220.
- Fleming, T.R. and Harrington, D. P. (1991), *Counting Processes and Survival Analysis*, New York: Wiley.
- Fuller, W. (1987). *Measurement Error Models*. Wiley, New York.
- Higgins, K.M., Davidian, M. and Giltinan, D. M. (1997), "A two-step approach to measurement error in time-dependent covariates in nonlinear mixed-effects models, with application to IGF-I pharmacokinetics," *Journal of the American Statistical Association*, 92 ,436-448.
- Hu, C. and Lin, D.Y. (2002), "Cox regression with covariate measurement error," *Scandinavian*

Journal of Statistics, in press.

- Hu, P., Tsiatis, A. A. and Davidian, M. (1998), "Estimating the parameters in the Cox model when covariate variables are measured with error," *Biometrics*, 54, 1407-19.
- Huang, Y. and Wang, C.Y. (2000), "Cox regression with accurate covariates unascertainable: A nonparametric- correction approach," *Journal of the American Statistical Association*, 95, 1209-1219.
- Hughes, M.D. (1993), "Regression dilution in the proportional hazards model," *Biometrics*, 49, 1056-66.
- Kahn, J.O., Lagakos, S.W., et al. (1992), "A controlled trial comparing continued zidovudine with didanosine in human immunodeficiency virus infection," *The New England Journal of Medicine*, 327, 581-587.
- Keiding, N, "Independent delayed entry," *Survival Analysis: State of the Art*, 309-326.
- Kong, F. and Slud, E. (1997), "Robust covariate-adjusted logrank tests," *Biometrika*, 84, 847-862.
- Kong, F. (1999), "Adjusting regression attenuation in the Cox proportional hazards model," *J STAT PLAN INFER*, 79, 31-44.
- Lagakos, S. W. (1988), "The loss in efficiency from misspecifying covariates in proportional hazards regression models", *Biometrika*, 75, 156-160.
- Lin, D.Y. and Wei, L.J. (1989), "The robust inference for the Cox proportional hazards model," *Journal of the American Statistical Association*, 89, 1074-1078.
- Nakamura, T. (1992), "Proportional Hazards Model With Covariates Subject to Measurement Error," *Biometrics*, 48, 829-838.
- Nielsen, G.G, Gill, R.D., Andersen, P.K. and Sorensen, T.I.A. (1992), "A counting process approach to maximum likelihood estimation in frailty models," *Scand. J. Statistics*, 19, 25-43.
- Paik, M. C. and Tsai, W. (1997), "On using the Cox proportional hazards model with missing

- covariates,” *Biometrika*, 84, 579-593.
- Prentice, R. L. (1982), “Covariate measurement errors and parameter estimation in a failure time regression model,” *Biometrika*, 69, 331-42.
- Schmoor, C. and Schumacher, M. (1997), “Effects of covariate omission and categorization when analysing randomized trials with the Cox model”, *Statistics in Medicine*, 16 , 225-237.
- Schoenfeld, D. A. (1983), “Sample-size formula for the proportional-hazards regression model”, *Biometrics*, 39, 499-503.
- Struthers, C. A. and Kalbfleisch, J. D. (1986), “Misspecified proportional hazard models”, *Biometrika*, 73 , 363-369.
- Therneau, T. M. and Grambsch, P. M. (2000), *Modeling survival data: extending the Cox model*. New York: Springer-Verlag.
- Tsiatis, A. A. and Davidian, M. (2001), “A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error,” *Biometrika*, 88, 447-458.
- Wang, C.Y., Xie, S. X., and Prentice, R. L. (2001), “Recalibration based on an approximate relative risk estimator in Cox regression with missing covariates,” *Statistica Sinica*, 11, (4), 1081-1104.
- Wang, N., Lin, X., Gutierrez, R. and Carroll R.J. (1998), “Bias analysis and SIMEX approach in generalized linear mixed error models,” *J. Am. Statist. Assoc.* 93, 249-61.
- Xie, S. X. , Wang, C. Y. and Prentice, R. L. (2001), “A risk set calibration method for failure time regression by using a covariate reliability sample,” *Journal of the Royal Statistical Society, Series B, Methodological*, 63, (4) , 855-870.
- Zhong, M., Sen, P. K. and Cai, J. (1996), “Cox Regression Model With Mismeasured Covariates Or Missing Covariate Data,” *ASA Proceedings of the Biometrics Section*, 323-328.
- Zhou, H. and Pepe, M. S. (1995), “Auxiliary Covariate Data in Failure Time Regression,” *Biometrika*, 82, 139-149.

Table 1. *Simulation results of the measurement error model based on 1000 replicates (true $\beta_x = \beta_z = 1$, $\text{Var}(X) = \text{Var}(Z) = 1$), where Var_e and Var_b are empirical variance and bootstrap variance respectively, when the variance function $\sigma(t) = t$.*

m	σ^2	Parameter	Naive			Corrected			
			Estimate	Var_e	MSE	Estimate	Var_e	Var_b	MSE
80	0.05	β_x	0.9546	0.0372	0.039	1.0434	0.0535	0.0627	0.055
		β_z	1.0099	0.0400	0.040	1.0307	0.0444	0.0502	0.045
	0.15	β_x	0.8453	0.0299	0.054	1.0485	0.0772	0.1055	0.080
		β_z	0.9722	0.0381	0.039	1.0344	0.0537	0.0800	0.055
	0.25	β_x	0.7604	0.0301	0.088	1.0550	0.1701	0.2166	0.173
		β_z	0.9520	0.0377	0.040	1.0513	0.0720	0.1120	0.075
	0.4	β_x	0.6565	0.0241	0.142	1.0876	0.3526	0.4615	0.360
		β_z	0.9267	0.0389	0.044	1.0612	0.1613	0.2348	0.165
	0.5	β_x	0.5996	0.0227	0.183	1.1487	0.5032	0.6264	0.525
		β_z	0.9094	0.0364	0.045	1.1012	0.1412	0.2687	0.151
	0.9	β_x	0.4534	0.0187	0.317	1.3157	2.4759	2.6075	2.576
		β_z	0.8701	0.0347	0.052	1.1849	0.6018	0.8003	0.636
200	0.05	β_x	0.9412	0.0124	0.016	1.0187	0.0164	0.0164	0.017
		β_z	0.9923	0.0129	0.013	1.0187	0.0142	0.0146	0.015
	0.15	β_x	0.8269	0.0107	0.041	1.0305	0.0239	0.0258	0.025
		β_z	0.9572	0.0134	0.015	1.0239	0.0178	0.0176	0.018
	0.25	β_x	0.7382	0.0104	0.079	1.0383	0.0349	0.0403	0.036
		β_z	0.9244	0.0131	0.019	1.0220	0.0214	0.0217	0.022
	0.4	β_x	0.6355	0.0094	0.142	1.0570	0.0639	0.0893	0.067
		β_z	0.8977	0.0134	0.024	1.0303	0.0275	0.0323	0.028
	0.5	β_x	0.5840	0.0084	0.181	1.0719	0.0830	0.1230	0.088
		β_z	0.8777	0.0132	0.028	1.0327	0.0328	0.0436	0.034
	0.9	β_x	0.4380	0.0065	0.322	1.1598	0.3369	0.4777	0.362
		β_z	0.8410	0.0134	0.039	1.0644	0.0879	0.1399	0.092
500	0.05	β_x	0.9322	0.0050	0.010	1.0046	0.0064	0.0058	0.006
		β_z	0.9755	0.0054	0.006	0.9998	0.0059	0.0051	0.006
	0.15	β_x	0.8186	0.0043	0.037	1.0045	0.0088	0.0083	0.009
		β_z	0.9447	0.0053	0.008	1.0053	0.0068	0.0061	0.007
	0.25	β_x	0.7295	0.0043	0.078	1.0036	0.0136	0.0118	0.014
		β_z	0.9230	0.0049	0.011	1.0095	0.0074	0.0072	0.008
	0.4	β_x	0.6280	0.0032	0.142	0.9970	0.0170	0.0188	0.017
		β_z	0.8828	0.0049	0.019	0.9955	0.0089	0.0089	0.009
	0.5	β_x	0.5701	0.0033	0.188	0.9918	0.0240	0.0248	0.024
		β_z	0.8691	0.0051	0.022	0.9928	0.0108	0.0102	0.011
	0.9	β_x	0.4331	0.0025	0.324	1.0003	0.0535	0.0879	0.054
		β_z	0.8309	0.0050	0.034	0.9942	0.0162	0.0206	0.016

Figure 1: Simulated Data Examples: measurement error against the time lag under different measurement error variance functions. The time lag was generated from $U[0, 1]$, the average measurement error variance was set to be 1.0 and 100 data points were simulated under model (3).

Figure 2: Asymptotic relative biases in naive estimates against the correlation between the observed exposure and the true exposure. The variance function is $\sigma(t) = t$. The true parameter values are $\lambda_0 = 1, \beta_x^{(0)} = \beta_z^{(0)} = 1$. The four curves in each plot correspond to censoring proportions, 10%, 30%, 60% and 80% with the solid line for 10%, and the longest dashed line for 80%.

Figure 3: Performance of the corrected and naive estimators when the variance function is $\sigma(t) = t$.