Journal of Statistical Research *ISSN 0256 - 422 X*
2011, Vol. 45, No. 2, pp. 111-130
Bangladesh

ANALYSIS OF ERROR-CONTAMINATED SURVIVAL DATA UNDER THE PROPORTIONAL ODDS MODEL

WENQING HE

Department of Statistical and Actuarial Sciences
University of Western Ontario. London, Ontario, Canada

Email: whe@stats.uwo.ca

JUAN XIONG

Department of Statistical and Actuarial Sciences
University of Western Ontario, London, Ontario, Canada and
and Department of Statistics and Actuarial Science
University of Waterloo Waterloo, Ontario, Canada

Email: xiongjuan2000@gmail.com

GRACE Y. YI

Department of Statistics and Actuarial Science
University of Waterloo, Waterloo, Ontario, Canada

Email: yyi@uwaterloo.ca

SUMMARY

There has been extensive research interest in analysis of survival data with covariates subject to measurement error. The focus of the most discussions is on the proportional hazards (PH) model, although there are some work concerning the accelerated failure time (AFT) model and the additive hazards (AH) model. Relatively little attention has been directed to studying the impact of measurement error on other models, such as the proportional odds (PO) model. The proportional odds model is an important alternative when PH, AFT or AH models are not appropriate to fit data. In this paper we discuss two inference methods to accommodate measurement error effects under the PO model, in contrast to the naive analysis that ignores the covariate measurement error. Numerical studies are conducted to evaluate the performance of the proposed methods.

Keywords and phrases: Censoring; Likelihood; Measurement error; Regression calibration; Proportional odds model; Survival data.

1 Introduction

Covariate measurement error is a typical feature that is present with survival data. There has been extensive research on addressing covariate measurement error in survival analysis, where the focus is on proportional hazards (PH) models. Prentice (1982) pioneered the so-called regression calibration method to correct for measurement error effects under the rare event assumption. A large body of various inference methods have since then been developed by many authors, including Nakamura (1992), Hu, Tsiatis and Davidian (1998), Zhou and Wang (2000), Huang and Wang (2000), Xie, Wang and Prentice (2001), Song and Huang (2005), Li and Ryan (2004, 2006), and Yi and Lawless (2007), among many others.

Recently, research attention has been extended to alternative models when the proportionality assumption attached to the PH model fails. For instance, Giménez, Bolfarine and Colosimo (1999, 2006), Yi and He (2006), and He, Yi and Xiong (2007), among others, considered covariate measurement error problems under the accelerated failure time (AFT) models; while with additive hazards (AH) regression Kulich and Lin (2000), Sun, Zhang and Sun (2006), and Sun and Zhou (2008) developed inference methods to accommodate covariate measurement error.

Relatively little attention has been directed to studying the impact of measurement error on other models, such as the proportional odds (PO) model (Bennett 1983). The proportional odds model is an important alternative when PH, AFT or AH models are not appropriate to fit data (Yang and Prentice 1999). In this paper, in contrast to the naive analysis that ignores the covariate measurement error, we discuss two inference methods to accommodate measurement error effects under the PO model.

The remainder is organized as follows. Notation and model formulation are introduced in Section 2. In Section 3 we describe two inference methods to account for measurement error in covariates. The performance of the proposed methods is assessed empirically in Section 4. General discussion is included in Section 5.

2 Notation and Model Formulation

For the survival process, let T_i and C_i be the survival and censoring times for subject i, respectively, and δ_i be the censoring indicator variable taking 1 if $T_i \leq C_i$ and 0 otherwise, $i = 1, 2, ..., n$. Let $Y_i = \min(T_i, C_i)$ for $i = 1, ..., n$. Let $\mathbf{X}_i$ be the vector of covariates subject to possible measurement error, and $\mathbf{Z}_i$ be the vector of covariates free of error. As a common practice, independent censoring is assumed here. That is, conditional on covariates, T_i and C_i are independent.

2.1 Proportional Odds Model

Let $F(t)$ be the distribution function of T_i. Response variable T_i is characterized by the proportional odds model, given by

$$\frac{F(T_i)}{1 - F(T_i)} = \frac{F_0(T_i)}{1 - F_0(T_i)} \exp\{\boldsymbol{\beta}_x^{\mathrm{T}} \mathbf{X}_i + \boldsymbol{\beta}_z^{\mathrm{T}} \mathbf{Z}_i\}, \tag{2.1}$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}_x^{\mathrm{T}}, \boldsymbol{\beta}_z^{\mathrm{T}})^{\mathrm{T}}$ is the vector of regression parameters, and $F_0(t)$ represents the baseline distribution function that is known up to unknown parameter $\boldsymbol{\alpha}$, say. If $R_0(t) = F_0(t)/(1 - F_0(t))$, then model (2.1) becomes

$$F(T_i) = \frac{R_0(T_i)\exp\{\boldsymbol{\beta}_x^{\mathrm{T}}\mathbf{X}_i + \boldsymbol{\beta}_z^{\mathrm{T}}\mathbf{Z}_i\}}{1 + R_0(T_i)\exp\{\boldsymbol{\beta}_x^{\mathrm{T}}\mathbf{X}_i + \boldsymbol{\beta}_z^{\mathrm{T}}\mathbf{Z}_i\}}, \tag{2.2}$$

yielding, equivalently, the density function

$$f(T_i) = \frac{r_0(T_i)\exp\{\boldsymbol{\beta}_x^{\mathrm{T}}\mathbf{X}_i + \boldsymbol{\beta}_z^{\mathrm{T}}\mathbf{Z}_i\}}{[1 + R_0(T_i)\exp\{\boldsymbol{\beta}_x^{\mathrm{T}}\mathbf{X}_i + \boldsymbol{\beta}_z^{\mathrm{T}}\mathbf{Z}_i\}]^2}, \tag{2.3}$$

where $r_0(t) = R_0'(t) = f_0(t)/\{1 - F_0(t)\}^2$ with $f_0(t) = F_0'(t)$.

2.2 Measurement Error Model

Let $\mathbf{X}_i^*$ be an observed version of covariate $\mathbf{X}_i$. $\mathbf{X}_i$ and $\mathbf{X}_i^*$ are assumed to follow a classical additive measurement error model, a model which is perhaps the most widely used in practice (e.g., Li and Lin 2003; Greene and Cai 2004; Carroll et al. 2006). That is, conditional on covariates $\mathbf{X}_i$ and $\mathbf{Z}_i$, response T_i and censoring time C_i,

$$\mathbf{X}_i^* = \mathbf{X}_i + \mathbf{e}_i, \tag{2.4}$$

where $\mathbf{e}_i$ follows, independent of other variables, a normal distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}_e = [\sigma_{jk}]$. To emphasize estimation on the response parameters which are of prime interest, here we treat the parameters in $\boldsymbol{\Sigma}_e$ as known or estimated from an independent sample. Discussion on this treatment is given in Section 5.

2.3 Covariate Process

Assume that $\mathbf{X}_i$ is modeled by the linear mixed model

$$\mathbf{X}_i = \boldsymbol{\gamma}\mathbf{Z}_i + \mathbf{S}_i\mathbf{u}_i, \tag{2.5}$$

where $\boldsymbol{\gamma}$ is an array of regression parameters, $\mathbf{S}_i$ is an array of covariates, and $\mathbf{u}_i = (u_{i1}, ..., u_{ir})^{\mathrm{T}}$ is a vector of random effects that are independent of the errors $\mathbf{e}_i$. The model is flexible to cover a broad class of random effects models. For instance, with $\boldsymbol{\gamma}$ set to be zero, setting $\mathbf{S}_i = (1, t)^{\mathrm{T}}$ leads to the linear growth-curve model that is often discussed in the literature (e.g., Wulfsohn and Tsiatis 1997; Tseng, Heieh and Wang 2005), usually together with a multivariate normal distributional assumption $N_r(\boldsymbol{\mu}_u, \boldsymbol{\Sigma}_u)$ for random effects $\mathbf{u}_i$. Other specifications of $\mathbf{S}_i$ can describe more complex nonlinear form of $\mathbf{X}_i$.

3 Inference Methods

3.1 Regression Calibration

Let $\boldsymbol{\eta} = (\boldsymbol{\beta}^{\mathrm{T}}, \boldsymbol{\alpha}^{\mathrm{T}}, \boldsymbol{\gamma}^{\mathrm{T}})^{\mathrm{T}}$ be the vector containing all the related parameters, where $\boldsymbol{\beta}$ is the vector of regression parameters, $\boldsymbol{\alpha}$ is the vector of parameters characterizing the baseline distribution function

$F_0(t)$, and $\boldsymbol{\gamma}$ is comprised of the parameters associated with the random effect model in the covariate process. The likelihood function of $\boldsymbol{\eta}$ contributed from subject i is then given by

$$L_i(\boldsymbol{\eta}) = \{f(Y_i|\mathbf{X}_i, \mathbf{Z}_i)\}^{\delta_i}\{1 - F(Y_i|\mathbf{X}_i, \mathbf{Z}_i)\}^{1-\delta_i}, \tag{3.1}$$

and the likelihood function for $\boldsymbol{\eta}$ is

$$L(\boldsymbol{\eta}) = \prod_{i=1}^{n} L_i(\boldsymbol{\eta}). \tag{3.2}$$

If covariates $\mathbf{X}_i$ are free of measurement error, inference about the $\boldsymbol{\eta}$ parameter can be carried out by the maximum likelihood method. That is, maximizing $L(\boldsymbol{\eta})$ with respect to $\boldsymbol{\eta}$ leads to the maximum likelihood estimator of $\boldsymbol{\eta}$.

However, when $\mathbf{X}_i$ is subject to measurement error, the observed surrogate $\mathbf{X}_i^*$ may differ from $\mathbf{X}_i$, and replacing $\mathbf{X}_i$ with $\mathbf{X}_i^*$ in the likelihood function to perform inference about the parameter $\boldsymbol{\eta}$ may lead to biased results. One strategy to correct the resulting bias is to employ the so-called *regression calibration* method (Prentice 1982, Carroll et al. 2006). That is, we first replace $\mathbf{X}_i$ with its conditional expectation $E(\mathbf{X}_i|\mathbf{X}_i^*, \mathbf{Z}_i)$ in the likelihood function $L(\boldsymbol{\eta})$, and then maximize the resulting function with respect to $\boldsymbol{\eta}$. The bootstrap method can be used to determine the standard error of the resulting estimator of $\boldsymbol{\eta}$.

To implement the regression calibration method, we need to determine $E(\mathbf{X}_i|\mathbf{X}_i^*, \mathbf{Z}_i)$, which is a direct result from the availability of the conditional distribution $f(\mathbf{X}_i|\mathbf{X}_i^*, \mathbf{Z}_i)$. Given the model setup described in the previous section, the condition distribution is given by

$$f(\mathbf{X}_i|\mathbf{X}_i^*, \mathbf{Z}_i) = \frac{f(\mathbf{X}_i^*|\mathbf{X}_i, \mathbf{Z}_i)f(\mathbf{X}_i|\mathbf{Z}_i)}{\int f(\mathbf{X}_i^*|\mathbf{X}_i, \mathbf{Z}_i)f(\mathbf{X}_i|\mathbf{Z}_i)d\mathbf{X}_i},$$

where $f(\mathbf{X}_i^*|\mathbf{X}_i, \mathbf{Z}_i)$ is determined by (2.4) and (2.5), and $f(\mathbf{X}_i|\mathbf{Z}_i)$ is determined by (2.5). In particular, when $\mathbf{u}_i \sim N(\boldsymbol{\mu}_u, \boldsymbol{\Sigma}_u)$, we obtain that $\mathbf{X}_i|\mathbf{Z}_i \sim N(\mathbf{S}_i\boldsymbol{\mu}_u + \boldsymbol{\gamma}\mathbf{Z}_i, \mathbf{S}_i\boldsymbol{\Sigma}_u\mathbf{S}_i^{\mathrm{T}})$. By the assumption of additive measurement error model that $\mathbf{X}_i^*|(\mathbf{X}_i, \mathbf{Z}_i) \sim N(\mathbf{X}_i, \boldsymbol{\Sigma}_e)$, the conditional distribution of $\mathbf{X}_i$ given $(\mathbf{X}_i^*, \mathbf{Z}_i)$ is a normal distribution with mean $\mathbf{S}_i\boldsymbol{\mu}_u + \boldsymbol{\gamma}\mathbf{Z}_i + \mathbf{S}_i\boldsymbol{\Sigma}_u\mathbf{S}_i^{\mathrm{T}}(\mathbf{S}_i\boldsymbol{\Sigma}_u\mathbf{S}_i^{\mathrm{T}} + \boldsymbol{\Sigma}_e)^{-1}(\mathbf{X}_i^* - \mathbf{S}_i\boldsymbol{\mu}_u - \boldsymbol{\gamma}\mathbf{Z}_i)$.

3.2 Observed Likelihood

Under the layout of these models, one may carry out the likelihood-based analysis to perform estimation of the parameters. The likelihood function of $\boldsymbol{\eta}$ contributed from subject i is given by

$$L_i(\boldsymbol{\eta}) = \{f(Y_i|\mathbf{X}_i^*, \mathbf{Z}_i)\}^{\delta_i}\{1 - F(Y_i|\mathbf{X}_i^*, \mathbf{Z}_i)\}^{1-\delta_i},$$

where the probability density function is determined by

$$f(t_i|\mathbf{X}_i^*, \mathbf{Z}_i) = \int f(t_i|\mathbf{X}_i, \mathbf{Z}_i)f(\mathbf{X}_i|\mathbf{X}_i^*, \mathbf{Z}_i)d\mathbf{X}_i \tag{3.3}$$

and the observed cumulative distribution function is determined accordingly by

$$\begin{aligned} F(t_i|\mathbf{X}_i^*, \mathbf{Z}_i) &= \int_0^{t_i} f(v|\mathbf{X}_i^*, \mathbf{Z}_i) dv \\ &= \int_0^{t_i} \int f(v|\mathbf{X}_i, \mathbf{Z}_i) f(\mathbf{X}_i|\mathbf{X}_i^*, \mathbf{Z}_i) d\mathbf{X}_i dv \\ &= \int \left\{ f(\mathbf{X}_i|\mathbf{X}_i^*, \mathbf{Z}_i) \cdot \int_0^{t_i} f(v|\mathbf{X}_i, \mathbf{Z}_i) dv \right\} d\mathbf{X}_i \\ &= \int f(\mathbf{X}_i|\mathbf{X}_i^*, \mathbf{Z}_i) F(t_i|\mathbf{X}_i, \mathbf{Z}_i) d\mathbf{X}_i \end{aligned}$$

with the function $F(.)$ being determined by (2.2).

In the formulation here we assume nondifferential measurement error mechanism, i.e., $f(t_i|\mathbf{X}_i^*, \mathbf{X}_i, \mathbf{Z}_i) = f(t_i|\mathbf{X}_i, \mathbf{Z}_i)$. Consequently, the observed likelihood function is

$$L^*(\boldsymbol{\eta}) = \prod_{i=1}^{n} L_i^*(\boldsymbol{\eta}), \tag{3.4}$$

where

$$L_i^*(\boldsymbol{\eta}) = \left\{ \int f(t_i|\mathbf{X}_i, \mathbf{Z}_i) f(\mathbf{X}_i|\mathbf{X}_i^*, \mathbf{Z}_i) d\mathbf{X}_i \right\}^{\delta_i} \left\{ 1 - \int f(\mathbf{X}_i|\mathbf{X}_i^*, \mathbf{Z}_i) F(t_i|\mathbf{X}_i, \mathbf{Z}_i) d\mathbf{X}_i \right\}^{1-\delta_i}.$$

Inference about $\boldsymbol{\eta}$ can be performed based on the observed likelihood function $L^*(\boldsymbol{\eta})$. Maximizing $L^*(\boldsymbol{\eta})$ gives the maximum likelihood estimator of $\boldsymbol{\eta}$, denoted by $\widehat{\boldsymbol{\eta}}$. Under suitable conditions, $\sqrt{n}(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})$ has an asymptotic normal distribution with mean $\mathbf{0}$ and covariance matrix $\left[E\{(\partial \ell_i^*(\boldsymbol{\eta})/\partial \boldsymbol{\eta}^{\mathrm{T}})(\partial \ell_i^*(\boldsymbol{\eta})/\partial \boldsymbol{\eta})\}\right]^{-1}$, where $\ell_i^*(\boldsymbol{\eta}) = \log L_i^*(\boldsymbol{\eta})$. Some details are included in the appendix.

Directly maximizing $L^*(\boldsymbol{\eta})$ however is not possible, since (3.4) does not have a closed form due to the involvement of nonlinear functions in the integrations. A remedy to this complication is to invoke numerical approximations such as the Monte Carlo algorithm. To be specific, let $\ell^*(\boldsymbol{\eta}) = \sum_{i=1}^{n} \ell_i^*(\boldsymbol{\eta})$. Note that

$$\begin{aligned} \ell_i^*(\boldsymbol{\eta}) &= \log L_i^*(\boldsymbol{\eta}) \\ &= \delta_i \log \left\{ \int f(t_i|\mathbf{X}_i, \mathbf{Z}_i) f(\mathbf{X}_i|\mathbf{X}_i^*, \mathbf{Z}_i) d\mathbf{X}_i \right\} \\ &\quad + (1-\delta_i) \log \left\{ 1 - \int f(\mathbf{X}_i|\mathbf{X}_i^*, \mathbf{Z}_i) F(t_i|\mathbf{X}_i, \mathbf{Z}_i) d\mathbf{X}_i \right\}. \end{aligned}$$

Specify a large integer N. We generate $\mathbf{x}_i^{(1)}, ..., \mathbf{x}_i^{(N)}$ from the distribution $f(\mathbf{X}_i|\mathbf{X}_i^*, \mathbf{Z}_i)$ for $i = 1, ..., n$. Define

$$\widetilde{\ell}_i^*(\boldsymbol{\eta}) = \delta_i \log \left\{ \frac{1}{N} \sum_{j=1}^{N} f(t_i|\mathbf{x}_i^{(j)}, \mathbf{Z}_i) \right\} + (1-\delta_i) \log \left\{ 1 - \frac{1}{N} \sum_{j=1}^{N} F(t_i|\mathbf{x}_i^{(j)}, \mathbf{Z}_i) \right\},$$

and $\widetilde{\ell}^*(\boldsymbol{\eta}) = \sum_{i=1}^n \widetilde{\ell}_i^*(\boldsymbol{\eta})$. Then we use $\widetilde{\ell}^*(\boldsymbol{\eta})$ to approximate $\ell^*(\boldsymbol{\eta})$ and maximize $\widetilde{\ell}^*(\boldsymbol{\eta})$ to obtain an estimator of $\boldsymbol{\eta}$. An estimate of the covariance matrix of $\widehat{\boldsymbol{\eta}}$ is given by $\Big[\sum_{i=1}^n (\partial \widetilde{\ell}_i^* / \partial \boldsymbol{\eta}^{\mathrm{T}})(\partial \widetilde{\ell}_i^* / \partial \boldsymbol{\eta})\Big]^{-1}$.

4 Empirical Studies

In this section we conduct simulation studies to assess the performance of the proposed approaches, as opposed to the naive analysis which ignores covariate measurement error.

4.1 Design of Simulation

We consider a single covariate X_i for simplicity. Sample size is set to be 200, and 500 simulations are run for each of the parameter configurations. For each subject i, the covariate Z_i is generated from a binomial distribution $\mathrm{Bin}(1, 0.5)$, representing each subject is randomly assigned to either treatment or control arms, for instance. The covariate X_i is generated from the linear mixed model

$$X_i = u_{i0} + u_{i1} Z_i,$$

where we set $\mathbf{u}_i = (u_{i0}, u_{i1})^{\mathrm{T}}$ to follow a bivariate normal distribution with mean $\boldsymbol{\mu} = (4.173, -0.0103)$ and covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1.24 & -0.0114 \\ -0.0114 & 0.003 \end{pmatrix},$$

to mimic an HIV clinical trial as described in Tsiatis and Davidian (2004). Surrogate X_i^* is generated from the error model $X_i^* = X_i + e_i$ with $e_i \sim N(0, \sigma^2)$.

The survival times are generated based on the proportional odds model

$$\frac{F(T_i)}{1 - F(T_i)} = \frac{F_0(T_i)}{1 - F_0(T_i)} \exp(\beta_x X_i + \beta_z Z_i),\ i = 1, 2, ..., n$$

where $F_0(t)$ is the baseline distribution function. We particularly consider two parametric modeling for $F_0(t)$ by using (1) a log-logistic distribution and (2) a Weibull distribution.

To be specific, in the first case, we set $F_0(t)$ to be a log-logistic distribution

$$F_0(t) = 1 - \{1 + (\alpha_1 t)^{\alpha_2}\}^{-1}$$

with $\alpha_1 = 5$ and $\alpha_2 = 2$. That is, for the ith subject, we first generate a random variate $v_i \sim U[0, 1]$, and calculate the survival time T_i by

$$T_i = \Big[\frac{v_i}{\alpha_1^{\alpha_2}(1 - v_i)} \exp\{-(\beta_x X_i + \beta_z Z_i)\}\Big]^{1/\alpha_2}.$$

In the second case, we set $F_0(t)$ to be a Weibull distribution

$$F_0(t) = 1 - \exp\{-(\alpha_1 t)^{\alpha_2}\}$$

with $\alpha_1 = 1$ and $\alpha_2 = 1.5$. The survival times can be generated similarly like the log-logistic case:

$$T_i = \frac{1}{\alpha_1}\Big\{\log\Big[\frac{v_i}{1-v_i}\exp\{-(\beta_x X_i + \beta_z Z_i)\} + 1\Big]\Big\}^{1/\alpha_2},$$

where $v_i \sim U[0,1]$.

The censoring times are generated independently from an exponential distribution with a fixed mean to produce roughly 20% and 40% censoring rates. We assess the performance of the proposed methods under a variety of situations. Different configurations for the magnitude σ^2 of measurement error are considered, with $\sigma^2 = 0, 0.25,$ and 0.6 to feature increasing degrees of measurement error in covariate X_i. The number N of the Monte Carlo simulations is chosen as 5000.

4.2 Simulation Results

We compare the performance of the three methods: (1) the naive method with the true covariate X_i directly replaced by its surrogate X_i^*, (2) the regression calibration approach which replaces the true covariate X_i with its conditional expectation $E(X_i|X_i^*, Z_i)$, and (3) the observed likelihood method. Specifically, we report on the results of the bias of the estimates (Bias), the empirical standard error (SEE), the model based standard error (SEM), and the coverage rate (CR) for 95% confidence intervals of the parameters. In particular, for the case that the baseline distribution is modeled by a log-logistic distribution, Tables 1.1 and 1.2 report the results for the $\boldsymbol{\beta}$ parameters by varying the magnitude of the covariate effects, and Tables 1.3 and 1.4 report the results for the $\boldsymbol{\alpha}$ parameter related to the baseline distribution. Similarly, Tables 2.1 to 2.4 display the results for the case that the baseline distribution is set as a Weibull distribution.

The impact of ignoring measurement error depends on the magnitude of error as well as the covariate effects, and this is evident from the results obtained from the naive analysis. If the true error-prone covariate effect is not zero, it is seen that as the magnitude of the measurement error becomes severe, the bias of the estimate for the error prone covariate effect increases while the associated standard error tends to get smaller; consequently, the corresponding coverage rate for the 95% confidence intervals becomes farther off the nominal level 95%. When the covariate effect of the error prone covariate is zero, then ignoring measurement error does not appear to have impact on estimation. The point estimate, standard error and the coverage rate all seem to be satisfactory, and they are not shown to be obviously affected by varying degrees of measurement error. The impact on estimation of the error-free covariate effect is not apparent. The finite sample bias and the associated standard error for the estimate of β_z are fairly reasonable and stable as the degree in error changes, and the coverage rate for the confidence intervals agrees well with the nominal level 95%.

On the other hand, it is interesting to see that ignoring covariate measurement error has an effect on estimation of the parameters of the baseline distribution function. If the true error-prone covariate effect exists, then the finite sample bias tends to increase as the magnitude of error increases, whereas the standard error seems to be relatively stable to the change of the error degree. The joint impact is also evident from the discrepancy between the coverage rate of the 95% confidence intervals and the nominal level. However, when the true error-prone covariate effect does not exist, the naive analysis does not appear to influence estimation of the $\boldsymbol{\alpha}$ parameter. When the censoring percentage and the baseline distribution change, we observe the similar impact of measurement error on estimation of the parameters.

In implementing the regression calibration approach, we employ the bootstrap method with 1000 runs for each setting to calculate the model-based standard error (SEM). As the error effect is partially adjusted through the replacement of the error contaminated covariate by its expected value given the observed covariate, the performance of the regression calibration method on estimation of the $\boldsymbol{\beta}$ parameters is greatly improved in contrast to that of the naive method. In the case that β_x is not zero, the regression calibration produces much smaller finite sample biases for the estimates of both the β_x and β_z parameters than the naive method does; if β_x is zero, the estimates from both methods are quite comparable. The standard errors obtained from the regression calibration are greater than those obtained from the naive analysis, but the corresponding coverage rates are reasonably close to the nominal value, regardless of the value of β_x. The effect of the censoring percentage and the magnitude of measurement error on the performance of the regression calibration seems quite similar to that on the naive analysis. Although it is observed that the regression calibration method can correct or partially correct for the bias induced by measurement error when estimating the $\boldsymbol{\beta}$ parameters, it is interesting to see that it does not necessarily do this for estimation of the parameters of the baseline distribution. See the settings with $\sigma^2 = 0.6$ in Table 1.3, for example.

Finally, we examine the performance of the observed likelihood approach which fully accommodates the measurement error effects in inferential procedures. In actual implementation, 5000 simulations are run for each configuration when using the Monte Carlo integration algorithm. It is seen that the performance of this method is satisfactory in all the settings. This method further improves the results obtained from the regression calibration method. Unlike the regression calibration approach, the observed likelihood method gives good estimates for the parameters of the baseline distribution functions. The observed likelihood approach produces reasonably small finite sample biases for both the $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ parameters and good coverage rates for the 95% confidence intervals. The standard errors resulted from the observed likelihood and the regression calibration methods are fairly close in most settings. The censoring percentage and the degree of measurement error have the similar effects on the performance of the observed likelihood method to those for the regression calibration approach.

In summary, the naive analysis with covariate measurement error ignored would often lead to biased results, whereas the regression calibration approach improves the results a lot, and its performance is reasonably satisfactory for many settings but not all. The observed likelihood method behaves the best and outperforms the other two methods. It can not only adjust for the measurement error effects on estimation of the covariate effects, but also produce reasonable estimates for the parameters of the baseline distribution function.

5 Discussion

The impact of covariate measurement error is well documented for survival models such as proportional hazards, accelerated failure time and additive hazards models. Various methods have been developed to correct for effects induced by mismeasured covariates. In this paper, we consider an alternative, the proportional odds model, which is useful for survival analysis. Our numerical studies demonstrate that the naive analysis with covariate error ignored would lead to biased results. To correct for error effects, we describe two methods: the regression calibration and the observed likelihood methods. Empirical studies demonstrate satisfactory performance of the observed likelihood method. They also suggest that the regression calibration approach dramatically outperforms the naive method, and the performance of the regression calibration is quite comparable to the observed likelihood in most of the settings we consider.

Covariate measurement error is a common feature in many applications. In specific applications, often it is known that some covariates are subject to measurement error, but there is a lack of additional data sources, such as a validation subsample or replicated measurements for those covariates, that can be used feasibly to facilitate the estimation of the measurement error parameters. Under such a circumstance, a viable strategy is to conduct sensitivity analyses to evaluate how sensitive the estimates of the response parameters are affected by different degrees of measurement error. Typically, one may choose a set of given values of the parameters for the measurement error model, and repeatedly conduct estimation on the response parameters to see what patterns may be unfolded. This strategy can enhance our understanding of the measurement error effects on individual data analysis.

Acknowledgement

This research was supported by the Natural Sciences and Engineering Research Council of Canada.

References

[1] Bennett, S. (1983). Log-logistic regression models for survival data. *Applied Statistics*, **32**, 165-171.

[2] Carroll, R. J., Ruppert, D., Stefanski, L. A. and Crainiceanu, C. M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*, Second Edition. Chapman and Hall CRC Press.

[3] Giménez, P., Bolfarine, H. and Colosimo, E. A. (1999). Estimation in Weibull regression model with measurement error. *Communications in Statistics - Theory and Method*, **28**, 495-510.

[4] Giménez, P., Bolfarine, H. and Colosimo, E. A. (2006). Asymptotic relative efficiency of score tests in Weibull models with measurement errors. *Statistical Papers*, **47**, 461-470.

[5] Greene, W. F. and Cai, J. (2004). Measurement error in covariates in the marginal hazards model for multivariate failure time data. *Biometrics*, **60**, 987-996.

[6] He, W., Yi, G. Y. and Xiong, J. (2007). Accelerated failure time models with covariates subject to measurement error. *Statistics in Medicine*, **26**, 4817-4832.

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

[8] 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.

[9] Kulich, M. and Lin, D. Y. (2000). Additive hazards regression with covariate measurement error. *Journal of the American Statistical Association*, **95**, 238-248.

[10] Lehmann, E. L. (1983). *Theory of Point Estimation*. New York: John Wiley.

[11] Li, Y. and Lin, X. (2003). Functional inference in frailty measurement error models for clustered survival data using the SIMEX approach. *Journal of the American Statistical Association*, **98**, 191-203.

[12] Li, Y. and Ryan, L. (2004). Survival analysis with heterogeneous covariate measurement Error. *Journal of the American Statistical Association*, **99**, 724-735.

[13] Li, Y. and Ryan, L. (2006). Inference on survival data with covariate measurement error - A imputation-based approach. *The Scandinavian Journal of Statistics*, **33**, 169-190.

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

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

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

[17] Sun, L. Zhang, Z. and Sun, J. (2006). Additive hazards regression of failure time data with covariate measurement errors. *Statistica Neerlandica*, **60**, 497-509.

[18] Sun, L. and Zhou, X. (2008). Inference in the additive risk model with time-varying covariates subject to measurement errors. *Statistics and Probability Letters*, **78**, 2559-2566.

[19] Tseng, Y. K., Hsieh, F. and Wang, J. L. (2005). Joint modeling of accelerated failure time and longitudinal data. *Biometrika*, **92**, 587-603.

[20] Tsiatis, A. A. and Davidian, M. (2004). Joint modeling of longitudinal and time-to-even data: An overview. *Statistica Sinica*, **14**, 809-834.

[21] Wulfsohn, M. S. and Tsiatis, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics*, **53**, 330-339.

[22] Xie, S. H., 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 B*, **63**, 855-870.

[23] Yang, S. and Prentice, R. L. (1999). Semiparametric inference in the proportional odds model. *Journal of the American Statistical Association*, **94**, 125-136.

[24] Yi, G. Y. and He, W. (2006). Methods for bivariate survival data with mismeasured covariates under an accelerated failure time model. *Communications in Statistics - Theory and Methods*, **35**, 1539-1554.

[25] Yi, G. Y. and Lawless, J. F. (2007). A corrected likelihood method for the proportional hazards model with covariates subject to measurement error. *Journal of Statistical Planning and Inference*, **137**, 1816-1828.

[26] Zhou, H. and Wang C. Y. (2000). Failure time regression with continuous covariates measured with error. *Journal of the Royal Statistical Society B*, **62**, 657-665.

Table 1.1: Simulation Results on the Regression Coefficients Obtained from the Parametric Approach. The True Baseline Distribution Follows A Log-logistic Distribution, and the Error-Prone Covariate has An Effect with $\beta_x = -1$.

Censoring	Measurement		Naive Method				Regression Calibration				Likelihood Method			
Rate	Error	Parameter	Bias [a]	SEM	SEE	CR	Bias	SEM	SEE	CR	Bias	SEM	SEE	CR
20%	$\sigma^2 = 0$	$\beta_x = -1$	-0.012	0.134	0.133	0.952	-0.012	0.139	0.133	0.962	-0.012	0.134	0.133	0.952
		$\beta_z = 0.5$	0.016	0.262	0.259	0.950	0.016	0.270	0.259	0.956	0.016	0.262	0.259	0.950
	$\sigma^2 = 0.25$	$\beta_x = -1$	0.187	0.119	0.121	0.614	0.021	0.148	0.145	0.948	-0.017	0.160	0.161	0.954
		$\beta_z = 0.5$	0.019	0.262	0.268	0.952	0.018	0.271	0.270	0.958	0.037	0.273	0.275	0.952
	$\sigma^2 = 0.6$	$\beta_x = -1$	0.363	0.104	0.105	0.086	0.051	0.161	0.157	0.946	-0.023	0.192	0.194	0.964
		$\beta_z = 0.5$	0.004	0.261	0.271	0.944	0.002	0.271	0.270	0.950	0.040	0.283	0.286	0.946
40%	$\sigma^2 = 0$	$\beta_x = -1$	-0.020	0.147	0.146	0.948	-0.020	0.154	0.145	0.962	-0.020	0.147	0.146	0.948
		$\beta_z = 0.5$	0.017	0.285	0.278	0.956	0.016	0.295	0.278	0.958	0.017	0.285	0.278	0.956
	$\sigma^2 = 0.25$	$\beta_x = -1$	0.187	0.130	0.132	0.668	0.021	0.162	0.157	0.950	-0.017	0.175	0.176	0.954
		$\beta_z = 0.5$	0.009	0.284	0.284	0.958	0.005	0.295	0.286	0.962	0.026	0.296	0.295	0.956
	$\sigma^2 = 0.6$	$\beta_x = -1$	0.362	0.114	0.115	0.130	0.050	0.175	0.170	0.940	-0.028	0.211	0.211	0.966
		$\beta_z = 0.5$	-0.006	0.283	0.285	0.960	-0.008	0.294	0.286	0.962	0.032	0.307	0.305	0.964

[a]Bias: bias of the estimate; SEM: model-based standard error; SEE: empirical standard error; CR: 95% coverage rate. The SEM from regression calibration approach is calculated by the bootstrap method with B = 1000. The likelihood method uses Monte Carlo integration algorithm with N = 5000.

Table 1.2: Simulation Results on the Regression Coefficients Obtained from the Parametric Approach. The True Baseline Distribution Follows A Log-logistic Distribution, and the Error-Prone Covariate has No Effect with $\beta_x = 0$.

Censoring	Measurement		Naive Method				Regression Calibration				Likelihood Method			
Rate	Error	Parameter	Bias [a]	SEM	SEE	CR	Bias	SEM	SEE	CR	Bias	SEM	SEE	CR
20%	$\sigma^2 = 0$	$\beta_x = 0$	0.003	0.117	0.119	0.950	0.003	0.122	0.119	0.956	0.003	0.117	0.119	0.950
		$\beta_z = 0.5$	0.016	0.261	0.259	0.946	0.016	0.269	0.259	0.952	0.016	0.261	0.259	0.946
	$\sigma^2 = 0.25$	$\beta_x = 0$	0.002	0.107	0.109	0.952	0.004	0.133	0.132	0.946	0.003	0.129	0.130	0.954
		$\beta_z = 0.5$	0.033	0.262	0.260	0.958	0.031	0.270	0.263	0.964	0.026	0.262	0.260	0.958
	$\sigma^2 = 0.6$	$\beta_x = 0$	0.003	0.096	0.098	0.948	0.006	0.147	0.147	0.952	0.005	0.144	0.148	0.944
		$\beta_z = 0.5$	0.033	0.262	0.263	0.956	0.033	0.270	0.262	0.964	0.036	0.262	0.263	0.956
40%	$\sigma^2 = 0$	$\beta_x = 0$	0.000	0.127	0.130	0.948	0.000	0.133	0.130	0.956	0.000	0.127	0.130	0.948
		$\beta_z = 0.5$	0.014	0.283	0.270	0.962	0.014	0.293	0.270	0.966	0.014	0.283	0.270	0.962
	$\sigma^2 = 0.25$	$\beta_x = 0$	0.001	0.115	0.116	0.950	0.004	0.144	0.139	0.958	0.004	0.139	0.140	0.948
		$\beta_z = 0.5$	0.025	0.283	0.288	0.948	0.032	0.293	0.285	0.964	0.037	0.283	0.287	0.946
	$\sigma^2 = 0.6$	$\beta_x = 0$	0.002	0.104	0.104	0.950	0.002	0.161	0.156	0.956	0.004	0.155	0.157	0.952
		$\beta_z = 0.5$	0.030	0.283	0.286	0.948	0.032	0.293	0.284	0.966	0.032	0.283	0.285	0.948

[a]Bias: bias of the estimate; SEM: model-based standard error; SEE: empirical standard error; CR: 95% coverage rate. The SEM from regression calibration approach is calculated by the bootstrap method with B = 1000. The likelihood method uses Monte Carlo integration algorithm with N = 5000.

Table 1.3: Simulation Results on The Baseline Function Parameters Obtained from The Parametric Approach. The True Baseline Distribution Follows A Log-logistic Distribution. The True Values are $\beta_x = -1$, $\beta_z = 0.5$, $\alpha_1 = 5$ and $\alpha_2 = 2$.

Censoring	Measurement		Naive Method				Regression Calibration				Likelihood Method			
Rate	Error	Parameter	Bias [a]	SEM	SEE	CR	Bias	SEM	SEE	CR	Bias	SEM	SEE	CR
20%	$\sigma^2 = 0$	$\log(\alpha_1)$	-0.004	0.260	0.254	0.954	-0.004	0.266	0.254	0.952	-0.004	0.260	0.254	0.954
		$\log(\alpha_2)$	0.010	0.065	0.062	0.972	0.010	0.066	0.062	0.964	0.010	0.065	0.062	0.972
	$\sigma^2 = 0.25$	$\log(\alpha_1)$	-0.362	0.246	0.254	0.666	-0.011	0.296	0.300	0.944	-0.012	0.291	0.298	0.940
		$\log(\alpha_2)$	-0.021	0.065	0.066	0.938	-0.021	0.065	0.066	0.932	0.017	0.069	0.071	0.938
	$\sigma^2 = 0.6$	$\log(\alpha_1)$	-0.695	0.232	0.238	0.144	-0.010	0.337	0.338	0.944	-0.011	0.330	0.339	0.946
		$\log(\alpha_2)$	-0.053	0.065	0.066	0.874	-0.052	0.064	0.066	0.870	0.019	0.076	0.077	0.942
40%	$\sigma^2 = 0$	$\log(\alpha_1)$	0.007	0.281	0.275	0.952	0.008	0.290	0.275	0.952	0.008	0.281	0.275	0.952
		$\log(\alpha_2)$	0.012	0.074	0.074	0.952	0.012	0.075	0.074	0.952	0.011	0.074	0.074	0.952
	$\sigma^2 = 0.25$	$\log(\alpha_1)$	-0.361	0.266	0.273	0.706	-0.010	0.320	0.320	0.940	-0.013	0.314	0.319	0.936
		$\log(\alpha_2)$	-0.019	0.074	0.076	0.932	-0.019	0.074	0.076	0.938	0.019	0.079	0.081	0.940
	$\sigma^2 = 0.6$	$\log(\alpha_1)$	-0.689	0.249	0.256	0.204	-0.006	0.363	0.361	0.944	-0.005	0.356	0.361	0.940
		$\log(\alpha_2)$	-0.051	0.074	0.074	0.904	-0.050	0.074	0.075	0.908	0.022	0.085	0.087	0.946

[a] Bias: bias of the estimate; SEM: model-based standard error; SEE: empirical standard error; CR: 95% coverage rate. The SEM from regression calibration approach is calculated by the bootstrap method with B = 1000. The likelihood method uses Monte Carlo integration algorithm with N = 5000.

Table 1.4: Simulation Results on The Baseline Function Parameters Obtained from The Parametric Approach. The True Baseline Distribution Follows A Log-logistic Distribution. The True Values are $\beta_x = 0$, $\beta_z = 0.5$, $\alpha_1 = 5$ and $\alpha_2 = 2$.

Censoring	Measurement		Naive Method				Regression Calibration				Likelihood Method			
Rate	Error	Parameter	Bias [a]	SEM	SEE	CR	Bias	SEM	SEE	CR	Bias	SEM	SEE	CR
20%	$\sigma^2 = 0$	$\log(\alpha_1)$	-0.008	0.260	0.256	0.946	-0.008	0.267	0.256	0.952	-0.008	0.260	0.256	0.946
		$\log(\alpha_2)$	0.011	0.065	0.064	0.960	0.011	0.066	0.064	0.960	0.011	0.065	0.064	0.960
	$\sigma^2 = 0.25$	$\log(\alpha_1)$	-0.009	0.238	0.237	0.946	-0.013	0.285	0.285	0.942	-0.009	0.280	0.278	0.950
		$\log(\alpha_2)$	0.015	0.065	0.066	0.932	0.016	0.066	0.065	0.934	0.016	0.065	0.066	0.938
	$\sigma^2 = 0.6$	$\log(\alpha_1)$	-0.010	0.218	0.220	0.946	-0.019	0.314	0.313	0.952	-0.016	0.308	0.311	0.952
		$\log(\alpha_2)$	0.015	0.065	0.066	0.938	0.014	0.066	0.066	0.934	0.018	0.065	0.067	0.934
40%	$\sigma^2 = 0$	$\log(\alpha_1)$	-0.005	0.282	0.281	0.944	-0.005	0.291	0.281	0.942	-0.005	0.282	0.281	0.944
		$\log(\alpha_2)$	0.010	0.074	0.073	0.958	0.010	0.075	0.073	0.952	0.010	0.074	0.073	0.958
	$\sigma^2 = 0.25$	$\log(\alpha_1)$	-0.004	0.258	0.257	0.952	-0.014	0.311	0.304	0.954	-0.012	0.304	0.303	0.948
		$\log(\alpha_2)$	0.016	0.074	0.076	0.934	0.015	0.075	0.076	0.936	0.015	0.074	0.075	0.942
	$\sigma^2 = 0.6$	$\log(\alpha_1)$	-0.006	0.237	0.234	0.950	-0.009	0.342	0.336	0.944	-0.013	0.335	0.335	0.940
		$\log(\alpha_2)$	0.015	0.074	0.076	0.938	0.015	0.075	0.075	0.936	0.016	0.074	0.077	0.940

[a] Bias: bias of the estimate; SEM: model-based standard error; SEE: empirical standard error; CR: 95% coverage rate. The SEM from regression calibration approach is calculated by the bootstrap method with B = 1000. The likelihood method uses Monte Carlo integration algorithm with N = 5000.

Table 2.1: Simulation Results on the Regression Coefficients Obtained from the Parametric Approach. The True Baseline Distribution Follows A Weibull Distribution, and the Error-Prone Covariate has An Effect with $\beta_x = -1$.

Censoring Rate	Measurement Error	Parameter	Naive Method Bias [a]	SEM	SEE	CR	Regression Calibration Bias	SEM	SEE	CR	Likelihood Method Bias	SEM	SEE	CR
20%	$\sigma^2 = 0$	$\beta_x = -1$	-0.004	0.136	0.136	0.944	-0.004	0.142	0.136	0.946	-0.004	0.136	0.136	0.944
		$\beta_z = 0.5$	-0.014	0.269	0.259	0.952	-0.014	0.277	0.259	0.958	-0.015	0.269	0.259	0.952
	$\sigma^2 = 0.25$	$\beta_x = -1$	0.179	0.121	0.127	0.660	0.028	0.150	0.148	0.942	-0.018	0.162	0.165	0.950
		$\beta_z = 0.5$	-0.037	0.268	0.273	0.938	-0.018	0.278	0.275	0.944	-0.006	0.280	0.285	0.936
	$\sigma^2 = 0.6$	$\beta_x = -1$	0.350	0.106	0.111	0.130	0.066	0.161	0.158	0.930	-0.021	0.193	0.198	0.948
		$\beta_z = 0.5$	-0.060	0.267	0.268	0.942	-0.029	0.279	0.275	0.948	-0.002	0.291	0.290	0.948
40%	$\sigma^2 = 0$	$\beta_x = -1$	0.001	0.152	0.147	0.956	0.001	0.161	0.147	0.962	0.000	0.152	0.147	0.956
		$\beta_z = 0.5$	-0.019	0.299	0.282	0.972	-0.019	0.311	0.282	0.974	-0.019	0.299	0.282	0.972
	$\sigma^2 = 0.25$	$\beta_x = -1$	0.173	0.135	0.143	0.710	0.024	0.170	0.165	0.952	-0.024	0.180	0.188	0.934
		$\beta_z = 0.5$	-0.035	0.298	0.302	0.940	-0.016	0.313	0.304	0.950	-0.008	0.311	0.313	0.946
	$\sigma^2 = 0.6$	$\beta_x = -1$	0.345	0.117	0.126	0.190	0.058	0.182	0.179	0.930	-0.027	0.215	0.230	0.934
		$\beta_z = 0.5$	-0.066	0.296	0.300	0.938	-0.031	0.313	0.304	0.956	0.003	0.323	0.330	0.946

[a] Bias: bias of the estimate; SEM: model-based standard error; SEE: empirical standard error; CR: 95% coverage rate. The SEM from regression calibration approach is calculated by the bootstrap method with B = 1000. The likelihood method uses Monte Carlo integration algorithm with N = 5000.

Table 2.2: Simulation Results on the Regression Coefficients Obtained from the Parametric Approach. The True Baseline Distribution Follows A Weibull Distribution, and the Error-Prone Covariate has No Effect with $\beta_x = 0$.

Censoring	Measurement		Naive Method				Regression Calibration				Likelihood Method			
Rate	Error	Parameter	Bias [a]	SEM	SEE	CR	Bias	SEM	SEE	CR	Bias	SEM	SEE	CR
20%	$\sigma^2 = 0$	$\beta_x = 0$	0.002	0.098	0.099	0.954	0.002	0.101	0.099	0.962	0.002	0.098	0.099	0.954
		$\beta_z = 0.5$	-0.016	0.260	0.254	0.948	-0.016	0.268	0.254	0.960	-0.016	0.260	0.254	0.948
	$\sigma^2 = 0.25$	$\beta_x = 0$	0.001	0.092	0.095	0.938	0.001	0.109	0.109	0.934	0.002	0.105	0.108	0.936
		$\beta_z = 0.5$	-0.015	0.260	0.267	0.956	-0.020	0.268	0.267	0.950	-0.019	0.260	0.264	0.956
	$\sigma^2 = 0.6$	$\beta_x = 0$	0.001	0.085	0.089	0.938	0.001	0.118	0.117	0.940	0.001	0.113	0.117	0.940
		$\beta_z = 0.5$	-0.015	0.259	0.267	0.954	-0.017	0.268	0.266	0.954	-0.011	0.260	0.267	0.958
40%	$\sigma^2 = 0$	$\beta_x = 0$	-0.002	0.108	0.109	0.948	-0.002	0.112	0.109	0.954	-0.002	0.108	0.109	0.948
		$\beta_z = 0.5$	-0.023	0.280	0.273	0.958	-0.023	0.290	0.273	0.964	-0.023	0.280	0.273	0.958
	$\sigma^2 = 0.25$	$\beta_x = 0$	0.000	0.101	0.100	0.954	0.001	0.121	0.115	0.958	0.001	0.115	0.115	0.960
		$\beta_z = 0.5$	-0.015	0.280	0.287	0.954	-0.014	0.290	0.287	0.954	-0.013	0.281	0.291	0.948
	$\sigma^2 = 0.6$	$\beta_x = 0$	0.001	0.093	0.094	0.940	0.001	0.131	0.124	0.954	0.001	0.124	0.124	0.952
		$\beta_z = 0.5$	-0.014	0.280	0.286	0.954	-0.013	0.291	0.286	0.958	-0.013	0.281	0.287	0.950

[a] Bias: bias of the estimate; SEM: model-based standard error; SEE: empirical standard error; CR: 95% coverage rate. The SEM from regression calibration approach is calculated by the bootstrap method with B = 1000. The likelihood method uses Monte Carlo integration algorithm with N = 5000.

Table 2.3: Simulation Results on The Baseline Function Parameters Obtained from The Parametric Approach. The True Baseline Distribution Follows A Weibull Distribution. The True Values are $\beta_x = -1$, $\beta_z = 0.5$, $\alpha_1 = 1$ and $\alpha_2 = 1.5$.

Censoring	Measurement		Naive Method				Regression Calibration				Likelihood Method			
Rate	Error	Parameter	Bias [a]	SEM	SEE	CR	Bias	SEM	SEE	CR	Bias	SEM	SEE	CR
20%	$\sigma^2 = 0$	$\log(\alpha_1)$	0.004	0.195	0.193	0.940	0.004	0.201	0.193	0.942	0.005	0.195	0.193	0.940
		$\log(\alpha_2)$	0.009	0.114	0.113	0.938	0.009	0.118	0.113	0.948	0.009	0.114	0.114	0.938
	$\sigma^2 = 0.25$	$\log(\alpha_1)$	-0.233	0.164	0.170	0.630	-0.015	0.218	0.218	0.944	0.016	0.223	0.227	0.942
		$\log(\alpha_2)$	0.146	0.114	0.116	0.760	0.009	0.129	0.127	0.942	0.007	0.125	0.126	0.940
	$\sigma^2 = 0.6$	$\log(\alpha_1)$	-0.434	0.133	0.143	0.164	-0.039	0.241	0.241	0.930	0.019	0.257	0.264	0.942
		$\log(\alpha_2)$	0.280	0.109	0.116	0.286	0.012	0.144	0.142	0.944	0.010	0.140	0.143	0.940
40%	$\sigma^2 = 0$	$\log(\alpha_1)$	0.000	0.218	0.213	0.946	0.000	0.227	0.213	0.950	0.001	0.218	0.213	0.946
		$\log(\alpha_2)$	0.014	0.129	0.126	0.956	0.014	0.134	0.126	0.958	0.013	0.129	0.126	0.956
	$\sigma^2 = 0.25$	$\log(\alpha_1)$	-0.228	0.181	0.195	0.686	-0.008	0.247	0.245	0.934	0.024	0.248	0.261	0.942
		$\log(\alpha_2)$	0.147	0.126	0.133	0.780	0.008	0.146	0.142	0.946	0.007	0.140	0.144	0.940
	$\sigma^2 = 0.6$	$\log(\alpha_1)$	-0.431	0.146	0.160	0.222	-0.025	0.273	0.276	0.926	0.023	0.286	0.304	0.922
		$\log(\alpha_2)$	0.282	0.120	0.129	0.362	0.007	0.162	0.161	0.942	0.013	0.156	0.161	0.936

[a]Bias: bias of the estimate; SEM: model-based standard error; SEE: empirical standard error; CR: 95% coverage rate. The SEM from regression calibration approach is calculated by the bootstrap method with B = 1000. The likelihood method uses Monte Carlo integration algorithm with N = 5000.

Table 2.4: Simulation Results on The Baseline Function Parameters Obtained from The Parametric Approach. The True Baseline Distribution Follows A Weibull Distribution. The True Values are $\beta_x = 0$, $\beta_z = 0.5$, $\alpha_1 = 1$ and $\alpha_2 = 1.5$.

Censoring	Measurement		Naive Method				Regression Calibration				Likelihood Method			
Rate	Error	Parameter	Bias [a]	SEM	SEE	CR	Bias	SEM	SEE	CR	Bias	SEM	SEE	CR
20%	$\sigma^2 = 0$	$\log(\alpha_1)$	0.010	0.176	0.175	0.948	0.010	0.183	0.175	0.954	0.010	0.176	0.175	0.948
		$\log(\alpha_2)$	0.004	0.100	0.096	0.962	0.004	0.101	0.096	0.962	0.004	0.100	0.096	0.962
	$\sigma^2 = 0.25$	$\log(\alpha_1)$	0.010	0.166	0.172	0.950	0.013	0.195	0.193	0.946	0.010	0.186	0.193	0.952
		$\log(\alpha_2)$	0.011	0.096	0.097	0.948	0.008	0.107	0.105	0.946	0.010	0.104	0.105	0.946
	$\sigma^2 = 0.6$	$\log(\alpha_1)$	0.010	0.155	0.164	0.942	0.013	0.213	0.208	0.950	0.012	0.198	0.207	0.948
		$\log(\alpha_2)$	0.011	0.093	0.093	0.950	0.009	0.113	0.111	0.946	0.010	0.108	0.111	0.948
40%	$\sigma^2 = 0$	$\log(\alpha_1)$	0.019	0.199	0.200	0.958	0.019	0.207	0.200	0.966	0.019	0.199	0.200	0.958
		$\log(\alpha_2)$	0.002	0.105	0.104	0.948	0.002	0.107	0.104	0.952	0.002	0.105	0.104	0.948
	$\sigma^2 = 0.25$	$\log(\alpha_1)$	0.009	0.187	0.188	0.952	0.009	0.220	0.211	0.954	0.008	0.209	0.211	0.956
		$\log(\alpha_2)$	0.010	0.102	0.099	0.954	0.010	0.112	0.106	0.954	0.011	0.109	0.106	0.954
	$\sigma^2 = 0.6$	$\log(\alpha_1)$	0.008	0.174	0.180	0.948	0.010	0.242	0.225	0.964	0.011	0.222	0.225	0.960
		$\log(\alpha_2)$	0.011	0.098	0.096	0.960	0.010	0.119	0.111	0.960	0.010	0.114	0.110	0.960

[a] Bias: bias of the estimate; SEM: model-based standard error; SEE: empirical standard error; CR: 95% coverage rate. The SEM from regression calibration approach is calculated by the bootstrap method with B = 1000. The likelihood method uses Monte Carlo integration algorithm with N = 5000.

A Appendix

To establish the asymptotic distribution, we need the following conditions, including regularity conditions in Lehmann (1993):

- The measurement errors are independent and identically distributed.
- Given covariates, the survival process is independent of censoring process.
- Let Ω be the parameter space for parameter $\boldsymbol{\eta}$. The true parameter value $\boldsymbol{\eta}^0$ belongs to a compact set within Ω.
- There exists an open subset ω of parameter space Ω that contains the true parameter value $\boldsymbol{\eta}^0$ such that almost all observations with log likelihood $\ell^*(\boldsymbol{\eta})$ have all third derivatives for $\boldsymbol{\eta} \in \omega$. There exist a function M which may depend on $\boldsymbol{\eta}^0$ such that

$$\Big|\frac{\partial^3}{\partial \boldsymbol{\eta}^3}\ell^*(\boldsymbol{\eta})\Big| \leq M \quad \text{for all} \quad \boldsymbol{\eta} \in \omega$$

  and $E_{\boldsymbol{\eta}^0}\big[M\big] < \infty$.

- The log likelihood $\ell^*(\boldsymbol{\eta})$ can be differentiated twice under the integral sign.
- The covariance matrix $I(\boldsymbol{\eta})$ is positive definite for all $\boldsymbol{\eta} \in \omega$.

$$I(\boldsymbol{\eta}) = E\Big[\Big(\frac{\partial \ell^*(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}^{\mathrm{T}}}\Big)\Big(\frac{\partial \ell^*(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\Big)\Big] = E\Big[-\frac{\partial^2 \ell^*(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}^{\mathrm{T}} \partial \boldsymbol{\eta}}\Big].$$

Following the arguments in Lehman (1983, p.409), it can be shown that the maximum likelihood estimator $\widehat{\boldsymbol{\eta}}$ exists and is unique with probability 1. In addition, $\sqrt{n}(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}^0)$ has an asymptotic normal distribution with mean $\mathbf{0}$ and covariance matrix $[I(\boldsymbol{\eta}^0)]^{-1}$.