Journal of Statistical Planning and Inference 153 (2014) 87–99

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Additive mean residual life model with covariate measurement errors

Zhibin Xu [a,*], Liuquan Sun [b], Hong Ji [a]

[a] School of Statistics, Capital University of Economics and Business, Flower-Town, Fengtai District, Beijing 100070, PR China
[b] Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, PR China

ARTICLE INFO

Article history:
Received 19 February 2014
Received in revised form
11 May 2014
Accepted 14 May 2014
Available online 28 May 2014

Keywords:
Additive measurement error model
Error-in-variables
Estimating equation
Replicate measurement
Survival data

ABSTRACT

In this paper, we consider the additive mean residual life model when a proportion of covariates cannot be observed accurately but repeated measurements on covariates are available. The number of repeated measurements is allowed to vary across different subjects. A nonparametric-correction method is developed for inference about regression parameters and the baseline mean residual life function. The resultant estimators are shown to be consistent and asymptotically normal, and consistent standard error estimators are also provided. Simulation studies are carried out to examine the performance of the proposed approach. We illustrate the method by application to data from an HIV clinical trial.

1. Introduction

In many biomedical studies, measurement error issues occur frequently for various reasons such as financial limitations, underdeveloped technology, failings of human memory, substantial instrumental contamination, physical location, and logistical infeasibility. Typical examples include tumor size, dietary measurements, cholesterol level, and blood pressure. A further example involving mismeasured covariates is the HIV clinical trial AIDS Clinical Trial Group (ACTG) 175 data (Hammer et al., 1996). In this clinical trial, it is often of interest to evaluate the prognostic effect of CD4 counts on time to AIDS of HIV-positive subjects. It is well known that the observation of CD4 counts can only be conducted with a substantial measurement error owing to instrumental contamination and temporal biological fluctuation.

Covariate measurement error problems have been extensively studied in the context of survival data for different situations. Sometimes, the covariate of interest is only precisely observed for a subset of study subjects, which is called a validation set, whereas mismeasured covariates are available for all the study subjects (Prentice, 1982; Zhou and Pepe, 1995; Zhou and Wang, 2000; Kulich and Lin, 2000; Chen, 2002; Jiang and Zhou, 2007; Yu and Nan, 2010; Li and Wu, 2010). In other situations, the covariate can never be measured accurately, there exist only repeated measurements on the covariates with measurement errors (Huang and Wang, 2000; Xie et al., 2001; Hu and Lin, 2002, 2004; Gorfine et al., 2004; Sun et al., 2006; Ma and Yin, 2008; Yi and Lawless, 2012). There is a vast literature dealing with measurement error issues and excellent review on methods can be found in Carroll et al. (2006).

In the context of the failure time data with a measurement error in covariates, there exist extensive discussion in the literature with the focus being on the hazards model framework. To the best of our knowledge, there is very little work

* Corresponding author.
E-mail address: xzbzxcasdf@163.com (Z. Xu).

http://dx.doi.org/10.1016/j.jspi.2014.05.007
0378-3758/

concerning the mean residual life model in the presence of covariate measurement error to date. In this paper, we study the additive mean residual life model (Chen and Cheng, 2006; Chen, 2007), where a part of covariates are not accurately ascertainable, there exist only replicate measurements. Moreover, the number of repeated measurements is allowed to vary across different subjects and can even depend on survival time. We apply a nonparametric correction approach to an additive mean residual life model with covariates subject to measurement error. The approach does not impose any additional assumptions on distributions of the measurement errors and covariates.

The rest of this paper is organized as follows. In Section 2, we introduce some notations and review the additive mean residual life model. Then, an inference procedure is developed for both the regression parameters and the baseline mean residual life function under the additive measurement error model. The consistency and asymptotic normality property of the proposed estimator are established in Section 3. In Section 4, we conduct extensive simulation studies to examine the finite sample behavior of the proposed method and illustrate the method by applying it to the ACTG 175 data. Some concluding remarks are provided in Section 5. Sketched proof of the asymptotic properties is outlined in the Appendix.

2. Model and estimation

2.1. Notation and review

The additive mean residual life model for the failure time $\tilde{T}$ associated with covariate $\mathbf{W}$ is specified as follows:

$$m(t|\mathbf{W}) = m_0(t) + \boldsymbol{\beta}_0^T \mathbf{W}, \tag{1}$$

where $\boldsymbol{\beta}_0$ is a p-vector of regression parameter to be estimated and $m_0(t)$ is an unspecified baseline mean residual life function. Let C denote the censoring time which is assumed to be noninformative in the sense that C is conditionally independent of $\tilde{T}$ given W. Under random censorship, we observe $T = \min(\tilde{T}, C)$. To avoid lengthy technical discussion of the tail behavior of the limiting distributions, we further assume that $0 < \tau \equiv \inf\{t: \Pr(T > t) = 0\} < \infty$. The failure indicator is denoted as $\Delta = I(T \le C)$, where $I(\cdot)$ is the indicator function. In addition, we define $Y(t) = I(T \ge t)$ and $N(t) = I(T \le t, \ \Delta = 1)$.

Let $\{T_i, \Delta_i, \mathbf{W}_i\}$, $i=1,\ldots,n$, be n independent and identically distributed copies of $(T, \Delta, \mathbf{W})$. When the covariates can be precisely measured, a consistent estimator of the regression parameter $\boldsymbol{\beta}_0$ can be obtained by solving the following estimating equation (Chen and Cheng, 2006):

$$\mathbf{U}(\boldsymbol{\beta}) = n^{-1} \sum_{i=1}^{n} \int_0^{\tau} \{\mathbf{W}_i - \overline{\mathbf{W}}(t)\}\{\widehat{m}(t;\boldsymbol{\beta}) + \boldsymbol{\beta}^T \mathbf{W}_i\}\, dN_i(t) = 0, \tag{2}$$

where $\overline{\mathbf{W}}(t) = \sum_{i=1}^{n} Y_i(t)\mathbf{W}_i / \sum_{i=1}^{n} Y_i(t)$ and

$$\widehat{m}(t;\boldsymbol{\beta}) = A_n(t)^{-1} \int_t^{\tau} A_n(u) B_n(u;\boldsymbol{\beta})\, du$$

with

$$A_n(t) = \exp\left\{ - \int_0^t \sum_{i=1}^{n} dN_i(u) \Big/ \sum_{i=1}^{n} Y_i(u). \right\}$$

and

$$B_n(t;\boldsymbol{\beta})\, dt = \sum_{i=1}^{n} \{Y_i(t)\, dt - \boldsymbol{\beta}^T \mathbf{W}_i\, dN_i(t)\} \Big/ \sum_{i=1}^{n} Y_i(t).$$

In the following, we suppose that covariates are measured with error. Specifically, a subset of covariates cannot be accurately measured and instead only repeated measurements are available. Furthermore, the number of repeated measurements can vary from subject to subject and can even be correlated with the survival time.

2.2. Inference with mismeasured covariates

Suppose that $\mathbf{W}$ can be partitioned into two parts $(\mathbf{Z}^T, \mathbf{X}^T)^T$, where $\mathbf{Z}$ denotes the p_1 vector of covariates subject to measurement errors, whereas $\mathbf{X}$ is the p_2 vector of covariates which can be measured accurately. We assume that for ith subject, $\mathbf{Z}_i$ is measured n_i times ($n_i \ge 1$) and consider the following classical measurement error model:

$$\mathbf{V}_{ir} = \mathbf{Z}_i + \boldsymbol{\epsilon}_{eir} \ (r = 1, \ldots, n_i),$$

where the error terms $\boldsymbol{\epsilon}_{eir}$ ($i = 1, \ldots, n;\ r = 1, \ldots, n_i$) are iid with mean $\mathbf{0}$ and variance matrix $\boldsymbol{\Sigma}_e$ and are independent of the data $\{\mathbf{W}_i, T_i, \Delta_i\}_{i=1}^n$. Thus, the observed data is of the form

$$\{T_i, \Delta_i, \mathbf{V}_i, \mathbf{X}_i\}_{i=1}^n,$$

where $\mathbf{V}_i = \{\mathbf{V}_{ir};\ r = 1, \ldots, n_i\}$.