Partially Linear Models with Missing Response Variables and Error-prone Covariates

HUA LIANG

Department of Biostatistics and Computational Biology, University of Rochester Medical Center, Rochester, NY 14642, USA

hliang@bst.rochester.edu

SUOJIN WANG AND RAYMOND J. CARROLL

Department of Statistics, Texas A&M University, College Station, TX 77843, USA

sjwang@stat.tamu.edu carroll@stat.tamu.edu

SUMMARY

In this paper, we consider partially linear models in the form $Y = X^{\mathrm{T}}\beta + \nu(Z) + \varepsilon$ when the response variable Y is sometimes missing with missingness probability π depending on (X, Z), and the covariate X is measured with error, where $\nu(z)$ is an unspecified smooth function. The missingness structure is therefore missing not at random (NMAR), rather than the usual missing at random (MAR). We propose a class of semiparametric estimators for parameter of interest β, as well as for the population mean $E(Y)$. The resulting estimators are shown to be consistent and asymptotically normal under general assumptions. To construct a confidence region for β, we also propose an empirical likelihood based statistic, which is shown to have an asymptotic chi-squared distribution. The proposed methods are applied to analyze an AIDS clinical trial data set. A simulation study is also reported to illustrate our approach.

Key Words: Confidence Region; Empirical likelihood; Estimating equation; Measurement error; Missing data; Missing not at random; Nonparametric regression; Partially linear models; Semiparametric estimation.

Short title: Partially Linear Measurement Error Models

1 INTRODUCTION

The partially linear model assumes that the response variable Y depends on variable X in a linear way but is related to another independent variable Z in an unspecified form, i.e., it is of the form

$$Y = X^{\mathrm{T}}\beta + \nu(Z) + \varepsilon, \tag{1}$$

where X is a p-vector covariate, Z is a scalar covariate, the function $\nu(\cdot)$ is unknown, and the model error ε has mean zero conditional on (X, Z).

There is a substantial literature on kernel-based methods for partially linear models and their generalizations, see for example, Engle et al. (1986), Speckman (1988), Robinson (1988), Severini & Staniswalis (1994), Zeger & Diggle (1994), Opsomer & Ruppert (1999) and Härdle, Liang & Gao (2000), among many others. Liang, Härdle, & Carroll (1999) considered model (1) with error-prone X. Recently, missing data issues have been considered, with Liang, Wang, Robins & Carroll (2004) considering the case that X in (1) is missing at random, while Wang, Linton & Härdle (2004) consider the case that the response Y is missing at random.

In this paper, we consider the missing response case, but in addition we allow some of the components of X to be measured with error. Our motivation is from AIDS clinical trials, where the response, variable viral load RNA, can be missing. In addition, the covariates, CD4 measurements, are measured with error. Measurement error in predictors causes a bias in the estimated regression coefficient. While Liang, et al. considered the measurement error problem, they did not allow for missing responses.

We deal with the case that X is measured with additive errors in the following sense: we cannot observe X directly but can observe a surrogate W related to X by

$$W = X + U. \tag{2}$$

If $\delta = 1$ indicates that Y is observed and $\delta = 0$ indicates that Y is missing, we assume that the measurement error U is independent of (Y, Z, X, δ) and with $E(U) = 0$, cov$(U) = \Sigma_{uu}$, and $E\{U(UU^{\mathrm{T}}\beta)^{\mathrm{T}}\} = 0$ for all β. We first assume that Σ_{uu} is known, and later extend the results to the general case.

It is important to stress that under the setting considered in this paper, see (3) below, the missingness of Y is allowed to depend on (X, Z), but not otherwise on W. Since the true X is not observable, Y is therefore not missing at random (NMAR). Since we will make no further assumptions, such as on the distribution of X or on the missing data probabilities, what we are dealing with here is conceptually quite different from most studies of missing data in which missing at random or missing completely at random is assumed.

The paper is arranged as follows. In Section 2 we define the missing data mechanism for the problem and propose our estimation methods. These are weighted estimating equations leading to computationally simple methods. Asymptotic theory and inference is also derived. In Section 3, following Wang, et al. (2004), we give three equivalent estimators of the mean $E(Y)$, which are shown to be semiparametric efficient in certain circumstances. To improve confidence intervals based on the standard error estimates, in Section 4 we develop an empirical log-likelihood ratio statistic, and show that it has an asymptotic chi-square distribution. Section 5 contains the results of a small simulation study, while Section 6 presents the results of our statistical analysis of an AIDS study. Some concluding remarks are given in Section 7. All proofs are given in the Appendix.

2 ESTIMATION AND MAIN RESULTS

2.1 Known Measurement Error Covariance Matrix

As described previously, we assume that if X were observed, the missing data mechanism follows the missing at random mechanism in the sense that

$$\text{pr}(\delta = 1|X, Z, Y) = \text{pr}(\delta = 1|X, Z) = \pi(X, Z), \tag{3}$$

for some unknown $\pi(X, Z)$. As stated previously, we also assume that the measurement errors U are independent of (Y, Z, X, δ), so that $\text{pr}(\delta = 1|X, Z, Y, W) = \pi(X, Z)$. Furthermore, assume that $\{(Y_i, W_i, Z_i, \delta_i), i = 1, \cdots, n\}$ are independent and identically distributed.

In this paper, we use local constant smoothers with fixed bandwidths for simplicity in presenting the derivations of the theoretical results: extensions to local polynomial estimation are straightforward, with no change in the limiting distribution of the estimate of β. In what follows, we denote $A^{\otimes 2} = AA^{\mathrm{T}}$. Also, define $m_x(z) = E(\delta X|Z = z)/E(\delta|Z = z)$, $m_w(z) = E(\delta W|Z = z)/E(\delta|Z = z)$ and $m_y(z) = E(\delta Y|Z = z)/E(\delta|Z = z)$. Let $\widetilde{X}_i = X_i - m_x(Z_i)$, $\widetilde{W}_i = W_i - m_w(Z_i)$, $\widetilde{Y}_i = Y_i - m_y(Z_i)$, and denote $\Sigma_{X|Z} = \text{cov}(\widetilde{X}\delta)$.

Note that $\delta Y = \delta X^{\mathrm{T}}\beta + \delta\nu(Z) + \delta\varepsilon$. From our assumptions, it follows that $E(\delta Y|Z) = E(\delta X^{\mathrm{T}}|Z)\beta + E(\delta|Z)\nu(Z)$. If X is observed, and $m_x(z)$ and $m_y(z)$ are known, one can obtain a least squares type estimator of β as

$$\left[\sum_{i=1}^{n} \delta_i \{X_i - m_x(Z_i)\}^{\otimes 2}\right]^{-1} \left[\sum_{i=1}^{n} \delta_i \{X_i - m_x(Z_i)\}\{Y_i - m_y(Z_i)\}\right].$$

It is easily shown that this estimator is consistent, asymptotically normal, etc.; see Wang, et al. (2004).

The formula above cannot be applied directly when X is measured with error, and $m_x(z)$ and $m_y(z)$ are unknown. However, by our assumptions, $E(\delta W|Z) = E(\delta X|Z)$. We thus propose a correction for attenuation estimator of β:

$$\widehat{\beta}_n = \left(\sum_{i=1}^{n} \delta_i \left[\{W_i - \widehat{m}_w(Z_i)\}^{\otimes 2} - \Sigma_{uu}\right]\right)^{-1} \left[\sum_{i=1}^{n} \delta_i \{W_i - \widehat{m}_w(Z_i)\}\{Y_i - \widehat{m}_y(Z_i)\}\right], \qquad (4)$$

where $\widehat{m}_w(z)$ and $\widehat{m}_y(z)$ are nonparametric regression estimators. Let $K(\cdot)$ be a symmetric density function, let h be a suitable bandwidth and define$K_h(z) = K(z/h)/h$. These estimators take the form

$$\widehat{m}_w(z) = \frac{\sum_{i=1}^{n} K_h(Z_i - z)\delta_i W_i}{\sum_{i=1}^{n} \delta_i K_h(Z_i - z)} \text{ and } \widehat{m}_y(z) = \frac{\sum_{i=1}^{n} K_h(Z_i - z)\delta_i Y_i}{\sum_{i=1}^{n} \delta_i K_h(Z_i - z)}. \qquad (5)$$

Remark 1 Alternative estimators are readily constructed, but generally suffer from complications. For example, since $Y - E(Y|Z) = \{X - E(W|Z)\}^{\mathrm{T}}\beta + \varepsilon$, an obvious approach is to estimate $E(W|Z)$ and $E(Y|Z)$ using all the data. The former is easy: any nonparametric regression will do. The latter though is problematic, because of the missing responses, the possibility that missingness depends on X, and the fact that X is unobserved. There does not appear to be an easy way to estimate $E(Y|Z)$ consistently under the current conditions. Note that one of the most important features of the proposed approach is that, by using the standard measurement error model (2), it can handle the not missing at random case with ease and still provide $\sqrt{n}$-consistent estimators, as shown in the theorems that follow.

Before presenting our first main result, we note that throughout the paper we make some general assumptions that are listed in the Appendix.

Theorem 1 Assume that $\{(Y_i, W_i, Z_i, \delta_i),\ i = 1, \cdots, n\}$ are independent and identically distributed. Under Assumption 1 in the Appendix, $n^{1/2}(\widehat{\beta}_{\mathrm{n}} - \beta)$ is asymptotically normally distributed with mean 0 and covariance matrix $\Sigma_\beta = \Sigma_{X|Z}^{-1}\Gamma\Sigma_{X|Z}^{-1}$, where

$$\Gamma = E[\delta\{(\varepsilon - U^{\mathrm{T}}\beta)\widetilde{X}\}^{\otimes 2}] + E(\delta U U^{\mathrm{T}}\varepsilon^2) + E[\delta\{(UU^{\mathrm{T}} - \Sigma_{uu})\beta\}^{\otimes 2}].$$

The proof of Theorem 1 is given in the Appendix.

Remark 2 In typical nonparametric kernel regression, bandwidth selection plays a key role in the performance of nonparametric estimators in terms of their bias and variance. In partially linear models, β is of main interest, and $\nu(z)$ is a nuisance function. Based on Assumption 1 (b), only the rate of order $n^{-1/5}$ is needed to lead to the same limit distribution for estimating β. In implementing our proposed estimation procedure, we adopt Ruppert, Sheather, and Wand's (1995) approach to search for the bandwidth. Our limited experience indicates that the numerical performance of the resulting estimators of β is stable around the selected bandwidth.

Checking the proof of Theorem 1, we see that Σ_β can be estimated via a standard sandwich method as follows. Let

$$\widehat{\Sigma}_{X|Z} = n^{-1}\sum_{i=1}^{n}\delta_i[\{W_i - \widehat{m}_w(Z_i)\}^{\otimes 2} - \Sigma_{uu}];$$

$$\widehat{\Gamma} = n^{-1}\sum_{i=1}^{n}\delta_i\Big(\{W_i - \widehat{m}_w(Z_i)\}[Y_i - \widehat{m}_y(Z_i) - \{W_i - \widehat{m}_w(Z_i)\}^{\mathrm{T}}\widehat{\beta}_{\mathrm{n}}] + \Sigma_{uu}\widehat{\beta}_{\mathrm{n}}\Big)^{\otimes 2}$$

and $\widehat{\Sigma}_\beta = \widehat{\Sigma}_{X|Z}^{-1}\widehat{\Gamma}\widehat{\Sigma}_{X|Z}^{-1}$. Then it is easily shown that $\widehat{\Sigma}_\beta$ is a consistent estimator of Σ_β.

2.2 Estimated Measurement Error Covariance Matrix

The covariance matrix Σ_{uu} is generally unknown and needs to be estimated. The usual method of doing so (Carroll, et al., 1995, Chapter 3) is by partial replication, so that we observe $W_{ij} = X_i + U_{ij},\ \ j = 1, ..., m_i$. For notational simplicity, we assume that $m_i \equiv 2$. Extension to more general settings is straightforward, see Liang, et al. (1999) for a related discussion. Let $\overline{W}_i$ be the sample mean of the replicates W_{ij}. A consistent, unbiased method of moments estimate for Σ_{uu} is

$$\widehat{\Sigma}_{uu} = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{2}(W_{ij} - \overline{W}_i)(W_{ij} - \overline{W}_i)^{\mathrm{T}}.$$

The corresponding estimator of β is

$$\widehat{\beta}_{n,2} = \left(\sum_{i=1}^{n} \delta_i \left[\{\overline{W}_i - \widehat{m}_{\overline{w}}(Z_i)\}^{\otimes 2} - (1/2)\widehat{\Sigma}_{uu}\right]\right)^{-1} \left[\sum_{i=1}^{n} \delta_i \{\overline{W}_i - \widehat{m}_{\overline{w}}(Z_i)\}\{Y_i - \widehat{m}_y(Z_i)\}\right], \quad (6)$$

where $\widehat{m}_{\overline{w}}(z)$ is the local constant estimate of $m_w(z)$ based on the data $\{(\overline{W}_i, Z_i), i = 1, \ldots, n\}$. We now present the following theorem.

Theorem 2 Under the general conditions of Theorem 1, the estimator $\widehat{\beta}_{n,2}$ given in (6) is consistent and asymptotically normal with covariance matrix $\Sigma_{X|Z}^{-1}\Gamma^{*}\Sigma_{X|Z}^{-1}$, where

$$\Gamma^{*} = E[\delta\{(\varepsilon - \overline{U}^{\mathrm{T}}\beta)\widetilde{X}\}^{\otimes 2}] + E(\delta\overline{U}\overline{U}^{\mathrm{T}}\varepsilon^2) + E[\delta\{(\overline{U}\overline{U}^{\mathrm{T}} - \Sigma_{uu}/2)\beta\}^{\otimes 2}].$$

By a straightforward but tedious derivation, Theorem 2 can be proved in a manner similar to Theorem 1: we omit the details.

The standard error estimates can also be derived. A consistent estimate of $\Sigma_{X|Z}$ in this case is defined as

$$n^{-1}\sum_{i=1}^{n} \delta_i \left[\left\{\overline{W}_i - \widehat{m}_{\overline{w}}(Z_i)\right\}^{\otimes 2} - (1/2)\widehat{\Sigma}_{uu}\right].$$

The Γ^{*} can be estimated as follows. Let

$$\begin{aligned} R_i = \{\overline{W}_i - \widehat{m}_{\overline{w}}(Z_i)\}\left[Y_i - \widehat{m}_y(Z_i) - \{\overline{W}_i - \widehat{m}_{\overline{w}}(Z_i)\}^{\mathrm{T}}\widehat{\beta}_{n,2}\right] \\ +(1/2)\left\{(W_{i1} - W_{i2})^{\otimes 2} - \widehat{\Sigma}_{uu}\right\}\widehat{\beta}_{n,2}. \end{aligned}$$

Then a consistent estimate of Γ^{*} is the sample covariance matrix of the $R_i\delta_i$'s. See Liang, Härdle, and Carroll (1999) for a detailed discussion.

3 Estimation of the Mean $E(Y)$

It is of interest to estimate the mean $E(Y) = \theta$. Cheng (1994) studied this problem in the purely nonparametric regression case, while Wang, et al. (2004) studied the partially linear model with X observed. Here we construct three estimators of θ when X is not observed. The methods are analogous to those of Wang, et al. in the case that X is observed. Like them, we show that the three estimators are asymptotically equivalent.

In a manner similar to Cheng (1994), we can construct two estimators of θ as follows:

$$\widehat{\theta}_{n,ave} = n^{-1}\sum_{i=1}^{n}\delta_i Y_i + n^{-1}\sum_{i=1}^{n}(1-\delta_i)\{W_i^{\mathrm{T}}\widehat{\beta}_n + \widehat{\nu}_n(Z_i)\},$$

or

$$\widehat{\theta}_{n,est} = n^{-1}\sum_{i=1}^{n}\{W_i^{\mathrm{T}}\widehat{\beta}_n + \widehat{\nu}_n(Z_i)\},$$

where $\widehat{\nu}_n(z) = \widehat{m}_y(z) - \widehat{m}_w^{\mathrm{T}}(z)\widehat{\beta}_{\mathrm{n}}$ is a nonparametric regression estimator of $\nu(z)$ based on the completely observed data of $\{(Z_i, Y_i - W_i^{\mathrm{T}}\widehat{\beta}_{\mathrm{n}}), i = 1, \cdots, n\}$. One can easily show that $\widehat{\nu}_n(z) - \nu(z) = o_p(n^{-1/3})$ in a way similar to Liang, Härdle, and Carroll (1999). This rate satisfies our assumption to establish the asymptotic normality of the estimators of θ.

Denote $s_n(z) = \sum_{i=1}^{n}\delta_i K_h(Z_i - z)/\sum_{i=1}^{n}K_h(Z_i - z)$, $s(z) = E(\delta|Z = z)$ and $P(Z,\delta) = \delta/s(Z)$. We define a third estimator of θ as

$$\widehat{\theta}_{n,wei} = n^{-1}\sum_{i=1}^{n}\frac{\delta_i}{s_n(Z_i)}Y_i + n^{-1}\sum_{i=1}^{n}\left\{1 - \frac{\delta_i}{s_n(Z_i)}\right\}\{W_i^{\mathrm{T}}\widehat{\beta}_n + \widehat{\nu}_n(Z_i)\}.$$

Note that if we try to substitute $s_n(z)$ by an estimator of $\pi(x,z)$, a problem arises because X is measured with error, so that exact X is not available for estimating $\pi(X,Z)$. In the following theorem, we give asymptotic normality of the three estimators, showing that they are asymptotically equivalent.

Theorem 3 In addition to the assumptions of Theorem 1, assume that $nh^4 \to 0$. Then $n^{1/2}(\widehat{\theta}_{n,\bullet} - \theta)$ has an asymptotically normally distribution with mean 0 and variance $E[P(Z,\delta)\varepsilon + \{1 - P(Z,\delta)\}U^{\mathrm{T}}\beta + E(\widetilde{W}^{\mathrm{T}})\Sigma_{X|Z}^{-1}\delta\{\widetilde{W}(\varepsilon - U^{\mathrm{T}}\beta) + \Sigma_{uu}\beta\}]^2 + E\{X^{\mathrm{T}}\beta + \nu(Z) - \theta\}^2$, where $\bullet$ indicates "ave", "est", or "wei".

4 INFERENCE BASED ON EMPIRICAL LIKELIHOOD PRINCIPLE

Based on our estimators of the covariance or its bootstrap version, one can give a confidence region for either β or $\theta = E(Y)$. Although we have confirmed that the estimator $\widehat{\Sigma}_\beta$ given in Section 2 is consistent, its finite-sample behavior may be affected by the need to plug in several estimated terms. Furthermore, the confidence region derived by this procedure is based on a normal approximation, which may be optimistic in small samples. An alternative method is to use the empirical likelihood principle, see Owen (1988, 1990, 1991, 2000), Qin (1994, 1999), Qin and Lawless (1994) and Chen (1993, 1994). In the remainder of this section, we assume ε_i are independent and identically distributed and independent of (W_i, Z_i). We propose our empirical likelihood ratio statistic, and show that the statistic is asymptotically chi-squared distributed. We need only to study the empirical likelihood based confidence interval for β since the situation for θ is similar and simpler.

Let F be the distribution function which assigns probability p_i at points (W_i, Y_i, Z_i). Then $\sum_{i=1}^n p_i = 1$ and $p_i \geq 0$ for each i. Our semiparametric empirical likelihood ratio is defined as follows. Note that $E[\delta\{\widetilde{W}(\widetilde{Y} - \widetilde{W}^{\mathrm{T}}\beta) + \Sigma_{uu}\beta\}] = 0$. The empirical likelihood ratio function for β may be defined as

$$\mathcal{R}(\beta) = \max\left\{\prod_{i=1}^n np_i \;\Big|\; \sum_{i=1}^n p_i\delta_i\{\widetilde{W}_i(\widetilde{Y}_i - \widetilde{W}_i^{\mathrm{T}}\beta) + \Sigma_{uu}\beta\} = 0, p_i \geq 0, \sum_{i=1}^n p_i = 1\right\},$$

if $m_w(z)$ and $m_y(z)$ are known. In our model setting, an modified empirical likelihood ratio function is defined as

$$\begin{aligned}\mathcal{R}_n(\beta) = \max\Bigg\{\prod_{i=1}^n np_i \;\Big|\; &\sum_{i=1}^n p_i\delta_i\Big(\{W_i - \widehat{m}_w(Z_i)\}[Y_i - \widehat{m}_y(Z_i) - \\ &\{W_i - \widehat{m}_w(Z_i)\}^{\mathrm{T}}\beta] + \Sigma_{uu}\beta\Big) = 0, p_i \geq 0, \sum_{i=1}^n p_i = 1\Bigg\}. \qquad (7)\end{aligned}$$

Theorem 4 Under Assumption 1, $-2\log\{\mathcal{R}_n(\beta)\}$ converges to a chi-squared distribution with p degrees of freedom.

Based on this result, a confidence region of β can be given as $\{\beta : -2\log\{\mathcal{R}_n(\beta)\} \leq c_\alpha\}$, where c_α denotes the α quantile of the chi-squared distribution. When Σ_{uu} is unknown, we need replication data in the usual way. Assuming the special case of $m_i \equiv 2$ as in Section 2, we can then replace W_i by $\overline{W}_i$ and $\Sigma_{uu}\beta$ by $1/2\widehat{\Sigma}_{uu}\beta$. The resulting statistic still has the property given in Theorem 4. A justification of this last assertion can be easily obtained by using the fact that

$$E[\delta\{\widetilde{\overline{W}}_i(\widetilde{Y}_i - \widetilde{\overline{W}}_i^{\mathrm{T}}\beta) + 1/2\widehat{\Sigma}_{uu}\beta\}] = 0,$$

where $\widetilde{\overline{W}}_i = \overline{W}_i - m_w(Z_i)$.

5 A SIMULATION STUDY

To evaluate the performance of the proposed methods, we conducted a small scale simulation experiment. We generated $n = 100$ and $n = 500$ observations from model (1), assuming that $Y|X, Z \sim$ Normal$\{\beta_0 + \beta_1 X + \nu(Z), \sigma^2(X, Z)\}$ and the probability of Y being observed equals pr$(\delta = 1|Y, X, Z) = \Phi\{\alpha_0 + \alpha_1 X + \nu_1(Z)\}$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function. We also assume that the measurement error follows $W = X + U$, where $U \sim$ Normal$(0, 0.2)$. In our simulations, we set $\alpha_0 = \beta_0 = 0$, $\beta_1 = 1$, $\alpha_1 = 2$, $X \sim$ Uniform$(0, 1)$, $Z \sim$ Uniform$(0, 1)$ independent of X, and $\nu(z) = 4\{\exp(-3.25z) - 4\exp(-6.5z) + 3\exp(-9.75z)\}$. We considered four cases.

Case 1: $\nu_1(z) = 0.75z$ and $\sigma^2(x, z) = 0.25$;

Case 2: $\nu_1(z) = \sin(z^2)$ and $\sigma^2(x, z) = 0.25$;

Case 3: $\nu_1(z) = 0.75z$ and $\sigma^2(x, z) = 0.1 * \{\sin^2(2\pi x^3) + 0.5z + 0.3\}$. This case is meant to see the effect of heteroscedastic error on the estimators and confidence intervals.

Case 4: $\nu_1(z) = 0.75z$ and the error ε follows $0.25^2(\mathcal{X}_2^2 - 2)$, where $\mathcal{X}_2^2$ is a chi-squared variable with 2 degrees of freedom. This case is meant to see the effect of asymmetric error on the estimators and confidence intervals.

In our nonparametric estimation procedure, we selected bandwidths as in Remark 2. We used the quartic kernel: $K(u) = 15/16(1-u^2)^2 I_{(|u|\le 1)}$. We generated $1,000$ data sets in each of the four cases. For each case, approximately 35% of the Y's are missing. To estimate the variance of U, we generated double samples of W. We computed the naive and correction-for-attenuation estimates of the parametric components, and their asymptotic and empirical likelihood-based confidence intervals.

The results are given in Table 1. Column “Estimate” gives the average of 1000 estimated coefficients based on the naive and our proposed methods; Column “CI(ME)” gives the confidence intervals using the empirical likelihood and normal approximation methods when the measurement errors are accounted for. The lower and upper values are the averages of 1000 simulated corresponding lower and upper values; Column “Coverage (ME)” gives the corresponding coverage probabilities of the 1000 data sets. We summarize our findings as follows. The results are basically in accord with the theory. The impact of the measurement errors on the estimates is substantial. When ignoring measurement errors, the estimates are significantly biased and attenuate to zero. For moderate sample size, the empirical likelihood-based confidence intervals appear to be superior to these based on the normal approximation. The improvement is better when the error is non-normal or its variance is not constant.

6 Analysis of a Dataset from an AIDS Study

In this section we present an analysis of the pediatric AIDS clinical trial group (PACTG 338) study. One of the purposes of this study is to investigate the effectiveness of antiretroviral (anti-HIV) medicines, and to see how increasing CD4 cell counts decrease the amount of HIV in the blood (HIV viral load). We are interested in understanding the pathogenesis of HIV infection and in evaluation of antiretroviral therapies by characterizing the relation between viral load and CD4 cell counts. Our preliminary investigations suggested that viral load depends linearly on CD4 cell count but nonlinearly on treatment time. See Liang et al. (2004) for a related discussion. We therefore model the relation between viral load and CD4 cell counts by using model (1), in which Y represents viral load, X CD4 cell counts, and Z treatment time. Model (1) was applied by Zeger and Diggle (1994) to investigate the relation between CD4 cell counts and other covariates.

The PACTG 338 study consists of 297 children, who were clinically stable and who had not had prior treatment with protease inhibitors to a 2- or 3-drug protease inhibitor containing regimen (ritonavir plus 1 or 2 nucleoside analogs) or to a dual nucleoside analog regimen, with 2287 observations, of which 17.6% viral load RNA were missing. See Nachman, et al. (2000) for a detailed explanation of PACTG 338. The management of HIV infected patients mainly includes monitoring their CD4 cell counts, which reflect body immunity, and HIV viral load, a useful virologic marker. CD4 cell counts are used to follow response to HIV medications, as a measure of adherence to treatment and most importantly to guide decisions regarding opportunistic infection prophylaxis. Some patients may fail to go to clinical trial centers for a HIV viral load measurement when they feel that their immunity is strong enough or too weak. Therefore, the assumption that the missing RNA levels depends on true CD4 cell counts and not measured counts and treatment time appears to be at least somewhat reasonable.

In our data analysis, we used the working independence assumption for the error terms, ε, in the model. The original times ranged from 0 to 84 (in weeks). To reduce the marked skewness of CD4 cell counts, the variation of viral load, and the sparsity of treatment times, we take log-transformation for all three variables. We used the same kernel function as in the simulation study (Section 5), and obtained a bandwidth of $h = 0.124$ in the same way there. We assume that the measurement errors U_{ij} were independent and normally distributed with mean zero and variance σ_u^2. To estimate the measurement error variance σ_u^2, one needs to have validation or replication data. However, neither kind of data is available, so that similar to Lin and Carroll (2000), we conducted a sensitivity analysis by taking σ_u^2 to be the 1/4 and 1/2 of the variance of W.

We applied the methods proposed in Section 2 and Section 4 to the data set, assuming $\sigma_u^2 = 0$ (naive estimate ignoring measurement error), $\sigma_u^2 = 0.068$, and $\sigma_u^2 = 0.135$. For the parameter β, we give its estimated value, along with asymptotic and empirical likelihood confidence intervals (CI) in Table 2. The estimated values corresponding to $\sigma_u^2 = 0.068$ and $\sigma_u^2 = 0.135$ increased 31.2% and 48.8%, respectively, compared to the naive estimate, while the confidence intervals were widened accordingly, which reflects the increased uncertainty due to the measurement errors. As expected, when the possible measurement errors were taken into account, we found a somewhat stronger negative association between viral load and CD4 cell counts. The confidence intervals were obtained by 200 bootstrap replications, where in the bootstrap patients were resampled.

The curve of the estimated nonparametric function of treatment time and the corresponding confidence bands under the naive framework are shown in Figure 1. The curves for other two cases (not shown here) have a similar pattern as the plot in Figure 1. The confidence bands were obtained by 200 bootstrap replications, in which patients were resampled. The plot indicates that the viral load RNA levels rapidly decrease after initial antiviral treatment and become flat and even rebound a little bit. But finally the estimated curve goes down rapidly.

7 DISCUSSION

In this paper, we have studied a class of easily computable estimators in partially linear models with missing response variables and error-prone covariates. While the missingness of Y in our setting is not at random, our proposed estimators have been shown to be $\sqrt{n}$-consistent and asymptotically normally distributed. We have also established empirical likelihood inference for this problem, with numerical results that suggest this type of inference is preferred over asymptotic normal approximations. The estimation methods and normal limit distributions are readily extended to longitudinal and repeated measures contexts, if one uses working independence, i.e., ignores the correlation structure when computing the estimator, but uses it in computing asymptotic covariance matrices.

The proposed estimators are based on the observed data, but exclude the observed covariates (W, Z) when Y is missing except that all the Ws are used to estimate Σ_{uu} when its estimation is desired. Although we have not derived the efficiency bound for the estimator of β, we conjecture that little gain, if any, can be obtained if we include those observations (W, Z) associated with missing Y's. See Bickel et al. (1993, page 146) for a related result.

ACKNOWLEDGMENT

Liang's research was partially supported by two grants from the National Institute of Allergy and Infectious Diseases. Wang and Carroll's research was partially supported by a grant from the National Cancer Institute (CA–57030), and by the Texas A&M Center for Environmental and Rural Health via a grant from the National Institute of Environmental Health Sciences (P30-ES09106). The work of Raymond Carroll occurred during a visit to the Centre of Excellence for Mathematics and Statistics of Complex Systems at the Australian National University, whose support is gratefully acknowledged.

APPENDIX: ASSUMPTIONS AND PROOFS

A.1 Assumptions

We list the following conditions which are assumed to hold throughout the paper.

Assumption 1 (a) $E\{\pi(X, Z)\widetilde{X}\widetilde{X}^{\mathrm{T}}\}$ is a positive-definite matrix, $E(\varepsilon|X, Z) = 0$, and $E(|\varepsilon|^3|X, Z) < \infty$;

(b) The bandwidths in estimating $m_w(z)$ and $m_y(z)$ are of order $n^{-1/5}$;

(c) $K(\cdot)$ is a bounded symmetric density function with compact support and satisfies that $\int K(u)du = 1$, $\int K(u)udu = 0$ and $\int u^2K(u)du = 1$;

(d) The density function of Z, $f_Z(z)$, is bounded away from zero and has bounded continuous second (partial) derivatives;

(e) $m_y(z)$ and $m_w(z)$ have bounded and continuous second derivatives;

(f) The probability function $\pi(x,z)$ is bounded away from zero on the support of (X,Z), and has bounded continuous second partial derivatives;

(g) $E(\|U\|^3) < \infty$.

A.2 Proof of Theorem 1

We first point out the following fact, which can easily be shown by Assumption 1 (b)-(e),

$$\widehat{m}_w(z) - m_w(z) = o_p(n^{-1/4}) \text{ and } \widehat{m}_y(z) - m_y(z) = o_p(n^{-1/4}). \tag{A.1}$$

We use these two equations in what follows. Define

$$\Psi(m_w, m_y, \beta, Y, W, Z, \delta) = \left(\{W - m_w(Z)\}\left[Y - m_y(Z) - \{W - m_w(Z)\}^{\mathrm{T}}\beta\right] + \Sigma_{uu}\beta\right)\delta. \tag{A.2}$$

Our estimator $\widehat{\beta}_{\mathrm{n}}$ solves the estimating equations

$$0 = \sum_{i=1}^{n} \Psi(\widehat{m}_w, \widehat{m}_y, \beta, Y_i, W_i, Z_i, \delta_i).$$

Let

$$D(m_w^* - m_w, m_y^* - m_y, \beta, Y, W, Z, \delta) = \frac{\partial \Psi}{\partial m_w}(m_w^* - m_w) + \frac{\partial \Psi}{\partial m_y}(m_y^* - m_y),$$

where the partial derivatives are the Frechet partial derivatives. It is easy to obtain that

$$\frac{\partial \Psi}{\partial m_w} = [-Y + m_y(Z) - 2\{W - m_w(Z)\}^{\mathrm{T}}\beta]\delta \text{ and } \frac{\partial \Psi}{\partial m_y} = -\{W - m_w(Z)\}\delta.$$

A direct calculation yields that $E\left(\frac{\partial \Psi}{\partial m_w}|Z\right) = 0$ a.e. and $E\left(\frac{\partial \Psi}{\partial m_y}|Z\right) = 0$ a.e. In addition,

$$\begin{aligned}&\|\Psi(m_w^*, m_y^*, \beta, Y, W, Z, \delta) - \Psi(m_w, m_y, \beta, Y, W, Z, \delta) - D(m_w^* - m_w, m_y^* - m_y, \beta, Y, W, Z, \delta)\|\\ &\quad = O\left(\|m_w^* - m_w\|^2 + \|m_y^* - m_y\|^2\right),\end{aligned} \tag{A.3}$$

where $\|h\|$ denotes a norm for the function h, such as Sobolev norm, i.e, supremum norm for a function and its derivatives. Equation (A.3) is Newey's (1994) assumption 5.1(i). Furthermore, (A.1) ensures assumption 5.1(ii) of Newey. Again his assumption 5.2 holds by the expression of $D(\bullet, \beta, Y, W, Z, \delta)$. In addition, it follows from the above statements that for any (m_w^*, m_y^*),

$$E\left\{D(m_w^* - m_w, m_y^* - m_y, \beta, Y, W, Z, \delta)\right\} = 0,$$

thus verifying Newey's assumption 5.3: his $\alpha(z) = 0$, according to his discussion just above his Lemma 5.1. By that lemma, it follows that $\widehat{\beta}_{\mathrm{n}}$ has the same limit distribution as the solution, call $\widetilde{\beta}_n$, to the equation

$$0 = \sum_{i=1}^{n} \Psi(m_w, m_y, \beta, Y_i, W_i, Z_i, \delta_i).$$

A direct derivation gives that

$$\widetilde{\beta}_n = \left\{n^{-1}\sum_{i=1}^{n} \widetilde{W}_i \widetilde{W}_i^{\mathrm{T}} \delta_i - n^{-1}\sum_{i=1}^{n} \delta_i \Sigma_{uu}\right\}^{-1} n^{-1}\sum_{i=1}^{n} \widetilde{W}_i \widetilde{Y}_i \delta_i.$$

It follows that

$$n^{1/2}(\widetilde{\beta}_n - \beta) = \Delta_n^{-1} n^{-1/2} \sum_{i=1}^{n} \delta_i(\varepsilon_i \widetilde{X}_i - \widetilde{X}_i U_i^{\mathrm{T}}\beta + U_i\varepsilon_i - U_i U_i^{\mathrm{T}}\beta + \Sigma_{uu}\beta),$$

where

$$\Delta_n = n^{-1}\sum_{i=1}^{n} \widetilde{X}_i \widetilde{X}_i^{\mathrm{T}} \delta_i + n^{-1}\sum_{i=1}^{n} U_i U_i^{\mathrm{T}} \delta_i - n^{-1}\sum_{i=1}^{n} \delta_i \Sigma_{uu} + o_p(1) \to \Sigma_{X|Z}.$$

Therefore $\widetilde{\beta}_n$ has the same limit distribution as described in the statement of Theorem 1. This completes the proof.

A.3 Proof of Theorem 3

We will prove the theorem for $\widehat{\theta}_{n,wei}$ because the proofs for the other two cases are similar.

Define

$$\Phi(m_w, m_y, s, \beta, Y, W, Z, \delta, \theta) = \frac{\delta}{s(Z)}(Y - \theta) + \left\{1 - \frac{\delta}{s(Z)}\right\}\{W^{\mathrm{T}}\beta + m_y(Z) - m_w^{\mathrm{T}}(Z)\beta - \theta\}$$

The estimator $\widehat{\theta}_{n,wei}$ is the solution of the estimating equation

$$n^{-1}\sum_{i=1}^{n} \Phi(\widehat{m}_w, \widehat{m}_y, s_n, \widehat{\beta}_n, Y_i, W_i, Z_i, \delta_i, \theta) = 0.$$

Using the same approach as that in the proof of Theorem 1, we can show that $\widehat{\theta}_{n,wei}$ has the same distribution as the solution, $\widetilde{\theta}_n$ say, of the equation

$$n^{-1}\sum_{i=1}^{n}\Phi(m_w, m_y, s, \widehat{\beta}_n, Y_i, W_i, Z_i, \delta_i, \theta) = 0.$$

Note that

$$\begin{aligned}\widetilde{\theta}_n - \theta &= n^{-1}\sum_{i=1}^{n}\{X_i^{\mathrm{T}}\beta + \nu(Z_i) - \theta\} + n^{-1}\sum_{i=1}^{n}\frac{\delta_i}{s(Z_i)}\varepsilon_i \\ &\quad + n^{-1}\sum_{i=1}^{n}\left\{1 - \frac{\delta_i}{s(Z_i)}\right\}W_i^{\mathrm{T}}(\widehat{\beta}_n - \beta) + n^{-1}\sum_{i=1}^{n}\left\{1 - \frac{\delta_i}{s(Z_i)}\right\}U_i^{\mathrm{T}}\beta \\ &\quad - n^{-1}\sum_{i=1}^{n}\left\{1 - \frac{\delta_i}{s(Z_i)}\right\}m_w^{\mathrm{T}}(Z_i)(\widehat{\beta}_n - \beta).\end{aligned}$$

Because $E\left[\left\{1 - \frac{\delta_i}{s(Z_i)}\right\}m_w^{\mathrm{T}}(Z_i)\right] = 0$ and $\widehat{\beta}_n - \beta = O_p(n^{-1/2})$, the last term is $O_p(n^{-1})$.

Furthermore, it follows from the proof of Theorem 1 that

$$n^{1/2}(\widehat{\beta}_n - \beta) = \Sigma_{X|Z}^{-1} n^{-1/2}\sum_{i=1}^{n}\delta_i\{\widetilde{W}_i(\varepsilon_i - U_i^{\mathrm{T}}\beta) + \Sigma_{uu}\beta\} + o_p(1)$$

and $n^{-1}\sum_{i=1}^{n}\left\{1 - \frac{\delta_i}{s(Z_i)}\right\}W_i^{\mathrm{T}} \to E(\widetilde{W}^{\mathrm{T}})$ in probability. We have that

$$\begin{aligned}n^{1/2}(\widetilde{\theta}_n - \theta) &= n^{-1/2}\sum_{i=1}^{n}\left[\frac{\delta_i}{s(Z_i)}\varepsilon_i + \left\{1 - \frac{\delta_i}{s(Z_i)}\right\}U_i^{\mathrm{T}}\beta + E(\widetilde{W}^{\mathrm{T}})\Sigma_{X|Z}^{-1}\delta_i\{\widetilde{W}_i(\varepsilon_i - U_i^{\mathrm{T}}\beta) + \Sigma_{uu}\beta\}\right] \\ &\quad + n^{-1/2}\sum_{i=1}^{n}\{X_i^{\mathrm{T}}\beta + \nu(Z_i) - \theta\} + o_p(1).\end{aligned}$$

Noting that the two summation terms are uncorrelated, the proof follows.

A.4 Proof of Theorem 4

We first present a lemma, whose proof can be found in Liang, Härdle, and Carroll (1999).

Lemma A.1. Assume that random variables a_i and b_i satisfy $Ea_i = 0$ and $\|b_i\| = o_p(n^{-1/4})$. Then

$$\sum_{i=1}^{n} a_i b_i \xi_i = o_p(n^{1/2}),$$

where ξ_i are independent variables with zero conditional mean and finite variance.

Denote $\mathbf{\Omega}_i = \left(\{W_i - \widehat{m}_w(Z_i)\}[Y_i - \widehat{m}_y(Z_i) - \{W_i - \widehat{m}_w(Z_i)\}^{\mathrm{T}}\beta] + \Sigma_{uu}\beta\right)\delta_i$. A standard simplification as in Owen (2000, page 61) yields that

$$p_i = \frac{1}{n(1 + \mathbf{a}^{\mathrm{T}}\mathbf{\Omega}_i)} \text{ for } i = 1, \ldots, n, \tag{A.4}$$

where $\mathbf{a} = (a_1, \ldots, a_p)^{\mathrm{T}}$ is the solution of the equation

$$n^{-1} \sum_{i=1}^{n} \frac{\boldsymbol{\Omega}_i}{1 + \mathbf{a}^{\mathrm{T}} \boldsymbol{\Omega}_i} = 0. \tag{A.5}$$

Mimicking the proof Theorem 3.2 of Owen (2000), we have

$$\|\mathbf{a}\| = O_p(n^{-1/2}). \tag{A.6}$$

On the other hand, based on the assumptions, the result of Theorem 1 and the strong law of large number, we have

$$\max_{1 \le i \le n} \|\boldsymbol{\Omega}_i\| = o_p(n^{1/2}). \tag{A.7}$$

Note that

$$n^{-1} \sum_{i=1}^{n} \frac{\boldsymbol{\Omega}_i}{1 + \mathbf{a}^{\mathrm{T}} \boldsymbol{\Omega}_i} = n^{-1} \sum_{i=1}^{n} \boldsymbol{\Omega}_i (1 - \mathbf{a}^{\mathrm{T}} \boldsymbol{\Omega}_i) + n^{-1} \sum_{i=1}^{n} \frac{(\mathbf{a}^{\mathrm{T}} \boldsymbol{\Omega}_i)^2 \boldsymbol{\Omega}_i}{1 + \mathbf{a}^{\mathrm{T}} \boldsymbol{\Omega}_i}.$$

The second term is $o_p(n^{-1/2})$ since $|\mathbf{a}^{\mathrm{T}} \boldsymbol{\Omega}_i| = o_p(1)$ and $\sum_{i=1}^{n} (\mathbf{a}^{\mathrm{T}} \boldsymbol{\Omega}_i)^2 \boldsymbol{\Omega}_i \le \|\mathbf{a}\| \max_{1 \le i \le n} |\mathbf{a}^{\mathrm{T}} \boldsymbol{\Omega}_i| \sum_{i=1}^{n} \|\boldsymbol{\Omega}_i\|^2 = O_p(n^{-1/2}) o_p(1) O_p(n) = o_p(1)$. It then follows from (A.5) that

$$\mathbf{a} = (\sum_{i=1}^{n} \boldsymbol{\Omega}_i \boldsymbol{\Omega}_i^{\mathrm{T}})^{-1} \sum_{i=1}^{n} \boldsymbol{\Omega}_i + o_p(n^{-1/2}). \tag{A.8}$$

A similar argument using $\sum_{i=1}^{n} p_i = 1$ yields that

$$0 = n^{-1} \sum_{i=1}^{n} \frac{\mathbf{a}^{\mathrm{T}} \boldsymbol{\Omega}_i}{1 + \mathbf{a}^{\mathrm{T}} \boldsymbol{\Omega}_i} = n^{-1} \sum_{i=1}^{n} \mathbf{a}^{\mathrm{T}} \boldsymbol{\Omega}_i - n^{-1} \sum_{i=1}^{n} (\mathbf{a}^{\mathrm{T}} \boldsymbol{\Omega}_i)^2 + o_p(n^{-1}).$$

Therefore, we have

$$\sum_{i=1}^{n} \mathbf{a}^{\mathrm{T}} \boldsymbol{\Omega}_i = \sum_{i=1}^{n} (\mathbf{a}^{\mathrm{T}} \boldsymbol{\Omega}_i)^2 + o_p(1). \tag{A.9}$$

Consider $\mathcal{R}_n(\beta)$. Using a Taylor expansion of $\log(1+x)$ on x, we have

$$\begin{aligned} -\log\{\mathcal{R}_n(\beta)\} &= \sum_{i=1}^{n} \log(1 + \mathbf{a}^{\mathrm{T}} \boldsymbol{\Omega}_i) \\ &= \sum_{i=1}^{n} \Big\{ \mathbf{a}^{\mathrm{T}} \boldsymbol{\Omega}_i - (1/2)(\mathbf{a}^{\mathrm{T}} \boldsymbol{\Omega}_i)^2 \Big\} + Q_n. \end{aligned}$$

The remainder term Q_n is bounded by $\|\mathbf{a}\|^2 \max_{1 \le i \le n} |\mathbf{a}^{\mathrm{T}} \boldsymbol{\Omega}_i| \sum_{i=1}^{n} \|\boldsymbol{\Omega}_i\|^2 = O_p(n^{-1}) o_p(1) O_p(n) = o_p(1)$. Using (A.9) and (A.8), we have

$$\begin{aligned} -2\log\{\mathcal{R}_n(\beta)\} &= \sum_{i=1}^{n} \mathbf{a}^{\mathrm{T}} \boldsymbol{\Omega}_i \boldsymbol{\Omega}_i^{\mathrm{T}} \mathbf{a} + o_p(1) \\ &= (n^{-1/2} \sum_{i=1}^{n} \boldsymbol{\Omega}_i)^{\mathrm{T}} (n^{-1} \sum_{i=1}^{n} \boldsymbol{\Omega}_i \boldsymbol{\Omega}_i^{\mathrm{T}})^{-1} (n^{-1/2} \sum_{i=1}^{n} \boldsymbol{\Omega}_i) + o_p(1). \end{aligned}$$

Write $\widetilde{\mathbf{\Omega}}_i = \Big(\{W_i - m_w(Z_i)\}[Y_i - m_y(Z_i) - \{W_i - m_w(Z_i)\}^{\mathrm{T}}\beta] + \Sigma_{uu}\beta\Big)\,\delta_i$. Then $\widetilde{\mathbf{\Omega}}_i - \mathbf{\Omega}_i$ can be expressed as

$$\begin{aligned}
&\widetilde{W}_i[\{\widehat{m}_y(Z_i) - m_y(Z_i)\} - \{\widehat{m}_w(Z_i) - m_w(Z_i)\}^{\mathrm{T}}\beta]\delta_i \\
&-\{\widehat{m}_w(Z_i) - m_w(Z_i)\}[\{\widehat{m}_y(Z_i) - m_y(Z_i)\} - \{\widehat{m}_w(Z_i) - m_w(Z_i)\}^{\mathrm{T}}\beta]\delta_i \\
&+\{\widehat{m}_w(Z_i) - m_w(Z_i)\}(\widetilde{Y}_i - \widetilde{W}_i^{\mathrm{T}}\beta)\delta_i.
\end{aligned}$$

It follows from (A.1) that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\widehat{m}_w(Z_i) - m_w(Z_i)\}[\{\widehat{m}_y(Z_i) - m_y(Z_i)\} + \{\widehat{m}_w(Z_i) - m_w(Z_i)\}^{\mathrm{T}}\beta]\delta_i = o_p(1).$$

On the other hand, Lemma A.1 implies that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{W}_i[\{\widehat{m}_y(Z_i) - m_y(Z_i)\} + \{\widehat{m}_w(Z_i) - m_w(Z_i)\}^{\mathrm{T}}\beta]\delta_i = o_p(1),$$

and

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\widehat{m}_w(Z_i) - m_w(Z_i)\}(\widetilde{Y}_i - \widetilde{W}_i^{\mathrm{T}}\beta)\delta_i = o_p(1).$$

These results imply that $n^{-1/2}\sum_{i=1}^{n}\mathbf{\Omega}_i$ and $n^{-1/2}\sum_{i=1}^{n}\widetilde{\mathbf{\Omega}}_i$ asymptotically have the same limiting normal distribution, and $n^{-1}\sum_{i=1}^{n}\mathbf{\Omega}_i\mathbf{\Omega}_i^{\mathrm{T}}$ and $n^{-1}\sum_{i=1}^{n}\widetilde{\mathbf{\Omega}}_i\widetilde{\mathbf{\Omega}}_i^{\mathrm{T}}$ have the same limiting value. The proof is thus complete.

REFERENCES

BICKEL, P. J., KLAASSEN, C. J., RITOV, Y. & WELLNER, J. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Baltimore: Johns Hopkins University Press.

BICKEL, P. J. & RITOV, Y. (1987). Efficient estimation in the errors in variables model. *Annals of Statistics*, **15**, 513-540.

CARROLL, R. J., RUPPERT, D. & STEFANSKI, L. A. (1995). *Measurement Error in Nonlinear Models.* New York: Chapman and Hall.

CHEN, S. X. (1993). On the accuracy of empirical likelihood confidence regions for linear regression model. *Annals of the Institute of Statistical Mathematics*, **45**, 621-637.

CHEN, S. X. (1994). Empirical likelihood confidence intervals for linear regression coefficients. *Journal of Multivariate Analysis*, **49**, 24-40.

CHENG, P. E. (1994). Nonparametric estimation of mean functionals with data missing at random. *Journal of the American Statistical Association*, **89**, 81-87.

ENGLE, R. F., GRANGER, C. W. J., RICE, J. & WEISS, A. (1986). Semiparametric estimates of the relation between weather and electricity sales. *Journal of the American Statistical Association*, **81**, 310-320.

HÄRDLE, W., LIANG, H. & GAO, J. (2000). *Partially Linear Models*. Springer Physica-Verlag, Heidelberg.

LIANG, H., HÄRDLE, W. & CARROLL, R. J. (1999). Estimation in a semiparametric partially linear errors-in-variables model. *Annals of Statistics*, **27**, 1519-1535.

LIANG, H., WANG, S., ROBINS, J. M. & CARROLL, R. J. (2004). Estimation in partially linear models with missing covariates. *Journal of the American Statistical Association*, **99**, 357-367.

LIN, X. & CARROLL, R. J. (2000). Nonparametric function estimation for clustered data when the predictor is measured without/with error. *Journal of the American Statistical Association*, **95**, 520-534.

NACHMAN, S. A., STANLEY, K., YOGEV, R., ET AL. (2000). Nucleoside analogs plus ritonavir in stable antiretroviral therapy experienced HIV infected children: a randomized controlled trial. *Journal of the American Medical Association*, **283**, 492-498.

NEWEY, W. K. (1994). The asymptotic variance of semiparametric estimators. *Econometrica*, **62**, 1349-1382.

OPSOMER, J. D. & RUPPERT, D. (1999). A root-n consistent backfitting estimator for semiparametric additive modelling. *Journal of Computational and Graphical Statistics*, **8**, 715-732.

OWEN, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, **75**, 237-249.

OWEN, A. B. (1990). Empirical likelihood confidence regions. *Annals of Statistics*, **18**, 90-120.

OWEN, A. B. (1991). Empirical likelihood for linear models. *Annals of Statistics*, **19**, 1725-1747.

OWEN, A. B. (2000). *Empirical Likelihood*. London: Chapman and Hall/CRC.

QIN, J. (1994). Semi-empirical likelihood ratio confidence intervals for the difference of two sample means. *Annals of Statistics*, **46**, 117-126.

QIN, J. (1999). Empirical likelihood ratio based confidence intervals for mixture proportions. *Annals of Statistics*, **27**, 1368-1384.

QIN, J. & LAWLESS, J. (1994). Empirical likelihood and general estimating equations. *Annals of Statistics*, **22**, 300-325.

ROBINSON, P. M. (1988). Root-n-consistent semiparametric regression. *Econometrica*, **56**, 931-954.

RUPPERT, D., SHEATHER, S. J. & WAND, M. P. (1995). An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association*, **90**, 1257-1270.

SEVERINI, T. A. & STANISWALIS, J. G. (1994). Quasilikelihood estimation in semiparametric models. *Journal of the American Statistical Association*, **89**, 501-511.

SPECKMAN, P. (1988). Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society, Series B*, **50**, 413-436.

STEFANSKI, L. A. & CARROLL, R. J. (1987). Conditional scores and optimal scores in generalized linear measurement error models. *Biometrika*, **74**, 703-716.

WANG, Q. H., LINTON, O. & HÄRDLE, W. (2004). Semiparametric regression analysis with missing response at random. *Journal of the American Statistical Association*, **99**, 334-345.

ZEGER, S. L. & DIGGLE, P. J. (1994). Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics*, **50**, 689-699.

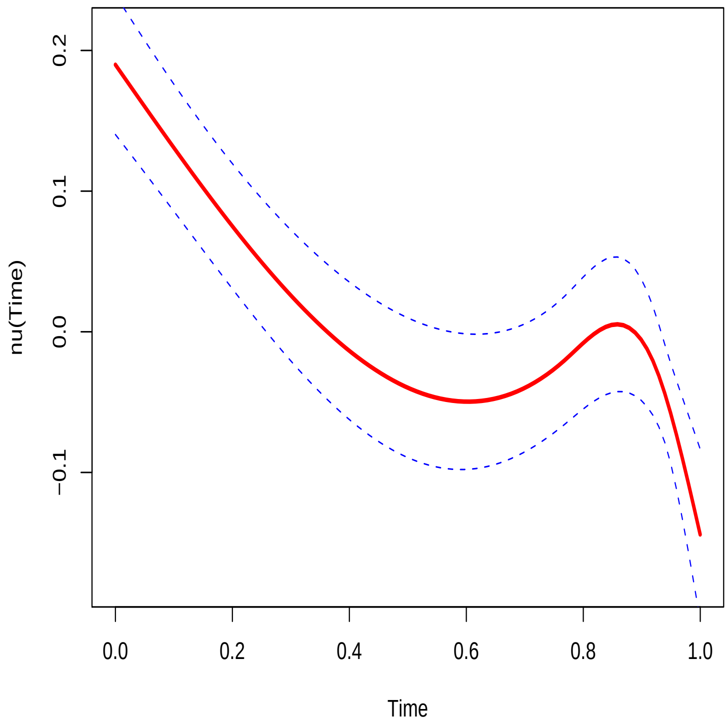

Figure 1: Estimates and the associated confidence interval of the nonparametric function $\nu(z)$ for PACTG 338 dataset. The solid curve represents the estimated values of $\nu(z)$ based on the complete observations, and the dotted lines are the confidence intervals.

Table 1: Point estimates for β_1 (=1), the 95% confidence intervals based on the empirical likelihood (EL) and normal approximation (Norm) methods, and the associated coverage probabilities for the simulated data. The four cases are discussed in the text.

		Estimate		CI (ME)		Coverage (ME)	
n	case	naive	ME	EL	Norm	EL	Norm
100	1	0.649	1.026	(0.479, 1.614)	(0.382, 1.670)	94.2	93.9
	2	0.649	1.024	(0.480, 1.609)	(0.446, 1.602)	95.1	93.1
	3	0.650	0.982	(0.784, 1.213)	(0.719, 1.245)	96.6	97.7
	4	0.650	1.029	(0.485, 1.612)	(0.451, 1.607)	94.1	92.4
500	1	0.662	1.001	(0.782, 1.235)	(0.778, 1.224)	95.2	95.7
	2	0.662	1.004	(0.779, 1.238)	(0.764, 1.244)	94.2	95.5
	3	0.662	1.001	(0.923, 1.102)	(0.913, 1.089)	96.7	96.4
	4	0.661	1.001	(0.829, 1.188)	(0.778, 1.224)	95.7	95.7

Table 2: Estimates of the parameter β, the 95% confidence intervals based on the empirical likelihood (EL) and normal approximation (Norm) methods for PACTG 338 dataset.

	$\sigma_u^2 = 0$	$\sigma_u^2 = 0.068$	$\sigma_u^2 = 0.135$
Estimates	-0.125	-0.164	-0.186
CI(Norm)	(-0.145, -0.105)	(-0.204,-0.124)	(-0.244, -0.128)
CI(EL)	(-0.132, -0.109)	(-0.183, -0.128)	(-0.221, -0.130)