Published in final edited form as:

Aust N Z J Stat. 2008 December ; 50(4): 347–359. doi:10.1111/j.1467-842X.2008.00521.x.

Empirical Likelihood Based Inferences for Partially Linear Models with Missing Covariates

Hua Liang and

Department of Biostatistics and Computational Biology, University of Rochester, Rochester, NY 14642, USA hliang@bst.rochester.edu

Yongsong Qin

School of Mathematical Sciences, Guangxi Normal University, Guilin, Guangxi 541004, P.R.China ysqin@mailbox.gxnu.edu.cn

Abstract

This paper considers statistical inference for partially linear models $Y = X^T\mu + v(Z) + \varepsilon$ when the linear covariate X is missing with missing probability π depending upon (Y, Z) . We propose empirical likelihood based statistics to construct confidence regions for β and $v(z)$. The resulting statistics are shown to be asymptotically chi-squared distributed. Finite sample performance of the proposed statistics is assessed by simulation experiments. The proposed methods are applied to a data set from an AIDS clinical trial.

Keywords

Confidence region; local linear regression; missing at random; semiparametric estimation

1 Introduction

The standard additive model, a generalization of the multiple linear regression models by introducing one-dimensional nonparametric functions in the place of the linear component, has been used to explore the complicated relation between the response to treatment and the predictors of interest (Stone 1985). Efforts have been also made to balance the interpretation of linear models and flexibility of additive models. Important results of those efforts are partially linear models (PLMs), which have been considered for analysis of cross-sectional and longitudinal data, and there is a substantial literature for PLMs and their generalizations; see for example Engle *et al.* (1986), Robinson (1988), Speckman (1988), Severini & Staniswalis (1994), Zeger & Diggle (1994), Gao & Anh (1999), Opsomer & Ruppert (1999), Härdle, Liang & Gao (2000), and Fan & Li (2004) among many others. Recently, Liang *et al.* (2004) considered the case in which the linear component is missing at random with missingness probability being independent of linear component. The authors developed new methods for estimating parameter and nonparametric function, and derived asymptotic distributions for the proposed estimators. Using their estimators of the covariance matrix or their bootstrap versions, one can give confidence regions of the parameters. However, the finite-sample performance is not optimistic because of the complexity of the covariance matrix and the need of plugging in several estimated terms (Liang *et al.* 2004 Theorem 2), and the confidence region derived by this procedure is based on a normal approximation. A valuable alternative method is the empirical likelihood principle, which has been systematically studied and developed by Owen (1988, 1990), Qin (1994, 1999), and Qin & Lawless (1994). The empirical likelihood ratio has a limiting chi-squared distribution, and it can be used to obtain tests and confidence intervals for a variety of settings, including linear models (Owen 1991, Chen 1993, 1994), generalized

linear models (Kolaczyk 1994), and general estimating equation (Qin & Lawless 1994). Empirical likelihood method has many advantages over its counterparts like the normal-approximation-based method and the bootstrap method (Hall & La Scala 1990). The most appealing features of the empirical likelihood method include improvement of the confidence region, increase of accuracy of coverage because of using auxiliary information, and ease of implementation.

Consider the following partially linear model:

$$Y = X^T \beta + v(Z) + \epsilon, \quad (1)$$

where $(X, Z, Y) \in R^p \times [0, 1] \times R$, β is an unknown parameter vector, $v(\cdot)$ is a smooth unknown function, $E(\epsilon|X, Z) = 0$. We study this model when there are missing data on X . Let $\delta = 1$ if X is observed, and $\delta = 0$ otherwise. Throughout this paper, we assume that X is missing at random (MAR) in the sense that

$$\pi(Y, Z) = \Pr(\delta=1|X, Y, Z) = \Pr(\delta=1|Y, Z).$$

Assume that $(Y_i, X_i, Z_i, \delta_i)$, $i = 1, 2, \dots, n$ are the independent and identically distributed data generated from (1). We are interested in constructing empirical likelihood confidence intervals of β and $v(z)$ for a fixed z . The empirical likelihood principle has been applied to PLMs. For example, Shi & Lau (1999) studied empirical likelihood method for PLMs. Qin & Jing (2001) and Wang & Li (2002) considered the case in which the response variables Y_i are randomly censored. The authors proposed an empirical likelihood ratio for β and derived an asymptotic distribution, which is a sum of independent chi-squared distributions with unknown weights. More recently, Liang, Wang & Carroll (2007) studied the PLMs with missing response variables.

This paper is organized as follows. In Section 2, the empirical likelihood ratio statistic for β is constructed. The limiting distribution of the statistic and an associated confidence interval for β are derived. Section 3 studies the same topic as in Section 2 for nonparametric component, and similar results are given. Section 4 reports the results from a simulation experiment, while Section 5 presents the results from a real data analysis. All technical derivations are given in the Appendix. Throughout this paper, we use the same notation as in Liang *et al.* (2004).

2 Confidence Interval of β

Let $m_1(Z) = E(X|Z)$, $m_2(Z) = E(Y|Z)$, $m_3(Y, Z) = E(X|Y, Z)$, and $m_4(Y, Z) = E(XX^T|Y, Z)$. For model (1), Liang *et al.* (2004) gave the estimating equation for estimating β after appropriately approximating $m_1(z)$, $m_2(z)$, $m_3(y, z)$, and $m_4(y, z)$. To deduce our methods, we estimate these four functions in the same way as Liang *et al.* (2004); i.e., using Horvitz-Thompson (HT) weighted local linear estimators (Liang *et al.* 2004, p.358). Denote the HT local linear estimators of $m_j(\cdot)$ and $E(y, z)$ as $\widehat{m}_j(\cdot)$ for $j = 1, 2, 3, 4$ and $\widehat{\pi}_j(y, z)$, respectively. Write

$$\begin{aligned}
\omega_1(m_1, m_2, m_3, m_4, \pi, \beta, Y, X, Z, \delta) = & (X - m_1(Z))(Y \\
& - m_2(Z) \\
& - (X - m_1(Z))^T \beta \frac{\delta}{\pi} \\
& - ((Y \\
& - m_2(Z)(m_3(Y, \\
& Z - m_1(Z) \\
& - (m_4(Y, \\
& Z - m_3(Y, \\
& Zm_1^T(Z) \\
& - m_1(Z)m_3^T(Y, \\
& Z + m_1(Z)m_1^T(Z))\beta \frac{\delta - \pi}{\pi}.
\end{aligned}$$

The estimating equation for β given by Liang *et al.* (2004) is of form:

$$\sum_{i=1}^n \omega_1(\widehat{m}_1, \widehat{m}_2, \widehat{m}_3, \widehat{m}_4, \widehat{\pi}, \beta, Y_i, X_i, Z_i, \delta_i) = 0.$$

This may motivate us to propose an empirical likelihood ratio statistic for β as follows.

$$\ell_1(\beta) = -2 \sup_{p_1, \dots, p_n} \sum_{i=1}^n \log(np_i),$$

subject to the constraints:

$$\sum_{i=1}^n p_i = 1 \quad \text{and} \quad \sum_{i=1}^n p_i \omega_1(\widehat{m}_1, \widehat{m}_2, \widehat{m}_3, \widehat{m}_4, \widehat{\pi}, \beta, Y_i, X_i, Z_i, \delta_i) = 0.$$

By the Lagrange multiplier method, it can be shown that

$$\ell_1(\beta) = 2 \sum_{i=1}^n \log(1 + \lambda_1 \omega_1(\widehat{m}_1, \widehat{m}_2, \widehat{m}_3, \widehat{m}_4, \widehat{\pi}, \beta, Y_i, X_i, Z_i, \delta_i)),$$

where λ_1 is determined by

$$\frac{1}{n} \sum_{i=1}^n \frac{\omega_1(\widehat{m}_1, \widehat{m}_2, \widehat{m}_3, \widehat{m}_4, \widehat{\pi}, \beta, Y_i, X_i, Z_i, \delta_i)}{1 + \lambda_1 \omega_1(\widehat{m}_1, \widehat{m}_2, \widehat{m}_3, \widehat{m}_4, \widehat{\pi}, \beta, Y_i, X_i, Z_i, \delta_i)} = 0. \quad (2)$$

We will examine statistical property of $\ell_1(\beta)$. Before stating the main results, we need regularity conditions. Assume that $\widehat{m}_{j\cdot}, j = 1, 2, 3, 4$, and $\widehat{\pi}_j$ converge uniformly at order $o_p(n^{-1/4})$. We claim the following conditions as in Liang *et al.* (2004) throughout the remainder of the article.

In what follows, we denote AA^T by $A^{\otimes 2}$, and let $\xi = \xi - E(\xi|Z)$ and $\hat{\xi} = \xi - \hat{E}(\xi|Z)$ for any random variable(vector) ξ , where $\hat{E}(\xi|Z)$ is a HT local linear estimator of $E(\xi|Z)$.

Assumption 2.1

- a. The matrix $E \begin{pmatrix} \tilde{X} & \tilde{X}^T \end{pmatrix}$ is positive-definite, $E(\epsilon|X, Z) = 0$, and $E(|\epsilon|^3|X, Z) < \infty$.
- b. The bandwidths for estimating $m_1(z)$ and $m_2(z)$ are of order $n^{-1/5}$, and the bandwidths for estimating $m_3(y, z)$ and $m_4(y, z)$ are of order $n^{-1/6}$.
- c. $K(\cdot)$ is a bounded symmetric density function with compact support and satisfies that $\int K(u) du = 0$, and $\int K(u) du = 1$.
- d. The density function of Z and the density function of (Y, Z) are bounded away from 0 and have bounded continuous second derivatives.
- e. The functions $m_1(z)$, $m_2(z)$ and $v(\cdot)$ have bounded and continuous second derivatives, and the functions $m_3(y, z)$ and $m_4(y, z)$ have bounded first derivatives.
- f. The probability function $\pi(y, z) > 0$ on the support of (Y, Z) , and has a bounded continuous second derivative.

The asymptotic property of the empirical likelihood ratio statistic $\ell_1(\beta)$ is established in Theorem 2.1, whose proof is given in the Appendix.

Theorem 2.1

Suppose that Assumption 2.1 is satisfied. Then, as $n \rightarrow \infty$,

$$\ell_1(\beta) \xrightarrow{d} \chi_p^2,$$

where χ_p^2 is the chi-squared distributed random variable with p degrees of freedom.

As a consequence, the approximate $100(1 - \alpha)\%$ confidence region of β can be formed as $\{\beta : \ell_1(\beta) \leq c_{\alpha,p}\}$, where $c_{\alpha,p}$ satisfies $\Pr(\chi_p^2 \leq c_{\alpha,p}) = 1 - \alpha$. To implement our method, we only need to estimate m_i and π , and to calculate $\ell_1(\beta)$ under the constraint (2). The algorithm developed by Owen (2001, p235) can be used for this goal.

3 Pointwise Confidence Interval of $v(z)$

Chen & Qin (2000) constructed confidence intervals for a nonparametric regression function with bounded support, and showed that the coverage error of the resulting confidence intervals has the same order throughout the support of the function. This feature remarkably improves the normal approximation based confidence intervals, which has a larger order of coverage error around the boundary. In this section, we are interested in empirical likelihood pointwise confidence interval of $v(z)$. To illustrate the motivation of empirical likelihood based confidence interval of $v(z)$ for a given $z \in (0, 1)$, we briefly review the results for nonparametric regression in the case of no missing data.

If $U_i = v(Z_i) + \epsilon_i$ are available, Chen & Qin (2000) proposed the following empirical likelihood statistic for $v(z)$,

$$\ell_0(\gamma) = -2 \sum_{i=1}^n \log(np_i)$$

subject to the following constraints:

$$\sum_{i=1}^n p_i = 1, \sum_{i=1}^n W_{ni}(z) (U_i - \gamma) p_i = 0,$$

Where $\gamma = v(z)$, $W_{ni}(z) = K_h(z - Z_i) s_{n,l} / h$ with $K_h(\cdot) = K(\cdot/h)$ and

$$s_{n,l} = \frac{1}{nh} \sum_{i=1}^n \frac{K_h(z - Z_i) (z - Z_i)^l}{h^l} \text{ for } l=1, 2.$$

If β in (1) were known and there is no missing data, one can obtain an empirical likelihood based confidence interval of $v(z)$ on the basis of the model $Y_i - X_i^T \beta = v(Z_i) + \varepsilon_i$. Let $\widehat{\beta}$ be a $\sqrt{n}$ -consistent estimator of β in the case of missing data discussed in this article (e.g. $\widehat{\beta}_{\widehat{\pi}, \text{all}}$ in Liang *et al.* 2004). Intuitively a confidence interval similar to Chen & Qin (2000) can be derived in a straightforward manner. To incorporate the missing information, we propose the following empirical likelihood statistic based on the local linear smoother:

$$\ell_2(\gamma) = -2 \sup_{p_1, \dots, p_n} \sum_{i=1}^n \log(np_i),$$

subject to the following constraints

$$\sum_{i=1}^n p_i = 1, \sum_{i=1}^n W_{ni}(z) \left(\frac{\delta_i}{\widehat{\pi}_i} (Y_i - X_i^T \widehat{\beta} - \gamma) + \left(1 - \frac{\delta_i}{\widehat{\pi}_i} \right) \widehat{\phi}_1(Y_i, Z_i, \widehat{\beta}) \right) p_i = 0,$$

where

$$\widehat{\phi}_1(Y, Z, \widehat{\beta}) = Y - (\widehat{E}(X|Y, Z))^T \widehat{\beta} - \gamma = Y - \widehat{m}_3^T \widehat{\beta} - \gamma.$$

Write

$$\omega_2(\widehat{m}_3, \widehat{\beta}, \widehat{\pi}, Y_i, X_i, Z_i, \delta_i) = W_{ni}(z) \left(\frac{\delta_i}{\widehat{\pi}_i} (Y_i - X_i^T \widehat{\beta} - \gamma) + \left(1 - \frac{\delta_i}{\widehat{\pi}_i} \right) \widehat{\phi}_1(Y_i, Z_i, \widehat{\beta}) \right). \quad (1)$$

By the Lagrange multiplier method, it can be shown that

$$\ell_2(\gamma) = 2 \sum_{i=1}^n \log \left(1 + \lambda_2 \omega_2(\widehat{m}_3, \widehat{\beta}, \widehat{\pi}, Y_i, X_i, Z_i, \delta_i) \right),$$

where λ_2 is determined by

$$\frac{1}{nh} \sum_{i=1}^n \frac{\omega_2(\widehat{m}_3, \widehat{\beta}, \widehat{\pi}, Y_i, X_i, Z_i, \delta_i)}{1 + \lambda_2 \omega_2(\widehat{m}_3, \widehat{\beta}, \widehat{\pi}, Y_i, X_i, Z_i, \delta_i)} = 0. \quad (3)$$

The asymptotic property of the empirical likelihood ratio statistic $\ell_2(\gamma)$ is established in Theorem 3.1. The proof of Theorem 3.1 is given in the Appendix.

Theorem 3.1

Suppose that $nh^5 \rightarrow 0$ and that Assumption 2.1 is satisfied. Then, as $n \rightarrow \infty$,

$$\ell_2(\gamma) \xrightarrow{d} \chi_1^2.$$

From Theorem 3.1, a $100(1 - \alpha)\%$ confidence interval of $v(z)$ for a given z can be constructed as $\{\gamma : \ell_2(\gamma) \leq c_{\alpha,1}\}$, where $c_{\alpha,1}$ satisfies $\Pr(\chi_1^2 \leq c_{\alpha,1}) = 1 - \alpha$.

4 Simulation Experiment

To evaluate the finite-sample performance of the proposed approach, we conducted a moderate sample simulation experiment. We generated $n = 100, 200$ observations from model (1), and assumed that $Y|X, Z \sim N\{X\beta + v(Z), \sigma^2\}$ and the probability of X to be observed equals $\Pr(\delta = 1|Y, X, Z) = \Phi\{\alpha_1 Y + v_1(Z)\}$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function. We set $\beta = 0.5$, $\sigma = 1$, $\alpha_1 = 2$, $Z \sim U(0, 1)$, and $v_1(z) = z^2$. We considered the following three cases.

Case 1 : $X|Z \sim U[0, 1]$, $v(Z) = 2Z$;

Case 2 : $X|Z \sim U[0, Z]$, $v(Z) = 2Z$;

Case 3 : $X|Z \sim U[0, Z]$, $v(Z) = \sin(Z)$.

The estimating procedure is the same as that in Liang *et al.* (2004, p.358); using the quartic kernel $K_0(u) = 15/16(1 - u^2)^2 I(|u| \leq 1)$ and the bandwidth $n^{-1/5}$ for estimating $m_1(z)$ and $m_2(z)$, and $K(u, v) = K_0(u)K_0(v)$ and bandwidths $n^{-1/6}$ for estimating $m_3(y, z)$ and $m_4(y, z)$. We generated 1000 datasets in each of configurations.

Table 1 gives the results for β , and Table 2 gives the results for the nonparametric function $v(z)$ at the three points $z = 0.3, 0.5, 0.7$, respectively. The column “CI” gives the confidence intervals using the empirical likelihood method. The lower and upper values are the averages of 1000 simulated lower and upper values, respectively. The column “len” gives the average length of the confidence intervals, while the column “CP(%)” gives the corresponding coverage probabilities of the 1000 simulated datasets. It can be seen from the tables that the coverage probabilities are close to the nominal level, and the larger the sample size, the closer to the nominal level. The lengths of the estimated confidence intervals of β decrease with the increase of sample size. The confidence intervals of β for case 3 are substantially wider than those for cases 1 and 2, while this feature disappears for the confidence intervals of the nonparametric components.

5 Real Data Analysis

In this section we present an illustrative analysis of the AIDS clinical trial group (ACTG 315) study. Our aim focuses on effectiveness of antiretroviral (anti-HIV) treatments, and 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. We model the relation between viral load and CD4 cell counts by using model (1). See Liang *et al.* (2004) for a detailed discussion. Let Y_{ij} be the viral load, and X_{ij} be the CD4 cell counts for subject i at treatment time Z_{ij} .

ACTG 315 was a single-arm clinical trial in which 53 enrolled subjects with moderately advanced HIV-1 infection received an antiretroviral therapy consisting of zidovudine, lamivudine, and ritonavir for 48 weeks. The primary objective of the study was to assess whether the treatment was associated with evidence of immunologic restoration. 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. HIV-1 RNA measurements were observed on days 0, 2, 7, 10, and weeks 2, 3, 4, 8, 12, 24, and 48. A total of 514 observations with 13.8% of CD4 cell count missing were obtained. Most of the missing CD4 cell counts occurred because some measurements of CD4 cell counts and viral load were obtained at different time points. Thus the missingness is at random.

In our data analysis, we ignored the correlation structure when computing the estimators and used the so-called working independence assumption because Lin & Carroll (2001) pointed that working independence has some model-robustness advantages over estimation methods that account for correlation, with a corresponding loss of efficiency. To reduce the marked skewness of CD4 cell counts, and the sparsity of treatment times, we took log-transformation for both variables. We used the same kernel function as in the simulation study in Section 4, and obtained a bandwidth of $h = 0.15$ in the same manner as described there. The empirical likelihood based confidence interval of β is $[-0.199, -0.025]$, which is slightly wider than that obtained by using normal approximation method given in Liang *et al.* (2004, p. 365). The difference between these two confidence intervals is not substantial but visible. We prefer the results based on the empirical likelihood method.

The confidence bands of the nonparametric function of treatment time are shown in Figure 1. Overall, the confidence intervals of the nonparametric function indicate that the viral load RNA levels decrease after initial antiviral treatment and then become flat. This finding is similar to that reflected in Liang *et al.*'s (2004) analysis for this dataset.

6 Discussion

To make inference for partially linear models with missing covariates, we proposed an empirical likelihood-based approach to construct confidence regions of β and $v(z)$. The proposed approach is remarkably simpler to implement than its counterpart, normal approximation method, and easily executable. The finite-sample performance of the proposed statistics shows promising. It may be of interest to extend the methods to generalized partially linear models:

$$\mu(E(Y|X, Z)) = X^T \beta + v(Z),$$

where $\mu(\cdot)$ is a known link function. This needs a further investigation, but it is beyond the scope of this paper.

Acknowledgments

The authors thank the Editor, an associate editor, and two referees for their constructive comments and suggestions. This research was supported by NIH/NIAID grants AI62247 and AI059773.

Appendix

The proofs of Theorems 2.1 and 3.1 follow the arguments similar to those used by Liang *et al.* (2004) to prove their Theorems 1 and 2, and by Owen (2001) to prove his Theorem 3.2. We give only the key steps and departures from their procedures.

Proof of Theorem 2.1.

Let

$$T_n = \max_{1 \leq i \leq n} \|\omega_1(\widehat{m}_1, \widehat{m}_2, \widehat{m}_3, \widehat{m}_4, \widehat{\pi}, \beta, Y_i, X_i, Z_i, \delta_i)\|.$$

To prove Theorem 2.1, we first show that

$$n^{-1/2} \sum_{i=1}^n \omega_1(\widehat{m}_1, \widehat{m}_2, \widehat{m}_3, \widehat{m}_4, \widehat{\pi}, \beta, Y_i, X_i, Z_i, \delta_i) \xrightarrow{d} N(0, C), \quad (\text{A.1})$$

$$n^{-1} \sum_{i=1}^n (\omega_1(\widehat{m}_1, \widehat{m}_2, \widehat{m}_3, \widehat{m}_4, \widehat{\pi}, \beta, Y_i, X_i, Z_i, \delta_i))^{\otimes 2} = C + o_p(1), \quad (\text{A.2})$$

and

$$T_n = o(n^{1/2}), a.s., \quad (\text{A.3})$$

$$\text{where } C = E\left(\epsilon^2 \tilde{X} \tilde{X}^T / \pi\right) - E\left((1 - \pi) \left(E(\tilde{X} \in |Y, Z)\right)\right)^{\otimes 2} / \pi$$

Using the arguments similar to the proofs of Theorems 1 and 2 in Liang *et al.* (2004) and Lemma 5.1 (with his $\alpha(z) = 0$) in Newey (1994), we know that

$$n^{-1/2} \sum_{i=1}^n \omega_1(\widehat{m}_1, \widehat{m}_2, \widehat{m}_3, \widehat{m}_4, \widehat{\pi}, \beta, Y_i, X_i, Z_i, \delta_i) \text{ and } n^{-1/2} \sum_{i=1}^n \omega_1(m_1, m_2, m_3, m_4, \pi, \beta, Y_i, X_i, Z_i, \delta_i)$$

have the same limiting distribution. The latter one is a sum of iid random vectors. In a consequence, (A.1) follows. On the other hand,

$$n^{-1} \sum_{i=1}^n (\omega_1(\widehat{m}_1, \widehat{m}_2, \widehat{m}_3, \widehat{m}_4, \widehat{\pi}, \beta, Y_i, X_i, Z_i, \delta_i))^{\otimes 2} = n^{-1} \sum_{i=1}^n (\omega_1(m_1, m_2, m_3, m_4, \pi, \beta, Y_i, X_i, Z_i, \delta_i))^{\otimes 2} + o_p(1).$$

It is readily seen that (A.2) and (A.3) hold.

From (A.1), (A.2), and (A.3), using the same arguments as the proof of Theorem 3.2 in Owen (1990), we have

$$\lambda_1 = O_p(n^{-1/2}).$$

It follows from (2) and $\sum_{i=1}^n p_i = 1$ that

$$\lambda_1 = \left(\frac{1}{n} \sum_{i=1}^n (\omega_1(\widehat{m}_1, \widehat{m}_2, \widehat{m}_3, \widehat{m}_4, \widehat{\pi}, \beta, Y_i, X_i, Z_i, \delta_i))^{\otimes 2} \right)^{-1} \times \left(\frac{1}{n} \sum_{i=1}^n \omega_1(\widehat{m}_1, \widehat{m}_2, \widehat{m}_3, \widehat{m}_4, \widehat{\pi}, \beta, Y_i, X_i, Z_i, \delta_i) \right) + o_p(n^{-1/2}).$$

By Taylor expansion we obtain that

$$\ell_1(\beta) = n \left(\frac{1}{n} \sum_{i=1}^n \omega_1(\widehat{m}_1, \widehat{m}_2, \widehat{m}_3, \widehat{m}_4, \widehat{\pi}, \beta, Y_i, X_i, Z_i, \delta_i) \right)^T \times \left(\frac{1}{n} \sum_{i=1}^n (\omega_1(\widehat{m}_1, \widehat{m}_2, \widehat{m}_3, \widehat{m}_4, \widehat{\pi}, \beta, Y_i, X_i, Z_i, \delta_i))^{\otimes 2} \right)^{-1} \times \left(\frac{1}{n} \sum_{i=1}^n \omega_1(\widehat{m}_1, \widehat{m}_2, \widehat{m}_3, \widehat{m}_4, \widehat{\pi}, \beta, Y_i, X_i, Z_i, \delta_i) \right) \quad (\text{A.4})$$

A combination of (A.1), (A.2), and (A.4) establishes Theorem 2.1. Let

$$\zeta(m_3, \beta, \pi, Y_i, X_i, Z_i, \delta_i) = h^{-1/2} W_{ni}(z) \left(\frac{\delta_i}{\pi_i} (Y_i - X_i^T \beta - \gamma) + \left(1 - \frac{\delta_i}{\pi_i} \right) \phi_1(Y_i, Z_i, \beta) \right),$$

and

$$\zeta_1(m_3, \beta, \pi, Y_i, X_i, Z_i, \delta_i) = h^{-1/2} W_{ni}(z) \left(\frac{\delta_i}{\pi_i} (m_3 - X_i)^T \beta + (Y_i - m_3^T \beta - \gamma(Z_i)) \right),$$

where $\phi_1(Y, Z, \beta) = Y - m^T \beta - \gamma$. We first prove the following lemma.

Lemma A.1

Under the conditions of Theorem 3.1,

$$n^{-1/2} \sum_{i=1}^n \zeta_1(\widehat{m}_3, \beta, \widehat{\pi}, Y_i, X_i, Z_i, \delta_i) = n^{-1/2} \sum_{i=1}^n \zeta_1(m_3, \beta, \pi, Y_i, X_i, Z_i, \delta_i) + o_p(1).$$

Proof

The proof can be finished using the arguments similar to the proofs of Theorems 1 and 2 in Liang *et al.* (2004) by utilizing Lemma 5.1 of Newey (1994) when $W_{ni}(z)$ is equal to $K_h(z - Z_i)$ or $K_h(z - Z_i)(z - Z_i)/h$.

Proof of Theorem 3.1

We first derive the following two expressions:

$$n^{-1/2} \sum_{i=1}^n \zeta(\widehat{m}_3, \widehat{\beta}, \widehat{\pi}, Y_i, X_i, Z_i, \delta_i) = (nh)^{-1/2} \sum_{i=1}^n W_{ni}(z) \left(\frac{\delta_i}{\pi_i} (m_3 - X_i)^T \beta + (Y_i - m_3^T \beta - \gamma) \right) + o_p(1), \quad (\text{A.5})$$

$$n^{-1} \sum_{i=1}^n \zeta^2(\widehat{m}_3, \widehat{\beta}, \widehat{\pi}, Y_i, X_i, Z_i, \delta_i) = (nh)^{-1} \sum_{i=1}^n W_{ni}^2(z) \left(\frac{\delta_i}{\pi_i} (m_3 - X_i)^T \beta + (Y_i - m_3^T \beta - \gamma) \right)^2 + o_p(1). \quad (\text{A.6})$$

Recall $\sqrt{n}(\widehat{\beta} - \beta) = O_p(1)$. A direct calculation yields that

$$\begin{aligned} n^{-1/2} \sum_{i=1}^n \zeta(\widehat{m}_3, \widehat{\beta}, \widehat{\pi}, Y_i, X_i, Z_i, \delta_i) &= n^{-1/2} \sum_{i=1}^n h^{-1/2} W_{ni}(z) \left(\frac{\delta_i}{\pi_i} (\widehat{m}_3 - X_i)^T \widehat{\beta} + (Y_i - \widehat{m}_3^T \widehat{\beta} - \gamma) \right) \\ &= n^{-1/2} \sum_{i=1}^n h^{-1/2} W_{ni}(z) \left(\frac{\delta_i}{\pi_i} (\widehat{m}_3 - X_i)^T \beta + (Y_i - \widehat{m}_3^T \beta - \gamma) \right) + \sqrt{h} \cdot (nh)^{-1} \sum_{i=1}^n W_{ni}(z) \left(\frac{\delta_i}{\pi_i} (\widehat{m}_3 - X_i) - \widehat{m}_3 \right)^T \cdot \sqrt{n}(\widehat{\beta} - \beta) \\ &= n^{-1/2} \sum_{i=1}^n h^{-1/2} W_{ni}(z) \left(\frac{\delta_i}{\pi_i} (\widehat{m}_3 - X_i)^T \beta + (Y_i - \widehat{m}_3^T \beta - \gamma) \right) + o_p(1). \end{aligned}$$

(A.5) follows from Lemma A.1. In a similar way, we can prove (A.6). Write

$$\bar{U}_j = (nh)^{-1} \sum_{i=1}^n \omega_2^j(\widehat{m}_3, \widehat{\beta}, \widehat{\pi}, Y_i, X_i, Z_i, \delta_i) \text{ for } j=1, 2.$$

Then

$$\sqrt{nh} \bar{U}_1 = n^{-1/2} \sum_{i=1}^n \zeta(\widehat{m}_3, \widehat{\beta}, \widehat{\pi}, Y_i, X_i, Z_i, \delta_i),$$

and

$$\bar{U}_2 = n^{-1} \sum_{i=1}^n \zeta^2(\widehat{m}_3, \widehat{\beta}, \widehat{\pi}, Y_i, X_i, Z_i, \delta_i). \quad (\text{A.7})$$

It follows from (A.5) and the central limiting theorem that

$$\bar{U}_1 = O_p((nh)^{-1/2} + h^2).$$

On the other hand, (3), (A.5), and (A.6) imply that

$$\lambda_2 = \left(\bar{U}_2 \right)^{-1} \bar{U}_1 + O_p \left(\left((nh)^{-1/2} + h^2 \right)^2 \right).$$

By Taylor expansion, we obtain

$$\begin{aligned} \ell_2(\gamma) &= 2\lambda_2 \sum_{i=1}^n \omega_2(\widehat{m}_3, \widehat{\beta}, \widehat{\pi}, Y_i, X_i, Z_i, \delta_i) - \lambda_2^2 \sum_{i=1}^n \omega_2^2(\widehat{m}_3, \widehat{\beta}, \widehat{\pi}, Y_i, X_i, Z_i, \delta_i) + O_p \left(\left((nh)^{-1/2} + h^2 \right)^3 \right) \\ &= (nh) \bar{U}_1 \left(\bar{U}_2 \right)^{-1} + O_p \left(\left((nh)^{-1/2} + h^2 \right)^3 \right). \end{aligned}$$

The proof is thus completed by combining (A.5)-(A.7).

References

- Chen SX. On the accuracy of empirical likelihood confidence regions for linear regression model. *Annals of the Institute of Statistical Mathematics* 1993;45:621–637.
- Chen SX. Empirical likelihood confidence intervals for linear regression coefficients. *Journal of Multivariate Analysis* 1994;49:24–40.
- Chen SX, Qin YS. Empirical likelihood confidence intervals for local linear smoothers. *Biometrika* 2000;87:946–953.
- Engle RF, Granger CWJ, Rice J, Weiss A. Semiparametric estimates of the relation between weather and electricity sales. *Journal of American Statistical Association* 1986;81:310–320.
- Fan J, Li R. New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. *Journal of American Statistical Association* 2004;99:710–723.
- Gao JT, Anh V. Semiparametric regression under long-range dependent errors. *Journal of Statistical Planning and Inference* 1999;80:37–57.
- Hall P, La Scala B. Methodology and algorithms of empirical likelihood. *Int. Statist. Rev* 1990;58:109–127.
- Härdle, W.; Liang, H.; Gao, J. *Partially Linear Models*. Springer Physica-Verlag; Heidelberg: 2000.
- Kolaczyk ED. Empirical likelihood for generalized linear models. *Statistica Sinica* 1994;4:199–218.
- Lederman MM, Connick E, Landay A, et al. Immunologic responses associated with 12 weeks of combination antiretroviral therapy consisting of Zidovudine, Lamivudine and Ritonavir: results of AIDS clinical trials group protocol 315. *The Journal of Infectious Diseases* 1998;178:70–79. [PubMed: 9652425]
- Liang H, Wang SJ, Carroll R. Partially linear models with missing response variables and error-prone covariates. *Biometrika* 2007;94:185–198. [PubMed: 20209027]
- Liang H, Wang S, Robins JM, Carroll RJ. Estimation in partially linear models with missing covariates. *Journal of American Statistical Association* 2004;99:357–367.
- Lin XH, Carroll RJ. Semiparametric regression for clustered data using generalized estimating equations. *Journal of American Statistical Association* 2001;96:1045–1056.
- Newey WK. The asymptotic variance of semiparametric estimators. *Econometrica* 1994;62:1349–1382.
- Opsomer JD, Ruppert D. A root- n consistent backfitting estimator for semiparametric additive modeling. *Journal of Computation and Graphical Statistics* 1999;8:715–732.
- Owen AB. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* 1988;75:237–249.
- Owen AB. Empirical likelihood ratio confidence regions. *Annals of Statistics* 1990;18:90–120.
- Owen AB. Empirical likelihood for linear models. *Annals of Statistics* 1991;19:1725–1747.
- Owen, AB. *Empirical likelihood*. Chapman & Hall; New York: 2001.
- Qin GS, Jing BY. Censored partial linear models and empirical likelihood. *Journal of Multivariate Analysis* 2001;78:37–61.

- Qin J. Semi-empirical likelihood ratio confidence intervals for the difference of two sample means. *Annals of Statistics* 1994;46:117–26.
- Qin J. Empirical likelihood ratio based confidence intervals for mixture proportions. *Annals of Statistics* 1999;27:1368–84.
- Qin J, Lawless J. Empirical likelihood and general estimating equations. *Annals of Statistics* 1994;22:300–325.
- Robinson PM. Root- n -consistent semiparametric regression. *Econometrica* 1988;56:931–954.
- Severini TA, Staniswalis JG. Quasilikelihood estimation in semiparametric models. *Journal of American Statistical Association* 1994;89:501–511.
- Shi J, Lau TS. Empirical likelihood for partially linear models. *Journal of Multivariate Analysis* 1999;72:132–148.
- Speckman P. Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society, Series B* 1988;50:413–436.
- Stone CJ. Additive regression and other nonparametric models. *Annals of Statistics* 1985;13:689–705.
- Wang QH, Li G. Empirical likelihood semiparametric regression analysis under random censorship. *Journal of Multivariate Analysis* 2002;83:469–486.
- Zeger SL, Diggle PJ. Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics* 1994;50:689–699. [PubMed: 7981395]

Figure 1. Empirical likelihood based confidence interval of the nonparametric function $v(z)$ for the data from the A315 study.

Table 1

The length (len) and coverage probability (CP) of the 95% confidence intervals (CI) for the parameter $\beta (= 0.5)$.

Case	n	CP(%)	CI	len
1	100	93.80	(0.3075, 0.6824)	0.3749
1	200	95.90	(0.3255, 0.6200)	0.2945
2	100	95.20	(0.2060, 0.7015)	0.4955
2	200	96.10	(0.3030, 0.6566)	0.3536
3	100	96.00	(0.0965, 0.8596)	0.7631
3	200	95.10	(0.1001, 0.8625)	0.7624

Table 2

The length (len) and coverage probability (CP) of the 95% confidence intervals (CI) for the nonparametric function $v(z)$ at $z = 0.3, 0.5$ and 0.7 .

Case	n	CP(%)	CI	len
$z = 0.3$				
1	100	96.00	(0.2415, 1.0216)	0.7801
1	200	94.50	(0.2848, 0.9655)	0.6807
2	100	99.00	(0.1921, 1.1322)	0.9401
2	200	96.00	(0.3010, 1.0185)	0.7175
3	100	99.00	(-0.7365, 0.3879)	1.1244
3	200	93.50	(-0.5284, 0.2515)	0.7799
$z = 0.5$				
1	100	96.50	(0.4411, 1.3555)	0.9144
1	200	94.50	(0.4150, 1.3543)	0.9393
2	100	97.00	(0.3082, 1.5142)	1.2060
2	200	96.10	(0.4511, 1.2058)	0.7547
3	100	98.20	(-0.5895, 0.5675)	1.1570
3	200	94.50	(-0.4502, 0.3911)	0.8413
$z = 0.7$				
1	100	96.00	(0.8490, 1.7400)	0.8910
1	200	95.50	(1.0238, 1.7948)	0.7710
2	100	97.50	(1.0722, 1.8027)	0.7305
2	200	93.50	(1.1621, 1.7611)	0.5990
3	100	99.00	(-0.4542, 0.3954)	0.8496
3	200	93.50	(-0.3312, 0.4044)	0.7356