Journal of the Korean Statistical Society 43 (2014) 91–103

Contents lists available at ScienceDirect

Journal of the Korean Statistical Society

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

Empirical likelihood for parameters in an additive partially linear errors-in-variables model with longitudinal data

Xiaoshuang Zhou [a,b], Peixin Zhao [c], Lu Lin [a,b,*]

[a] Shandong University Qilu Securities Institute for Financial Studies, Shandong University, Jinan 250014, China

[b] School of Mathematics, Shandong University, Jinan 250014, China

[c] Department of Mathematics, Hechi University, Yizhou 546300, China

ARTICLE INFO

Article history:
Received 16 March 2012
Accepted 10 April 2013
Available online 2 May 2013

AMS 2000 subject classifications:
primary 62J02
secondary 62E20

Keywords:
Empirical likelihood
Additive partially linear model
Errors-in-variables
Longitudinal data

ABSTRACT

Empirical likelihood inferences for the parameter component in an additive partially linear errors-in-variables model with longitudinal data are investigated in this article. A corrected-attenuation block empirical likelihood procedure is used to estimate the regression coefficients, a corrected-attenuation block empirical log-likelihood ratio statistic is suggested and its asymptotic distribution is obtained. Compared with the method based on normal approximations, our proposed method does not require any consistent estimator for the asymptotic variance and bias. Simulation studies indicate that our proposed method performs better than the method based on normal approximations in terms of relatively higher coverage probabilities and smaller confidence regions. Furthermore, an example of an air pollution and health data set is used to illustrate the performance of the proposed method.

1. Introduction

Longitudinal data arise frequently in biological and economic applications. In this article, we consider the following additive partially linear model with longitudinal data:

$$Y_{ij} = X_{ij}^T \beta + \sum_{d=1}^{D} f_d(Z_{ij}^d) + \varepsilon_{ij}, \quad i = 1, \dots, n,\ j = 1, \dots, n_i, \tag{1.1}$$

where Y_{ij} is the response variable for the jth measurement of the ith subject, X_{ij} and $Z_{ij} = (Z_{ij}^1, \dots, Z_{ij}^D)^T$ are covariates on R^p and R^D respectively, $f_1, f_2, \dots, f_D$ are unspecified smooth functions, $\beta = (\beta_1, \dots, \beta_p)^T$ is a p-dimensional vector of unknown parameters, and errors ε_{ij} have mean zero conditional on X_{ij} and Z_{ij}.

Semi-parametric additive partially linear models, which balance the interpretability of linear models and flexibility of additive models, have been studied by many authors recently. For example, Opsomer and Ruppert (1999) proposed a root-n consistent backfitting estimator for the parametric component of the model; Manzana and Zeromb (2005) introduced a marginal integration estimator for the parametric component; Jiang, Zhou, Jiang, and Peng (2007) extended the generalized likelihood ratio tests in Fan and Jiang (2005) to the model; Liang, Thurston, Ruppert, and Apanasovich (2008) investigated additive partially linear models with measurement errors. All these articles focus on inferences of the regression coefficients

* Corresponding author at: School of Mathematics, Shandong University, Jinan 250014, China.
E-mail address: linlu@sdu.edu.cn (L. Lin).

1226-3192/$ – see front matter
http://dx.doi.org/10.1016/j.jkss.2013.04.002

without longitudinal data. In this article we generalize the model of Liang et al. (2008) to the case with longitudinal data. Under the setting considered in this article, moreover, measurement errors in X_{ij} are allowed.

The aim of the present article is to investigate the issues of estimation and confidence region construction using the empirical likelihood approach (Owen, 1988, 1990, 1991). It is well known that the empirical likelihood method does not need to estimate the asymptotic covariance for constructing the confidence region. Furthermore, the confidence region's shape and orientation are determined entirely by the data itself. Many authors had used this method to deal with different kinds of problems. For example, Wang, Chen, and Lin (2010) considered the empirical likelihood inferences for the unknown parameter in an additive partially linear errors-in-variables model; Zhang and Zhu (2011) investigated empirical likelihood inferences for longitudinal data with missing response variables and error-prone covariates in a partially linear model. More recent works on other kinds of models with longitudinal data or errors-in-variables models can be found from Li, Tian, and Xue (2008); Li and Xue (2008); Li, Zhu, Xue, and Feng (2010); Wang, Li, and Lin (2011) among others. As stated above, in this article we consider the empirical likelihood inferences for the parameter in an additive partially linear errors-in-variables model with longitudinal data. Due to the within-subject correlation structure, the results of Liang et al. (2008) and Wang et al. (2010) cannot be applied directly to our model. Since subjects are independent, we take each subject as a whole and ignore the within-subject correlation structure using the so-called working independence assumption when we compute the estimates. As pointed out in Lin and Carroll (2001), the working independence has some model-robustness advantages over those estimation methods which have considered the within-subject correlation structure. Motivated by Zhang and Zhu (2011), we propose a corrected-attenuation block empirical likelihood method for β and prove that the estimated corrected-attenuation block empirical log-likelihood ratio statistics is asymptotically chi-squared.

The rest of this article is organized as follows. The corrected-attenuation block empirical likelihood method and the maximum empirical likelihood estimator for β are proposed in Section 2. Conditions and main results are given in Section 3. Simulation studies and an application of an air pollution and health data set are conducted in Section 4. The conclusion is conducted in Section 5. The proofs of the main results are relegated to the Appendix.

2. Empirical likelihood for the parameter component

For notational simplicity, we only consider the case of $D = 2$ in model (1.1). To ensure identifiability of the nonparametric functions, we assume that $E(f_1(Z^1)) = E(f_2(Z^2)) = 0$. We also assume that Y and X are centered, without loss of generality. Suppose that the observed data $\{Y_{ij}, Z_{ij}^1, Z_{ij}^2, W_{ij}, i = 1, \ldots, n, j = 1, \ldots, n_i\}$ are generated from the following model

$$\begin{cases} Y_{ij} = X_{ij}^T \beta + f_1(Z_{ij}^1) + f_2(Z_{ij}^2) + \varepsilon_{ij}, \\ W_{ij} = X_{ij} + U_{ij}, \end{cases} \tag{2.1}$$

where $E(\varepsilon_{ij}|X_{ij}, Z_{ij}^1, Z_{ij}^2) = 0$, $E(U_{ij}) = 0$ and $\mathrm{Cov}(U_{ij}) = \Sigma_{uu}$. If β is known, the first equation of model (2.1) can be rewritten as

$$Y_{ij} - X_{ij}^T \beta = f_1(Z_{ij}^1) + f_2(Z_{ij}^2) + \varepsilon_{ij}, \quad i = 1, 2, \ldots, n,\ j = 1, 2, \ldots, n_i. \tag{2.2}$$

Denote

$$\begin{aligned} &\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{in_i})^T, \qquad \mathbf{X}_i = (X_{i1}, \ldots, X_{in_i})^T, \qquad \varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{in_i})^T, \\ &\mathbf{Z}_i^1 = (Z_{i1}^1, \ldots, Z_{in_i}^1)^T, \qquad \mathbf{Z}_i^2 = (Z_{i1}^2, \ldots, Z_{in_i}^2)^T, \\ &F_i^1 = (f_1(Z_{i1}^1), \ldots, f_1(Z_{in_i}^1))^T, \qquad F_i^2 = (f_2(Z_{i1}^2), \ldots, f_2(Z_{in_i}^2))^T. \end{aligned}$$

Then, (2.2) can be written as

$$\mathbf{Y}_i - \mathbf{X}_i \beta = F_i^1 + F_i^2 + \varepsilon_i, \quad i = 1, 2, \ldots, n. \tag{2.3}$$

Similar to Liang et al. (2008) and Wang et al. (2010), let S_{i1,z^1}, S_{i2,z^2} denote the equivalent kernels for the local linear regression at z^1 and z^2, respectively. It is noted that z^1 and z^2 are only symbols here, and the following have the same meaning. Then

$$\begin{aligned} S_{i1,z^1} &= e_1^T [(\mathcal{Z}_i^1)^T \Omega_i^1 \mathcal{Z}_i^1]^{-1} (\mathcal{Z}_i^1)^T \Omega_i^1, \\ S_{i2,z^2} &= e_1^T [(\mathcal{Z}_i^2)^T \Omega_i^2 \mathcal{Z}_i^2]^{-1} (\mathcal{Z}_i^2)^T \Omega_i^2, \end{aligned}$$

where $e_1^T = (1, 0)$,

$$\Omega_i^1 = \mathrm{diag}\left(\frac{1}{h_1} K\left(\frac{Z_{i1}^1 - z^1}{h_1}\right), \ldots, \frac{1}{h_1} K\left(\frac{Z_{in_i}^1 - z^1}{h_1}\right)\right),$$

$$\Omega_i^2 = \mathrm{diag}\left(\frac{1}{h_2} K\left(\frac{Z_{i1}^2 - z^2}{h_2}\right), \ldots, \frac{1}{h_2} K\left(\frac{Z_{in_i}^2 - z^2}{h_2}\right)\right),$$