Journal of Multivariate Analysis 130 (2014) 275–288

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Empirical likelihood for partly linear models with errors in all variables

Li Yan, Xia Chen *

College of Mathematics and Information Science, Shaanxi Normal University, Xi'an 710062, China

ARTICLE INFO

Article history:
Received 11 June 2013
Available online 16 June 2014

AMS 2000 subject classifications:
62F12
62G05

Keywords:
Empirical likelihood
Measurement error
Confidence regions
Coverage probability
Maximum empirical likelihood estimate

ABSTRACT

In this paper, we consider the application of the empirical likelihood method to a partly linear model with measurement errors in possibly all the variables. It is shown that the empirical log-likelihood ratio at the true parameters converges to the standard chi-square distribution. Also, a class of estimators for the parameter are constructed, and the asymptotic distributions of the proposed estimators are obtained. Some simulations and an application are conducted to illustrate the proposed method.

1. Introduction

Let (X, T, Y) denote a triple of random variables (or vectors). Consider the following partly linear model

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

where X is a p-dimensional random vector, T is a random variable defined on $[0, 1]$, β is an unknown p-dimensional parameter vector. The function $g(\cdot)$ is an unknown smoothing function defined in $[0, 1]$, and ε is the random error with $\mathrm{E}(\varepsilon|X, T) = 0$.

However, because of the measuring mechanism or the nature of environment, the covariates are not always observable without error. Generally, for model (1), the following three cases are considered. Case one with measurement errors in the parametric part, Cui and Li [3], Liang et al. [10], He and Liang [6] studied the asymptotic normality of the estimators of the parameter and the convergence rate of the estimate of the nonparametric function in model (1). The empirical likelihood inference for parameter β in the semi-linear model can be found in Cui and Kong [2]. Case two with measurement errors in the nonparametric part, Fan and Truong [5] discussed the regression function estimate in the nonparametric regression model. Liang [9] studied the generalized least-squares estimator of the parameter β and obtained the asymptotic normality of the estimator. Chen and Cui [1] applied the empirical likelihood method to the partly linear model in this case. Case three with measurement errors in both the parametric part and the nonparametric part, Zhu and Cui [18] treated the strong convergence, optimal rate of weak convergence and the asymptotic normality of the estimators of the parameters.

In this paper, we discuss the empirical likelihood inference for partly linear models with measurement errors in both the parametric and the non-parametric part, simultaneously. The empirical likelihood as an alternative to the bootstrap for

* Corresponding author.
E-mail addresses: lyan@snnu.edu.cn (L. Yan), xchen80@snnu.edu.cn (X. Chen).

http://dx.doi.org/10.1016/j.jmva.2014.06.007
0047-259X/

constructing confidence regions was introduced by Owen [11,12]. The method defines an empirical likelihood ratio function to construct confidence regions. Important features of the empirical likelihood method are its automatic determination of the shape and orientation of the confidence region by the data. Qin and Lawless [14], Qin [13], and Shi and Lau [15] introduced this method into semi-parametric models and obtained the asymptotic efficiency of the maximum empirical likelihood estimate. More references can be found from Xue and Zhu's book [17] and references therein.

As noted earlier, the variables X and T are measured with error. That is, X and T are observed through

$$\begin{cases} Z = X + u, \\ W = T + v, \end{cases}$$

where u and v are random errors. Therefore, a partly linear models with errors in all possible variables is defined as follows:

$$\begin{cases} Y = X^{\mathrm{T}}\beta + g(T) + e, \\ Z = X + u, \\ W = T + v, \end{cases} \tag{2}$$

where u, v and $(X^{\mathrm{T}}, T, e)^{\mathrm{T}}$ are mutually independent, T has an unknown density $f(t)$.

For model identification with errors in nonparametric part, generally assume that v has a known distribution with characteristic function $\phi_v(t)$ (see Fan and Truong [5]). Further we assume that

$$\begin{cases} \mathrm{E}(u) = E(v) = 0, & \mathrm{Cov}(u) = \Sigma_u, \\ \mathrm{E}(e|X, T) = 0, & \mathrm{Var}(e|X, T) = \sigma_e^2, \end{cases}$$

where σ_e^2 is unknown, $\Sigma_u > 0$ is assumed known (for model identification with errors in parametric part, see Zhu and Cui [18]).

In the present paper, for measurement error in nonparametric part T, we use the deconvolution method to estimate the function $g(\cdot)$; for measurement error in parametric part X, apply the parametric correction for attenuation. Based on the deconvolution estimate of $g(\cdot)$, for parameter β, we propose a empirical log-likelihood ratio function with correction for attenuation. It is shown that the empirical log-likelihood ratio at the true parameters converges to the standard chi-square distribution. The construction procedure and main results are described in Section 2. Simulation studies and a real data analysis are presented in Section 3. A concluding remark is presented in Section 4. Proofs are delegated to an Appendix.

2. Methodology and main results

We have a sample $(X_i, W_i, Y_i),\ i = 1, \ldots, n$ of model (2). Let $g_1(t) = \mathrm{E}(X|T = t)$ and $g_2(t) = \mathrm{E}(Y|T = t)$. If covariates X and T are observable, in order to construct empirical likelihood ratio function, we introduce an auxiliary random vector

$$\eta_i(\beta) = \breve{X}_i\{\breve{Y}_i - \breve{X}_i^{\mathrm{T}}\beta\},$$

where $\breve{X}_i = X_i - g_1(T_i),\ \breve{Y}_i = Y_i - g_2(T_i)$. Note that $\mathrm{E}(\eta_i(\beta)) = 0$ if β is the true value of the parameter. Using this, an empirical log-likelihood ratio function evaluated at β is defined as

$$l(\beta) = -2\max\left\{\sum_{i=1}^{n}\log(np_i) :\ p_i \ge 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i\eta_i(\beta) = 0\right\}.$$

However, the covariates X and T are measured with errors, the above empirical likelihood ratio function cannot directly used. First, we need to deal with the two unknown functions $g_1(\cdot)$ and $g_2(\cdot)$ in $l(\beta)$, a natural way is to replace them in $l(\beta)$ by two estimators. In the following, we will use the deconvolution method to estimate them. Next, because of the measurement error in covariate X, in order to avoid the underestimate for parameter, we shall add the correction for attenuation in the construction of auxiliary random vector.

For non-parametric models with measurement errors, Stefanski and Carroll [16], Fan [4], Fan and Truong [5] used a deconvolution method to study the estimation of non-parametric regression function. Let $f(t)$ be the density of T. The deconvolution kernel estimator of the density $f(t)$ is defined as (see Fan and Truong [5], or Stefanski and Carroll [16])

$$\hat{f}_n(t) = \frac{1}{nh}\sum_{j=1}^{n} K_n\left(\frac{t - W_j}{h}\right),$$

where $h = h_n$ is the bandwidth, and $K_n(z)$ is a deconvolution kernel function defined by

$$K_n(z) = \frac{1}{2\pi}\int_{R^1}\exp(-itz)\frac{\phi_K(t)}{\phi_v(t/h)}dt,$$

$\phi_K(\cdot)$ is the Fourier transform of ordinary kernel function $K(\cdot)$ (see condition (C5)) and $\phi_v(\cdot)$ is the characteristic function of the error variable v. Denote

$$\omega_{ni}(\cdot) = K_n\left(\frac{\cdot - W_i}{h}\right)\Big/\sum_j K_n\left(\frac{\cdot - W_j}{h}\right) = \frac{1}{nh}K_n\left(\frac{\cdot - W_i}{h}\right)\Big/\hat{f}_n(\cdot).$$