

Empirical Likelihood Confidence Region for Parameters in Semi-linear Errors-in-Variables Models

HENGJIAN CUI and EFANG KONG

Department of Statistics and Financial Mathematics, Beijing Normal University

ABSTRACT. This paper proposes a constrained empirical likelihood confidence region for a parameter in the semi-linear errors-in-variables model. The confidence region is constructed by combining the score function corresponding to the squared orthogonal distance with a constraint on the parameter, and it overcomes that the solution of limiting mean estimation equations is not unique. It is shown that the empirical log likelihood ratio at the true parameter converges to the standard chi-square distribution. Simulations show that the proposed confidence region has coverage probability which is closer to the nominal level, as well as narrower than those of normal approximation of generalized least squares estimator in most cases. A real data example is given.

Key words: confidence regions, coverage probability, errors-in-variables, empirical likelihood ratio, interval length

1. Introduction

Consider the linear errors-in-variables (EV) regression model

$$Y = x^{\tau} \beta_0 + \epsilon, \quad X = x + u, \quad (1)$$

where β_0 is a $p \times 1$ vector of unknown parameters, x and u are the $p \times 1$ unobservable random covariates and measurement error vectors, respectively, Y is a scalar response and ϵ is the model error. It is assumed that x and $(\epsilon, u^{\tau})^{\tau}$ are independent. In the last two decades, the linear EV model (1) has frequently been used in practice and has attracted considerable attention in the statistical literature. Basic results for model (1) can be found in Fuller (1987). Recently, Carroll *et al.* (1995) discussed estimation of the regression parameter in linear and non-linear EV model, and He & Liang (2000) gave the asymptotic properties for an orthogonal regression approach for model (1).

The empirical likelihood as an alternative to the bootstrap for constructing confidence regions non-parametrically was introduced by Owen (1988). But instead of resampling with equal probability weights like the bootstrap, it works by profiling out a multinomial likelihood supported on the sample. The method defines an empirical likelihood ratio (ELR) 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. In the independent and identically distributed (i.i.d.) data settings, empirical likelihood and its associated properties have been well studied for different statistical models, see Owen (1990), Hall & La Scala (1990), Hall & Owen (1993), Qin & Lawless (1994) and references therein. For independent but not identically distributed variables, Owen (1991) was the first to apply empirical likelihood to regression models, and Qin (1999) and Shi & Lau (2000) introduced the empirical likelihood method into semi-parametric models. Some related work can be found in Chen (1993, 1994), Wang & Jing (1999), Wang & Zhu (2001), Wang & Li (2002), Wang & Rao (2002) and Chen & Cui (2003), among others.

As a generalization of the linear EV model (1), we consider the semi-linear EV model in which the response is assumed to be linearly related to one or more true variables and the

relation to an additional variable (say time t) is assumed to be non-parametric. The specific semi-linear EV model considered here is

$$Y = x^\tau \beta_0 + g(t) + \epsilon, \quad X = x + u, \tag{2}$$

where t is a real-valued variable such that it ranges over a non-degenerate compact interval of one dimension which, without loss of generality, can be the unit interval $[0, 1]$, ϵ is an unobservable error variable and u is a $p \times 1$ unobservable error vector with

$$E[(\epsilon, u^\tau)^\tau] = 0, \quad \text{Cov}[(\epsilon, u^\tau)^\tau] = \sigma^2 I_{p+1}, \tag{3}$$

where $\sigma^2 > 0$ is an unknown parameter, β_0 is a $p \times 1$ vector of unknown parameters, and g is an unknown smooth function of t .

Assumption (3) is needed for identification of model (2), see Fuller (1987, p. 30, pp. 39–42, p. 124). It is assumed that x and u are independent, and $\Sigma_x = \text{Cov}(x)$ and $\Sigma_u = \text{Cov}(u)$ are the covariance matrices of the covariates and the measurement error, respectively. In order to identify model (1), we usually assume Σ_x is a positive definite matrix (PDM) and $\Sigma_u/\text{var}(\epsilon) = \Sigma_0$ is a known $p \times p$ PDM. Without loss of generality (otherwise, transform X by $\Sigma_0^{-1/2}X$, the corresponding β_0 is $\Sigma_0^{1/2}\beta_0$), we assume (3) which means ϵ and u have the same dispersion parameter σ^2 . In some cases, y and x are measured in the same way, e.g. if they are both blood pressure measurements. Hence, it is often reasonable to assume that variances of ϵ and u are equal, see Carroll *et al.* (1995, Section 2.3.2). It is also the standard framework of orthogonal regression taken by Cui & Li (1998), Liang *et al.* (1999, p. 1523) and He & Liang (2000). Another way to identify model (1) is to assume that Σ_u is a known $p \times p$ PDM (or estimator of Σ_u is available). If there are replicated observations from model (2), then assumption (3) can be dropped.

Cui & Li (1998) has considered the estimates of the parameters β_0 , σ^2 and smooth function $g(t)$ for model (2). Under some weak conditions, they showed that the estimators of β_0 and σ^2 are strongly consistent and asymptotically normal. Liang *et al.* (1999) considered the orthogonal regression approach, Zhu & Cui (2003) studied a semi-linear EV model with errors in the linear and the non-linear parts. Qin (1999) and Shi & Lau (2000) constructed the empirical likelihood confidence regions of the parameter β_0 , respectively, in the semi-linear model $Y = x_i^\tau \beta_0 + g(t) + \epsilon$, and proved the confidence regions are consistent. Cui & Chen (2003) constructed a constrained ELR confidence region of β_0 in the linear EV model (1), and showed that the ELR confidence region is consistent as well as Bartlett correctable. They compared the coverage probability, the stability and the length of the constrained ELR confidence intervals with that of the classical generalized least squares (GLS) approach (i.e. based on normal approximation of GLS estimator). They concluded that the ELR confidence region has some superiority over its competitor.

Here we apply the ELR approach for the semi-linear EV model (2) and give a non-parametric version of Wilk’s theorem for $-2 \log$ ELR of β_0 , see section 2 for more details. The rest of the paper is organized as follows. In section 2, we introduce the methodology and main results. Some simulation results and an example are presented in sections 3 and 4 to compare the ELR confidence region with GLS approach, the proof of the main results is given in the appendix.

2. The construction of the constrained ELR confidence region

Suppose that $\{X_i = (X_{i1}, X_{i2}, \dots, X_{ip})^\tau, t_i, Y_i, i = 1, \dots, n\}$ is a sample of size n from model (2). As $g(t)$ is unknown, we define a series of probability weight function $w_{ni}(t)$, which satisfy $\sum_{j=1}^n w_{nj}(t_i) = 1$ ($1 \leq i \leq n$). The kernel weight function, for example, can be taken as

$$w_{ni}(t) = \frac{K((t - t_i)/a_n)}{\sum_{j=1}^n K((t - t_j)/a_n)},$$

where $K(\cdot)$ is a probability density function with $c_1I\{|x| \leq r_0\} \leq K(x) \leq c_2I\{|x| \leq r_0\}$ for some positive constants c_1, c_2 and $r_0, a_n > 0$ is a sequence of smooth parameter, $I\{\cdot\}$ denotes the indicator function. Then a ‘nominal’ estimator of g is defined as the form

$$\hat{g}_n(t) = \sum_{i=1}^n w_{ni}(t)(Y_i - x_i^\tau \beta_0).$$

Replacing $g(t)$ by $\hat{g}_n(t)$ in the first equation of model (2), we get

$$\tilde{Y}_i = \tilde{x}_i^\tau \beta_0 + \tilde{\epsilon}_i^*, \quad \tilde{X}_i = \tilde{x}_i + \tilde{u}_i, \tag{4}$$

where $\tilde{Y}_i = Y_i - \sum_{j=1}^n w_{nj}(t_i)Y_j$, $\tilde{X}_i = X_i - \sum_{j=1}^n w_{nj}(t_i)X_j$, $\tilde{x}_i = x_i - \sum_{j=1}^n w_{nj}(t_i)x_j$, $\tilde{u}_i = u_i - \sum_{j=1}^n w_{nj}(t_i)u_j$, $\tilde{\epsilon}_i^* = \tilde{g}(t_i) + \tilde{\epsilon}_i$ with $\tilde{\epsilon}_i = \epsilon_i - \sum_{j=1}^n w_{nj}(t_i)\epsilon_j$, $\tilde{g}(t_i) = g(t_i) - \sum_{j=1}^n w_{nj}(t_i)g(t_j)$. Let

$$d^2(\beta) = \frac{1}{n} \sum_{i=1}^n \frac{(\tilde{Y}_i - \tilde{X}_i^\tau \beta)^2}{1 + \|\beta\|^2}, \quad \tilde{Z}_{ni}(\beta) = \tilde{X}_i(\tilde{Y}_i - \tilde{X}_i^\tau \beta) + \frac{(\tilde{Y}_i - \tilde{X}_i^\tau \beta)^2 \beta}{1 + \|\beta\|^2}.$$

Note that model (4) is little different from model (1), and $\{\tilde{\epsilon}_i^*\}$ are still i.i.d. approximately with $E(\tilde{\epsilon}_i^*) \approx 0$ since $\tilde{g}(t_i) \approx 0$. Therefore, we treat model (4) as (1) approximately and use the orthogonal regression approach which is presented in Cui & Li (1998), Liang *et al.* (1999) and He & Liang (2000) to define the GLS estimator of β_0 as follows

$$\hat{\beta}_n = \arg \min_{\beta \in R^p} d^2(\beta) = \arg \min_{\beta \in R^p} \frac{1}{n} \sum_{i=1}^n \frac{(\tilde{Y}_i - \tilde{X}_i^\tau \beta)^2}{1 + \|\beta\|^2}.$$

In fact, if ϵ_i, x_i and u_i are assumed to be normal and independent in model (2) without $g(t)$, then the MLE of β is just a solution of minimizing the $d^2(\beta)$, see Fuller (1987, Section 2.2.1). It follows from $\partial d^2(\beta)/\partial \beta \Big|_{\beta=\hat{\beta}_n} = 0$ that

$$\sum_{i=1}^n \left[\tilde{X}_i(\tilde{Y}_i - \tilde{X}_i^\tau \hat{\beta}_n) + \frac{(\tilde{Y}_i - \tilde{X}_i^\tau \hat{\beta}_n)^2 \hat{\beta}_n}{1 + \|\hat{\beta}_n\|^2} \right] = 0, \tag{5}$$

it means that the $\hat{\beta}_n$ satisfies $\sum_{i=1}^n \tilde{Z}_{ni}(\hat{\beta}_n) = 0$, and $\sum_{i=1}^n \tilde{Z}_{ni}(\beta) = 0$ has at least one solution no matter if $E\tilde{Z}_{ni}(\beta)$ is 0 or not. Meanwhile, the ELR evaluated at β by the estimation function $\sum_{i=1}^n \tilde{Z}_{ni}(\beta) = 0$ is defined as

$$R(\beta) = \sup \left\{ \prod_{i=1}^n np_i : \sum_{i=1}^n p_i \tilde{Z}_{ni}(\beta) = 0, p_i \geq 0, \sum_{i=1}^n p_i = 1 \right\}. \tag{6}$$

Intuitively speaking, $\hat{\beta}_n$ should converge to β_0 if β_0 is a unique solution of $\lim_{n \rightarrow \infty} EZ_{ni}(\beta) = 0$, where $\sum_{i=1}^n \tilde{Z}_{ni}(\hat{\beta}_n) = 0$. However the solution is not unique, this implies that the ordinary empirical likelihood confidence region $\{\beta : R(\beta) \geq r\}$ of β_0 is not consistent. We then propose a constrained empirical likelihood confidence region of parameter β_0 with α level as follows:

$$CR_\alpha = \{\beta : R(\beta) \geq r, d^2(\beta) \leq \lambda_s(\hat{\Sigma})\} \tag{7}$$

where $\lambda_s(\hat{\Sigma})$ is the smallest eigenvalue of $\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n \tilde{X}_i \tilde{X}_i^\tau$, $0 \leq r \leq 1$ depends on α .

Shi & Lau (2000) constructed an empirical likelihood confidence region for linear semi-parametric models without errors in the variable (i.e. $u_i = 0$) as $C_{sL} = \{\beta : R(F) \geq r, F \ll F_n\}$. The relationship ($\mu = T(F)$) between the parameter β and the distribution F (discrete distribution on $\{Y_1, \dots, Y_n\}$ with probability p_i) is given by $\sum_{i=1}^n p_i \tilde{X}_i \tilde{\epsilon}_i(\beta) = 0$, where $\tilde{\epsilon}_i(\beta) = Y_i - X_i^T \beta - \hat{g}_n(t)$. Our CR_x is different from the one of Shi & Lau (2000), as it involves a constraint set $\{\beta : d^2(\beta) \leq \lambda_s(\hat{\Sigma})\}$. The restriction is necessary as there are at least two roots (no more $p + 1$ roots) for the equation $\lim_{n \rightarrow \infty} E \tilde{Z}_{ni}(\beta) = 0$ [see (16) in the appendix and $E \tilde{Z}_{ni}(\beta_0) = O(a_n^2) = o(n^{-1/2})$ uniformly in i under the assumptions of the paper]. Thus, if β_0 and $\beta_1 (\neq \beta_0)$ are such two roots and there is no constricton $d^2(\beta) \leq \lambda_s(\hat{\Sigma})$ in CR_x , then

$$P\{\beta_0 \in CR_x\} = P\{\beta_1 \in CR_x\} = 1 - \alpha + o(1)$$

indicating that the probability of confidence region covering the faulty value β_1 is equal to that the probability covers the true value β_0 , and the confidence region is not consistent. More details were given by Cui & Chen (2003). On the contrary, if we impose the restriction $d^2(\beta) \leq \lambda_s(\hat{\Sigma})$, then for $\forall \beta \neq \beta_0, \lim_{n \rightarrow \infty} P\{\beta \in CR_x\} = 0$ (see theorem 2), this means the confidence region is consistent.

The maximum in (6) may be found via Lagrange multipliers, the value for p_i is given by $p_i = 1/[n(1 + \lambda^\tau \tilde{Z}_{ni}(\beta))]$ for fixed β , where λ satisfies $\frac{1}{n} \sum_{i=1}^n \tilde{Z}_{ni}(\beta)/(1 + \lambda^\tau \tilde{Z}_{ni}(\beta)) = 0$. Hence, the $-2 \log$ ELR at β from (6) is

$$l(\beta) = -2 \log R(\beta) = 2 \sum_{i=1}^n (1 + \lambda^\tau \tilde{Z}_{ni}(\beta)).$$

Now we state the main results as follows.

Theorem 1

Under the conditions C1–C5 listed in the appendix, we have

$$l(\beta_0) \xrightarrow{d} \chi_p^2 \text{ as } n \rightarrow \infty.$$

where ‘ \xrightarrow{d} ’ denotes convergence in distribution.

According to theorem 1, we construct an α -level ELR confidence region of β based on (7) as

$$CR_x = \{\beta : l(\beta) \leq C_x, d^2(\beta) \leq \lambda_s(\hat{\Sigma})\}, \tag{8}$$

where C_x satisfies $P\{\chi_p^2 \leq C_x\} = 1 - \alpha$.

Remark 1. From the proof of theorem 2, we have $d^2(\hat{\beta}_n) \leq d^2(\beta_0) \leq \lambda_s(\hat{\Sigma})$, a.s., then $P(d^2(\beta_0) \leq \lambda_s(\hat{\Sigma})) \rightarrow 1$ and $P(\beta_0 \in CR_x) = P(l(\beta_0) \leq C_x, d^2(\beta_0) \leq \lambda_s(\hat{\Sigma})) \rightarrow 1 - \alpha$. The asymptotic properties of $l(\beta_0)$ are mainly dependent on those of the four quantities $n^{-1/2} \sum_i \tilde{Z}_{ni}(\beta_0)$, $n^{-1} \sum_i \tilde{Z}_{ni}(\beta_0) \tilde{Z}_{ni}^c(\beta_0)$, $n^{-1} \sum_i \|\tilde{Z}_{ni}(\beta_0)\|^2$ and $\max_i \|\tilde{Z}_{ni}(\beta_0)\|$. Moreover, these four quantities can be replaced by $n^{-1/2} \sum_i Z_{ni}(\beta_0)$, $n^{-1} \sum_i Z_{ni}(\beta_0) Z_{ni}^c(\beta_0)$, $n^{-1} \sum_i \|Z_{ni}(\beta_0)\|^2$ and $\max_i \|Z_{ni}(\beta_0)\|$ respectively (see smoothing approach in lemma 4), where $Z_{ni}(\beta_0) = X_i(Y_i - X_i^T \beta_0) + (Y_i - X_i^T \beta_0)^2 \beta_0 / (1 + \|\beta_0\|^2)$ in model (1). Rather than weighted chi-square distribution, $l_0(\beta_0) \xrightarrow{d} \chi_p^2$ and (16) in the appendix yield $l(\beta_0) \xrightarrow{d} \chi_p^2$ where $l_0(\beta_0)$ stands for $-2 \log$ ELR at β_0 corresponding to $Z_{ni}(\beta_0)$.

Theorem 2

$\{\beta : d^2(\beta) \leq \lambda_s(\hat{\Sigma})\}$ is a convex set and not empty as n is large enough. Moreover, if $\beta \in R^p$ is a constant vector and $\beta \neq \beta_0$, then $\lim_{n \rightarrow \infty} P\{\beta \in CR_x\} = 0$, where CR_x is in (7).

Theorem 3

If $\beta^* = \beta_0 - n^{-1/2} \Sigma_v^{-1} \Omega^{1/2} \gamma$ for $\gamma \in R^p$ is a constant vector, then under conditions C1–C5 listed in the appendix, we have

$$l(\beta^*) \xrightarrow{d} \chi_p^2(\|\gamma\|^2) \quad (n \rightarrow \infty),$$

where $\Sigma_v = \text{Cov}(v_i)$ and v_i is shown in condition C1, $\|\gamma\|^2$ is the non-central parameter and Ω is defined as

$$\Omega = (1 + \|\beta_0\|^2) \sigma^2 \Sigma + \text{Cov} \left[(\epsilon_i - u_i^\tau \beta_0) u_i + (\epsilon_i - u_i^\tau \beta_0)^2 \frac{\beta_0}{(1 + \|\beta_0\|^2)} \right]. \tag{9}$$

3. Simulations

In this section, we report results from a simulation study designed to evaluate the performance of the proposed ELR confidence region (8) and compare it with the GLS confidence region (i.e. based on the normal approximation of GLS estimator $\hat{\beta}_n$). Our simulations show that the ELR confidence region has coverage probability close to the nominal level and is narrower than that of GLS for most cases.

Cui & Li (1998, theorems 1 and 2) showed that the GLS estimator $\hat{\beta}_n$ of β_0 for linear semi-linear EV models with random covariate variable t is strongly consistent and asymptotically normal. We can adopt the same strategy as in Cui & Li (1998) and lemmas 1–4 to get that our GLS estimator $\hat{\beta}_n$ of β_0 for semi-linear EV models with deterministic covariate variable t in this paper also has the following properties:

$$\hat{\beta}_n \xrightarrow{a.s.} \beta_0, \quad \hat{\sigma}_n^2 \xrightarrow{a.s.} \sigma^2, \quad \sqrt{n} \Omega^{-1/2} \Sigma_v (\hat{\beta}_n - \beta_0) \xrightarrow{d} N(0, I_p),$$

$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n \tilde{X}_i \tilde{X}_i^\tau - \hat{\sigma}_n^2 I_p \xrightarrow{a.s.} \Sigma_v$ and $\frac{1}{n} \sum_{i=1}^n \tilde{Z}_{ni}(\beta_0) \tilde{Z}_{ni}^\tau(\beta_0) = \Omega + o(1)$ a.s., where $\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (\tilde{Y}_i - \tilde{X}_i^\tau \hat{\beta}_n)^2 / (1 + \|\hat{\beta}_n\|^2)$, see also Liang *et al.* (1999) for special case. We shall use $\hat{\Omega}_n = \frac{1}{n} \sum_{i=1}^n \tilde{Z}_{ni}(\hat{\beta}_n) \tilde{Z}_{ni}^\tau(\hat{\beta}_n)$ to estimate Ω , it can be derived that $\hat{\Omega}_n$ is consistent estimator of Ω from lemmas 3 and 4. In this sense, we have $\sqrt{n} \hat{\Omega}_n^{-1/2} \hat{\Sigma} (\hat{\beta}_n - \beta_0) \xrightarrow{d} N(0, I_p)$.

As $p = 1$, $\hat{\beta}_n$ has an explicit form,

$$\hat{\beta}_n = \frac{2 \sum_{i=1}^n \tilde{X}_i \tilde{Y}_i}{\sqrt{(\sum_{i=1}^n (\tilde{Y}_i^2 - \tilde{X}_i^2))^2 + 4(\sum_{i=1}^n \tilde{X}_i \tilde{Y}_i)^2 - \sum_{i=1}^n (\tilde{Y}_i^2 - \tilde{X}_i^2)}}.$$

Then, a two-side confidence interval for β_0 with level α is given by

$$(\hat{\beta}_n - Z_{1-\alpha/2} \hat{\Omega}_n^{1/2} / (\sqrt{n} \hat{\Sigma}), \hat{\beta}_n + Z_{1-\alpha/2} \hat{\Omega}_n^{1/2} / (\sqrt{n} \hat{\Sigma})), \tag{10}$$

where $\Phi(Z_\alpha) = \alpha$, $\Phi(\cdot)$ is the cdf of the standard normal distribution. The explicit form of $\hat{\beta}_n$ cannot be expressed when $p \geq 2$, but the numerical solution of $\hat{\beta}_n$ can be given by using the eigenvector decomposition. Based on the asymptotic normality of $\hat{\beta}_n$, we construct the GLS confidence region as follows:

$$\text{CR}_\alpha = \{\beta : n(\beta - \hat{\beta}_n)^\tau \hat{\Sigma}_1 (\beta - \hat{\beta}_n) \leq C_\alpha\}, \tag{11}$$

where C_α satisfied $P(\chi_p^2 \leq C_\alpha) = 1 - \alpha$ and $\hat{\Sigma}_1 = \hat{\Sigma} \hat{\Omega}_n^{-1} \hat{\Sigma}$.

We shall compare the performance of our ELR with GLS approach. Let $t_i = i/n$ be equal spaced points on $[0, 1]$, $g(t) = \sin(\pi t)$, $\epsilon \sim N(0, \sigma^2)$. The kernel method is used to determine a

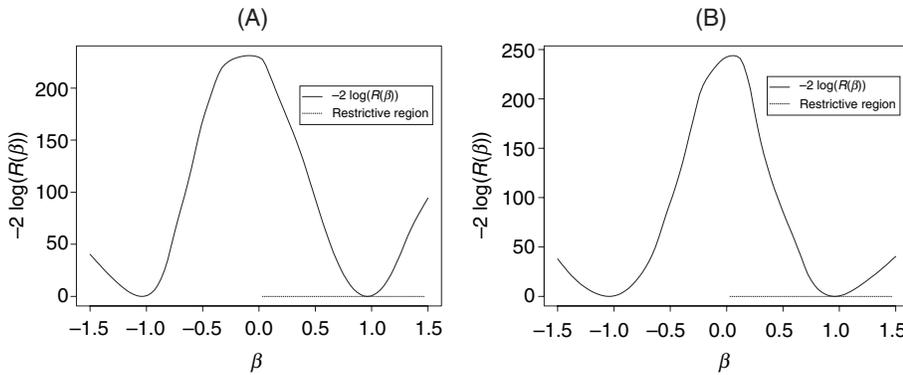


Fig. 1. (A) $-2 \log R(\beta)$ versus β for case 2. (B) $-2 \log R(\beta)$ versus β for case 4.

weight function $w_n(t)$ with $K(u) = (3/4)(1 - u^2)I\{|u| \leq 1\}$ and bandwidth $a_n = 1.06s_n n^{-1/3} \approx 0.306n^{-1/3}$ (s_n is the standard deviation of $\{t_i\}$), the bandwidth h used here is the histogram bandwidth, it may not be optimal but an experimental choice. Cross-validation criteria is helpful for choosing optimal h . The four cases of distributions for u and v below are considered, where cases 1 and 2 are for $p = 1$ with $\beta_0 = 1$ and $h(t) = t$, cases 3 and 4 are for $p = 2$ with $\beta_0 = (1, 1)^T$ and $h(t) = (t, t^2)^T$.

Case 1. $v \sim N(0, 1)$ and $u \sim N(0, \sigma^2)$.

Case 2. $v \sim U(-\sqrt{3}, \sqrt{3})$ and $u \sim U(-\sqrt{3}\sigma, \sqrt{3}\sigma)$.

Case 3. $v \sim U(D_1)$ and $u \sim U(D_2)$, where $U(D_i)$ denotes the uniform distribution on D_i ($i = 1, 2$) with $D_1 = [-\sqrt{3}, \sqrt{3}]^2$ and $D_2 = [-\sqrt{3}\sigma, \sqrt{3}\sigma]^2$.

Case 4. $v \sim N(0, I_2)$ and $u \sim N(0, \sigma^2 I_2)$, where I_2 is the 2×2 unit matrix.

First, we display the curves of $-2 \log(R(\beta))$ by ELR approach in Fig. 1A, B for cases 1 and 2, respectively, with $\sigma = 0.3$; indicating that the $-2 \log(R(\beta))$ reaches minimum value 0 at two points which are very close to the true value $\beta_0 = 1$ and the faulty value $\beta = -1$, respectively. Moreover, the curve near those two points is approximately in quadratic form and the set $\{\beta : -2 \log(R(\beta)) < C_\alpha\}$ is actually a union of two separate intervals. Our constrictive region $\{d^2(\beta) \leq \lambda_s(\hat{\Sigma})\}$ remains just the one covering the true value and gets rid of the faulty one. For $p = 2$, case 3 with $n = 50$ and 100 are considered, respectively. The 95% confidence regions of ELR and GLS are shown in Fig. 2(A) for $n = 100$, and Fig. 2(B) for $n = 50$. The two confidence regions are close for $n = 100$, but there are some differences for $n = 50$. The area of ELR confidence region seems more smaller than that of GLS confidence region and the true vector β_0 is more close to the centre of ELR confidence region. The ELR confidence region is overall rather reasonable.

Next, we shall make a comparison of coverage probability. Sample size is chosen to be 30, 50, 100 and 200 and the level α is fixed at 0.05. The coverage probabilities are calculated for the ELR and GLS methods based on 1000 replications. Tables 1 and 2 present coverage probability comparisons. It can be shown that the ELR outperforms the GLS method in all four cases considered as the coverage probability of ELR confidence region is close to designed probability than that of GLS. It is also interesting to note that the coverage accuracies for both generally increase to the designed probability as the sample size n becomes larger.

The case 1 is used for comparison of the length ratio (ELR confidence interval length/GLS confidence interval length) of ELR and GLS confidence intervals. The result is shown in

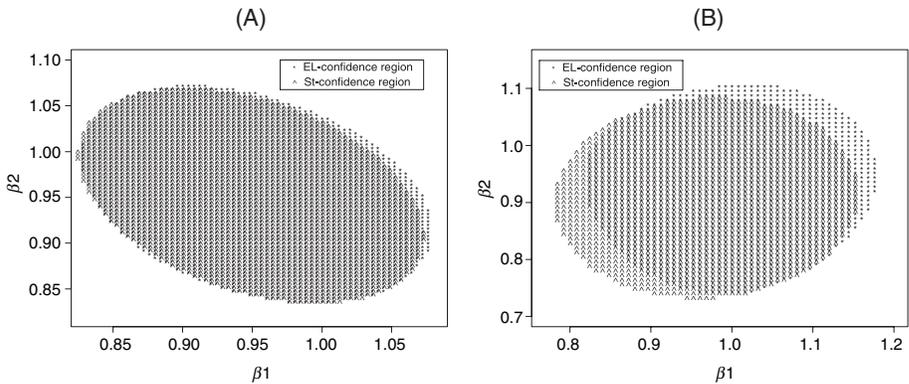


Fig. 2. (A) Confidence region for $p = 2$ with $n = 100$. (B) Confidence region for $p = 2$ with $n = 50$.

Table 1. The coverage probability of confidence regions when $\sigma = 0.3$

n	Case 1		Case 2		Case 3		Case 4	
	ELR	GLS	ELR	GLS	ELR	GLS	ELR	GLS
30	0.849	0.844	0.843	0.820	0.836	0.830	0.833	0.829
50	0.881	0.878	0.907	0.895	0.883	0.875	0.878	0.869
100	0.926	0.922	0.929	0.914	0.925	0.920	0.917	0.912
200	0.934	0.925	0.937	0.928	0.932	0.930	0.928	0.923

Table 2. The coverage probability of confidence regions when $\sigma = 0.9$

n	Case 1		Case 2		Case 3		Case 4	
	ELR	GLS	ELR	GLS	ELR	GLS	ELR	GLS
30	0.848	0.847	0.842	0.782	0.834	0.825	0.831	0.823
50	0.894	0.890	0.906	0.819	0.869	0.867	0.886	0.875
100	0.931	0.927	0.937	0.873	0.919	0.908	0.921	0.918
200	0.940	0.935	0.939	0.927	0.937	0.932	0.939	0.934

Table 3. The length ratio of confidence intervals for case 1

n	$\sigma = 0.3$	$\sigma = 0.6$	$\sigma = 0.9$
30	0.908 (0.009)	1.038 (0.183)	1.130 (0.398)
50	0.873 (0.078)	0.985 (0.090)	1.107 (0.215)
100	0.788 (0.112)	0.923 (0.055)	1.006 (0.065)
200	0.696 (0.145)	0.878 (0.074)	0.940 (0.047)

Value represent average (SD).

Table 3. The length ratio is smaller than 1 when $\sigma = 0.3$, $\sigma = 0.6$ and $n \geq 50$, and $\sigma = 0.9$ and $n \geq 200$. It means that the ELR confidence intervals has shorter length than that of GLS if σ is smaller relatively and the sample size is large. This is reasonable due to ELR method's automatic determination of shape and orientation of the confidence region by the data. Usually, the length of ELR confidence interval is shorter than that of GLS confidence interval in the linear EV model, see Gao & Cui (2001) and Cui & Chen (2003). Unlike linear EV model,

the smoothing method is applied to semi-linear EV model which leads a non-i.i.d. model (4), it might reduce the accuracy of ELR confidence interval when the variance of *error* is large relatively and the sample size is small. This phenomenon may be dependent on the second-order property of ELR approach.

4. Example

Finally, we demonstrate the comparisons of both confidence intervals through an example. We consider the relationship between body height and arm span (x) and age using a sample of $n = 32$ observations (the data set can be downloaded from the web-page: <http://math.bnu.edu.cn/~chj/data1.ps>). This data set comes from the junior class of statistical major in Beijing Normal University, China. Students measured their body height and arm span by the same roll ruler. The pairs scatter plots of observed body height (Y : in cm) versus age (in months), observed arm span (X : in cm) versus age, and Y versus X are shown in Fig. 3 A–C, respectively. We would like to establish the relationship of Y and (x , age). Fig. 3(C) shows us there is a statistical linear relationship between Y and X , where X is observed variable of x with error u whose variance is assumed to be approximately equal to the variance of error due to the same measurement tool. On the contrary, we seem not to

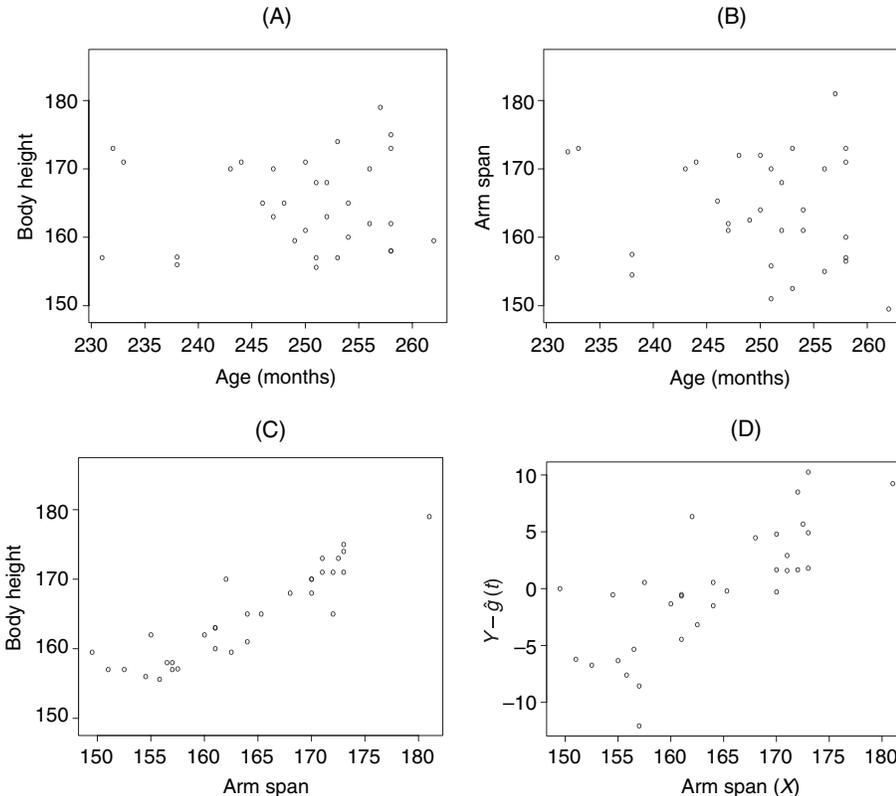


Fig. 3. The plots of (A) Body height versus age, (B) arm span versus age, (C) body height versus arm span, (D) $Y - \hat{g}(t)$ versus arm span (X).

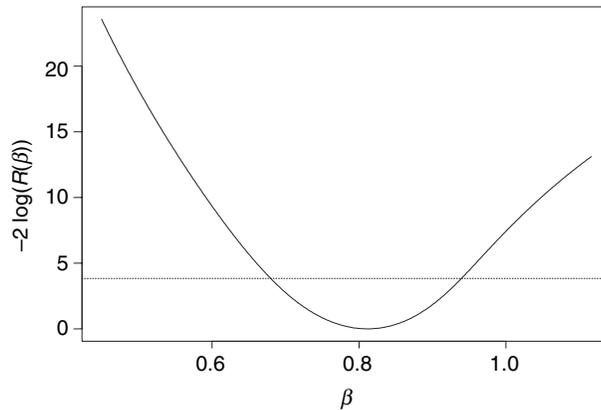


Fig. 4. The plot of $-2 \log R(\beta)$ versus β .

know exactly what relationship between Y and age from Fig. 3(A). So we adopt a working semi-linear EV model,

$$Y = x\beta + g(t) + \text{error}, \quad X = x + u.$$

where $t = m/31 \in [0, 1]$, $m = \text{age} - 231$ stands for the months with respect to minimum age (231 months and birth date is 9/1985), 31 is the range of such relative months. The bandwidth h is also taken as $h = 1.06s_n n^{-1/3}$. The plot of $Y - \hat{g}(t)$ versus observed arm span (X) is shown in Fig. 3(D). In this sense, we get GLS estimator $\hat{\beta} = 0.812$, the GLS confidence interval with level $\alpha = 0.05$ is (0.679, 0.944) and its confidence interval length is 0.265. Meanwhile, the ELR confidence interval with level $\alpha = 0.05$ is (0.682, 0.938) and the confidence interval length is 0.256. This implies the ELR confidence interval length is shorter than that of GLS method with the same level. Figure 4 shows the plot of $-2 \log(R(\beta))$ versus β .

Acknowledgements

We would like to thank the Associate Editor, the Editor and two referees for their truly helpful comments and suggestions which led to a much improved presentation. The research was partially supported by the NSFC and the RFDP of China.

References

- Carroll, R. J., Ruppert, D. & Stefanski, L. A. (1995). *Measurement error in nonlinear models*. Chapman & Hall, London.
- Chen, S. X. (1993). On the accuracy of empirical likelihood confidence regions for linear regression model. *Ann. Inst. Statist. Math.* **45**, 627–637.
- Chen, S. X. (1994). Empirical likelihood confidence intervals for linear regression coefficients. *J. Multivar. Anal.* **49**, 24–40.
- Chen, S. X. & Cui, H. J. (2003). An extend empirical likelihood for generalized linear model. *Statist. Sinica* **13**, 69–81.
- Cui, H. J. & Chen, S. X. (2003). Empirical likelihood confidence region for parameter in the errors-in-variables models. *J. Multivar. Anal.* **84**, 101–115.
- Cui, H. J. & Li, R. C. (1998). On parameter estimation for semi-linear errors-in-variables models. *J. Multivar. Anal.* **64**, 1–24.
- Fuller, W. A. (1987). *Measurement error models*. John Wiley, New York.

- Gao, J. T., Hong, S. Y. & Liang, H. (1995). Convergence rates of a class of estimators in partly linear models. *Acta Mathematica Sinica* **38**, 658–669.
- Gao, X. H. & Cui, H. J. (2001). Empirical likelihood ratio confidence regions in EV linear model. *J. Beijing Normal Univ. (Nat. Sci.)* **37**, 586–591.
- Hall, P. & La Scala, B. (1990). Methodology and algorithms of empirical likelihood. *Int. Statist. Rev.* **58**, 109–127.
- Hall, P. & Owen, A. (1993). Empirical likelihood confidence bands in density estimation. *J. Comput. Graph. Statist.* **2**, 273–289.
- He, X. & Liang, H. (2000). Quantile regression estimates for a class of linear and partially linear errors-in-variables models. *Statist. Sinica* **10**, 129–140.
- Liang, H., Härdle, W. & Carroll, R. J. (1999). Estimation in a semiparametric partially linear errors-in-variables model. *Ann. Statist.* **27**, 1519–1535.
- Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75**, 237–249.
- Owen, A. B. (1990). Empirical likelihood ratio confidence regions. *Ann. Statist.* **18**, 90–120.
- Owen, A. B. (1991). Empirical likelihood for linear models. *Ann. Statist.* **18**, 121–140.
- Qin, Y. S. (1999). Empirical likelihood ratio confidence regions in a partly linear model. *Chinese J. Appl. Probab. Statist.* **15**, 363–369.
- Qin, J. & Lawless, J. (1994). Empirical likelihood and general estimation equations. *Ann. Statist.* **22**, 300–325.
- Shi, J. & Lau, T. S. (2000). Empirical likelihood for partially linear models. *J. Multivar. Anal.* **72**, 132–148.
- Wang, Q. H. & Jing, B. Y. (1999). Empirical likelihood for partial linear models with fixed designs. *Statist. Probab. Lett.* **41**, 425–433.
- Wang, Q. H. & Li, G. (2002). Empirical likelihood semiparametric regression analysis under random censorship. *J. Multivar. Anal.* **83**, 469–486.
- Wang, Q. H. & Rao, J. N. K. (2002). Empirical likelihood-based inference in linear errors-in-covariables models with validation data. *Biometrika* **89**, 345–358.
- Wang, Q. H. & Zhu, L. X. (2001). Estimation in partial linear error-in-covariables models with censored data. *Comm. Statist. Theory Methods* **30**, 41–54.
- Zhu, L. X. & Cui, H. J. (2003). Semiparametric regression model with errors in variables. *Scand. J. Statist.* **30**, 429–442.

Received May 2004, in final form May 2005

Hengjian Cui, Department of Statistics and Financial Mathematics, School of Mathematical Sciences, Beijing Normal University, Beijing, 100875, China.
E-mail: hjcui@bnu.edu.cn

Appendix: Proof of main results

Some conditions used in the paper are listed below:

(C1) $x_i = h(t_i) + v_i$, where $h(t) = (h_1(t), h_2(t), \dots, h_p(t))^T$, $v_i = (v_{i1}, \dots, v_{ip})^T \in R^p$ are unknown random vectors with $E(v_i) = 0$, $E\|v_i\|^4 < \infty$ for $1 \leq i \leq n$ and $\|\cdot\|$ is Euclidean norm.

(C2) $g, h_j (1 \leq j \leq p) \in \text{Lip}$, where

$$\text{Lip} =: \{f : \sup_{s,t \in [0,1]} |f(s) - f(t)| \leq C|s - t| \text{ for some positive constant } C\}.$$

(C3) $0 \leq t_0 \leq t_1 \leq \dots \leq t_{n-1} \leq t_n = 1$, and $\max_{1 \leq i \leq n} |t_i - t_{i-1}| = O(\log n/n)$.

(C4) For some positive sequence a_n with $na_n^4 \rightarrow 0$ and $\sqrt{na_n}/\log^3 n \rightarrow \infty$, $w_{ni}(\cdot)$ satisfies:

- (1) $\sup_{t \in [0,1]} \max_{1 \leq i \leq n} w_{ni}(t) = O(\log n/(na_n))$.
- (2) $\sup_{t \in [0,1]} \sum_{i=1}^n w_{ni}(t) I\{|t_i - t| > a_n\} = O(a_n)$.

Note that the conditions (C1–C3) are used commonly in non-parametric estimation, see Gao *et al.* (1995) and Qin (1999).

$$(C5) E(\epsilon_i^4 + \|u_i\|^4) < \infty.$$

Remark 2. If T_1, T_2, \dots, T_n are i.i.d. $\sim U[0, 1]$ and take $t_i = T_{(i)} (1 \leq i \leq n)$, where $T_{(1)} \leq T_{(2)} \leq \dots \leq T_{(n)}$, then the condition C3 holds automatically. In this sense, the results in the paper remain true for random $t_i = T_{(i)}$. Under the condition C3, if we take the kernel weight function $w_n(t)$ as shown in section 2, then (1) and (2) hold in condition C4. C5 means that $\max_{1 \leq i \leq n} |\epsilon_i| = o(n^{\frac{1}{4}})$ a.s. and $\max_{1 \leq i \leq n} \|u_i\| = o(n^{\frac{1}{4}})$ a.s.

Lemma 1

Let $\xi_1, \xi_2, \dots, \xi_n$ be i.i.d. random variables with 0 mean and $E|\xi_i|^r < \infty (r \geq 1)$. Assume that $\{a_{ki} : k, i = 1, 2, \dots, n\}$ is a sequence of numbers. Then

$$\max_{1 \leq i \leq n} \left| \sum_{k=1}^n a_{ki} \xi_k \right| = O\left((n^{1/r} d_n \vee \sqrt{d_n s_n}) \log n \right) \text{ a.s.}$$

where $d_n = \max_{1 \leq k, i \leq n} |a_{ki}|$ and $s_n = \max_{1 \leq i \leq n} \sum_{k=1}^n |a_{ki}|$.

The proof of lemma 1 can follow completely the proof of lemma 1 in Cui & Li (1998) or Shi & Lau (2000), we omit it here. The following lemmas are necessary for the proof of theorem 1.

Lemma 2

Let $(\xi_1, \eta_1), (\xi_2, \eta_2), \dots, (\xi_n, \eta_n)$ be bivariate i.i.d. random vectors with 0 mean and $E(|\xi_i|^4 + |\eta_i|^4) < \infty$; $\zeta_{n1}, \zeta_{n2}, \dots, \zeta_{nm}$ be a random sequence with $\max_{1 \leq i \leq n} |\zeta_{ni}| = o(1)$ a.s.; b_1, b_2, \dots, b_n be a sequence with $\max_{1 \leq i \leq n} |b_i| = O(a_n)$, then under the condition (C3) and (C4),

$$(i) E \left(\sum_{i=1}^n b_i \bar{\xi}_i \right)^2 = O(n a_n \log n),$$

$$(ii) E \left(\sum_{i=1}^n \xi_i \bar{\eta}_i \right)^2 = O((a_n^{-1} \log n)^2),$$

$$(iii) \frac{1}{n} \sum_{i=1}^n \zeta_{ni} \xi_i = o(1) \text{ a.s.}$$

where $\bar{\xi}_i = \sum_{j=1}^n w_{nj}(t_i) \xi_j$, $\bar{\eta}_i = \sum_{j=1}^n w_{nj}(t_i) \eta_j$.

Proof. From the assumption of lemma 2, we have

$$(i) E \left(\sum_{i=1}^n b_i \bar{\xi}_i \right)^2 = \sum_{j=1}^n \left[\sum_{i=1}^n b_i w_{nj}(t_i) \right]^2 E \zeta_j^2 \leq C a_n^2 \sum_{j=1}^n \sum_{i=1}^n w_{nj}(t_i) n \log n / (n a_n) = O(n a_n \log n),$$

$$\begin{aligned}
 (ii) \ E \left(\sum_{i=1}^n \xi_i \bar{\eta}_i \right)^2 &= \sum_{i,j,k,l} w_{nj}(t_i) w_{nl}(t_k) E(\xi_i \xi_k \eta_j \eta_l) \\
 &\leq C \left[\sum_i w_{ni}^2(t_i) + \sum_{i,j} w_{nj}^2(t_i) + \left(\sum_i w_{ni}(t_i) \right)^2 \right] \\
 &= O((\log n/a_n)^2),
 \end{aligned}$$

$$(iii) \ \left| \frac{1}{n} \sum_{i=1}^n \zeta_{in} \zeta_i \right| \leq \max_{1 \leq i \leq n} |\zeta_{in}| \left| \frac{1}{n} \sum_{i=1}^n \zeta_i \right| = o(1) \text{ a.s.}$$

Lemma 3

Under the conditions C3 and C4, then

$$\sup_{0 \leq t \leq 1} |\tilde{f}(t)| = O(a_n) \quad \text{and} \quad \max_{1 \leq i \leq n} \left\| \sum_{j=1}^n w_{nj}(t_i) \eta_j \right\| = O\left(\frac{\log^{3/2} n}{\sqrt{na_n}} \right) \text{ a.s.}$$

where $\tilde{f}(t) = f(t) - \sum_{j=1}^n w_{nj}(t) f(t_j)$ with $f(t) \in Lip$, $\eta_1, \eta_2, \dots, \eta_n$ are i.i.d. random vectors with 0 mean and $E\|\eta_i\|^4 < \infty$.

Proof. Since $f \in Lip$,

$$\begin{aligned}
 \tilde{f}(t) &= \sum_{j=1}^n w_{nj}(t) [f(t) - f(t_j)] \\
 &= \sum_{j=1}^n w_{nj}(t) [f(t) - f(t_j)] I\{|t - t_j| \leq a_n\} + \sum_{j=1}^n w_{nj}(t) [f(t) - f(t_j)] I\{|t - t_j| > a_n\},
 \end{aligned}$$

it follows from conditions C3 and C4 that $|\tilde{f}(t)| \leq Ca_n$. From the condition C4, we obtain $d_n = \max_{1 \leq i,j \leq n} w_{nj}(t_i) = O(\log n/(na_n))$, $s_n = \max_{1 \leq i \leq n} \sum_{j=1}^n w_{nj}(t_i) = 1$, $n^{1/4} d_n \leq \sqrt{d_n}$ when n is large, then by lemma 1 with $r = 4$, we have

$$\max_{1 \leq i \leq n} \left\| \sum_{j=1}^n w_{nj}(t_i) \eta_j \right\| = O(\sqrt{d_n} \log n) = O\left(\log^{3/2} \frac{n}{\sqrt{na_n}} \right) \text{ a.s.}$$

The proof is complete.

Lemma 4

Under the above conditions C2–C5, we have

$$n^{-1/2} \sum_{i=1}^n \tilde{Z}_{ni}(\beta_0) \xrightarrow{d} (0, \Omega), \quad \frac{1}{n} \sum_{i=1}^n \tilde{Z}_{ni}(\beta_0) \tilde{Z}_{ni}^c(\beta_0) = \Omega + o(1) \text{ a.s.},$$

$$\max_{1 \leq i \leq n} \|\tilde{Z}_{ni}(\beta_0)\| = o(n^{1/2}) \text{ a.s.} \quad \text{and} \quad \frac{1}{n} \sum_{i=1}^n \|\tilde{Z}_{ni}(\beta_0)\|^3 = o(n^{1/2}) \text{ a.s.},$$

where $\Omega = (1 + \|\beta_0\|^2) \sigma^2 \Sigma + Cov[(\epsilon_i - u_i^c \beta_0) u_i + (\epsilon_i - u_i^c \beta_0)^2 \beta_0 / (1 + \|\beta_0\|^2)]$.

Proof. Since

$$\begin{aligned} \tilde{Z}_{ni}(\beta_0) &= \tilde{X}_i(\tilde{Y}_i - \tilde{X}_i^\tau \beta_0) + \frac{(\tilde{Y}_i - \tilde{X}_i^\tau \beta_0)^2 \beta_0}{1 + \|\beta_0\|^2} \\ &= [v_i + u_i + \tilde{h}(t_i) - (\bar{v}_i + \bar{u}_i)][\epsilon_i - u_i^\tau \beta_0 + \tilde{g}(t_i) - \bar{\epsilon}_i + \bar{u}_i^\tau \beta_0] \\ &\quad + [\epsilon_i - u_i^\tau \beta_0 + \tilde{g}(t_i) - \bar{\epsilon}_i + \bar{u}_i^\tau \beta_0]^2 \beta_0 / (1 + \|\beta_0\|^2) =: J_{1i} + J_{2i} \beta_0 / (1 + \|\beta_0\|^2), \end{aligned}$$

it is suffice to prove that

$$n^{-1/2} \sum_{i=1}^n J_{1i} = n^{-1/2} \sum_{i=1}^n (v_i + u_i)(\epsilon_i - u_i^\tau \beta_0) + o_p(1) \tag{12}$$

and

$$n^{-1/2} \sum_{i=1}^n J_{2i} = n^{-1/2} \sum_{i=1}^n (\epsilon_i - u_i^\tau \beta_0)^2 + o_p(1). \tag{13}$$

We only prove (12), (13) is proved by the similar way. Denote $\bar{u}_i = \sum_{j=1}^n w_{nj}(t_i) u_j$, $\bar{v}_i = \sum_{j=1}^n w_{nj}(t_i) v_j$, $\bar{\epsilon}_i = \sum_{j=1}^n w_{nj}(t_i) \epsilon_j$. It follows from the condition C5 and lemma 3 that $\max_{1 \leq i \leq n} |\tilde{h}(t_i) \tilde{g}(t_i)| = O(a_n^2)$, $\sum_{i=1}^n \tilde{h}(t_i)(\epsilon_i - u_i^\tau \beta_0) = O_p(n^{1/2} a_n)$, $\sum_{i=1}^n \tilde{g}(t_i)(v_i + u_i) = O_p(n^{1/2} a_n)$, and $\max_{1 \leq i \leq n} |(\bar{v}_i + \bar{u}_i)(\bar{\epsilon}_i - \bar{u}_i^\tau \beta_0)| = O(\log^3 n / (na_n))$ a.s.

Using (i) and (ii) in lemma 2, we get

$$\sum_{i=1}^n \tilde{h}(t_i)(\bar{\epsilon}_i - \bar{u}_i^\tau \beta_0) = O_p((na_n \log n)^{1/2}), \quad \sum_{i=1}^n \tilde{g}(t_i)(\bar{v}_i + \bar{u}_i) = O_p((na_n \log n)^{1/2}),$$

$$\sum_{i=1}^n (v_i + u_i)(\bar{\epsilon}_i - \bar{u}_i^\tau \beta_0) = O_p\left(\frac{\log n}{a_n}\right), \quad \sum_{i=1}^n (\bar{v}_i + \bar{u}_i)(\epsilon_i - u_i^\tau \beta_0) = O_p\left(\frac{\log n}{a_n}\right).$$

Therefore,

$$\begin{aligned} n^{-1/2} \sum_{i=1}^n J_{1i} &= n^{-1/2} \sum_{i=1}^n (v_i + u_i)(\epsilon_i - u_i^\tau \beta_0) + O_p(a_n) + O_p\left(\frac{\log^3 n}{(\sqrt{na_n})}\right) \\ &\quad + O_p(\sqrt{na_n^2}) + O_p((a_n \log n)^{1/2}) \\ &= n^{-1/2} \sum_{i=1}^n (v_i + u_i)(\epsilon_i - u_i^\tau \beta_0) + o_p(1). \end{aligned}$$

Moreover, it can be derived by using the condition C5 and lemma 3 again that

$$\max_{1 \leq i \leq n} |\tilde{h}(t_i)(\epsilon_i - u_i^\tau \beta_0)| = o(n^{1/4} a_n) \text{ a.s.}, \quad \max_{1 \leq i \leq n} |\tilde{g}(t_i)(v_i + u_i)| = o(n^{1/4} a_n) \text{ a.s.},$$

$$\max_{1 \leq i \leq n} |\tilde{h}(t_i)(\bar{\epsilon}_i - \bar{u}_i^\tau \beta_0)| = o(n^{-5/8} \log^{3/2} n) \text{ a.s.},$$

$$\max_{1 \leq i \leq n} |\tilde{g}(t_i)(\bar{v}_i + \bar{u}_i)| = o(n^{-5/8} \log^{3/2} n) \text{ a.s.},$$

$$\max_{1 \leq i \leq n} |(v_i + u_i)(\bar{\epsilon}_i - \bar{u}_i^\tau \beta_0)| = o\left(\sqrt{\frac{\log^3 n}{(\sqrt{na_n})}}\right) \text{ a.s.},$$

$$\max_{1 \leq i \leq n} |(\bar{v}_i + \bar{u}_i)(\epsilon_i - u_i^\tau \beta_0)| = o\left(\sqrt{\frac{\log^3 n}{(\sqrt{na_n})}}\right) \text{ a.s.}$$

Thus, we have from (iii) in lemma 2 that

$$J_{1i} = (v_i + u_i)(\epsilon_i - u_i^\tau \beta_0) + o(1) \text{ a.s. and } J_{2i} = (\epsilon_i - u_i^\tau \beta_0)^2 + o(1) \text{ a.s.}$$

uniformly in i . This means by (iii) in lemma 2 and the strong law of large numbers that

$$\begin{aligned} \frac{1}{n} \tilde{Z}_{ni}(\beta_0) \tilde{Z}_{ni}(\beta_0)^\tau &= \frac{1}{n} \sum_{i=1}^n \left[J_{1i} + \frac{J_{2i} \beta_0}{(1 + \|\beta_0\|^2)} \right] \left[J_{1i} + \frac{J_{2i} \beta_0}{(1 + \|\beta_0\|^2)} \right]^\tau \\ &= \frac{1}{n} \sum_{i=1}^n \left[(v_i + u_i)(\epsilon_i - u_i^\tau \beta_0) + \frac{(\epsilon_i - u_i^\tau \beta_0)^2 \beta_0}{1 + \|\beta_0\|^2} \right] \left[(v_i + u_i)(\epsilon_i - u_i^\tau \beta_0) \right. \\ &\quad \left. + \frac{(\epsilon_i - u_i^\tau \beta_0)^2 \beta_0}{1 + \|\beta_0\|^2} \right]^\tau + o(1) \\ &= \Omega + o(1) \text{ a.s.} \end{aligned}$$

and $\max_{1 \leq i \leq n} \|\tilde{Z}_{ni}(\beta_0)\| \leq \max_{1 \leq i \leq n} \|J_{1i}\| + \max_{1 \leq i \leq n} \|J_{2i}\| = o(n^{1/2})$.

From the proof above, we obtain $\frac{1}{n} \sum_{i=1}^n \|\tilde{Z}_{ni}(\beta_0)\|^2 = \text{tr}(\Omega) + o(1) \text{ a.s.}$ Then,

$$\frac{1}{n} \sum_{i=1}^n \|\tilde{Z}_{ni}(\beta_0)\|^3 \leq \max_{1 \leq i \leq n} \|\tilde{Z}_{ni}(\beta_0)\| \frac{1}{n} \sum_{i=1}^n \|\tilde{Z}_{ni}(\beta_0)\|^2 = o(n^{1/2}) \text{ a.s.}$$

This completes the proof of lemma 4.

Proof of theorem 1

We use lemma 4 and follow the standard proof of theorem 1 in Owen (1990), then theorem 1 can be proved.

Proof of Theorem 2

Let $D_n = \{\beta : d^2(\beta) \leq \lambda_s(\hat{\Sigma})\}$, where $d^2(\beta) = \frac{1}{n} \sum_{i=1}^n (\tilde{Y}_i - \tilde{X}_i^\tau \beta)^2 / (1 + \|\beta\|^2)$. Then,

$$\begin{aligned} \beta \in D_n &\iff \frac{1}{n} \sum_{i=1}^n (\tilde{Y}_i - \tilde{X}_i^\tau \beta)^2 - (1 + \beta^\tau \beta) \lambda_s(\hat{\Sigma}) \leq 0 \\ &\iff \beta^\tau \left(\frac{1}{n} \sum_{i=1}^n \tilde{X}_i \tilde{X}_i^\tau - \lambda_s(\hat{\Sigma}) I_p \right) \beta - \left(\frac{2}{n} \sum_{i=1}^n \tilde{Y}_i \tilde{X}_i \right)^\tau \beta + \frac{1}{n} \sum_{i=1}^n \tilde{Y}_i^2 - \lambda_s(\hat{\Sigma}) \leq 0. \end{aligned} \tag{14}$$

Since $\frac{1}{n} \sum_{i=1}^n \tilde{X}_i \tilde{X}_i^\tau - \lambda_s(\hat{\Sigma}) I_p$ is a non-negative positive matrix, then the quadratic form of the left of ' \leq ' in (14) is a convex function of β , it follows that D_n is a convex set.

Similar to the proof of lemma 3 (1), we have $d^2(\beta_0) \rightarrow \sigma^2 \text{ a.s.}$ and $\hat{\Sigma} \rightarrow \Sigma_v + \sigma^2 I_p \text{ a.s.}$ Furthermore, $\lambda_s(\hat{\Sigma}) \rightarrow \lambda_s(\Sigma_v) + \sigma^2 \text{ a.s.}$ ($n \rightarrow \infty$). It means that

$$d^2(\beta_0) < \sigma^2 + \frac{1}{2} \lambda_s(\Sigma) < \lambda_s(\hat{\Sigma}), \text{ a.s.} \tag{15}$$

provided n is large enough. Then D_n is not empty since β_0 in it when n is large enough.

Note that $\tilde{x}_i, \tilde{u}_i, \tilde{\epsilon}_i$ are independent, $E\tilde{x}_i = \tilde{h}(t_i), E\tilde{u}_i = 0, E\tilde{\epsilon}_i = 0, \sup_{0 \leq t \leq 1} |\tilde{g}(t)| = O(a_n)$ and $\sup_{0 \leq t \leq 1} |\tilde{h}(t)| = O(a_n)$, we can easily drive from the formulae of $\tilde{Z}_{ni}(\beta)$ before (12) that

$$E \tilde{Z}_{ni}(\beta) = \text{Cov}(\tilde{x}_i)(\beta_0 - \beta) + \frac{(\beta_0 - \beta)^\tau \text{Cov}(\tilde{x}_i)(\beta_0 - \beta)}{1 + \|\beta\|^2} \beta + O(a_n^2). \tag{16}$$

By using $\tilde{x}_i = v_i + \tilde{h}(t_i) - \bar{v}_i$, we get

$$\text{Cov}(\tilde{x}_i) = \Sigma_v + \left(\sum_{j=1}^n w_{nj}^2(t_i) - 2w_{ni}(t_i) \right) \Sigma_v + O(a_n^2) = \Sigma_v + o(1).$$

Therefore, on one hand, if $\beta \neq \beta_0$ and $\lim_{n \rightarrow \infty} E \tilde{Z}_{ni}(\beta) = 0$, then we have

$$\Sigma_v(\beta_0 - \beta) + \frac{(\beta_0 - \beta)^\tau \Sigma_v(\beta_0 - \beta)}{1 + \|\beta\|^2} \beta = 0. \tag{17}$$

Multiply $(\beta_0 - \beta)^\tau$ at both sides of (17), we get

$$(\beta_0 - \beta)^\tau \Sigma_v(\beta_0 - \beta) + \frac{(\beta_0 - \beta)^\tau \beta (\beta_0 - \beta)^\tau \Sigma_v(\beta_0 - \beta)}{1 + \|\beta\|^2} = 0.$$

Since Σ_v is a PDM, we get $1 + \beta^\tau \beta_0 = 0$. Applying the same method of proving Theorem 3.2 in Cui & Chen (2003), we get

$$d^2(\beta) = \frac{(\beta_0 - \beta)^\tau \Sigma_v(\beta_0 - \beta)}{1 + \|\beta\|^2} + \sigma^2 + o_p(1), \quad \text{and} \quad \lambda_s(\Sigma_v) + \sigma^2 + o_p(1).$$

It is easy to get ' $o_p(1) > \lambda_s(\Sigma_v)[(\|\beta - \beta_0\|^2)/(1 + \|\beta\|^2) - 1] =: \delta > 0$ ' from ' $d^2(\beta) < \lambda_s(\hat{\Sigma})$ ' and $1 + \beta^\tau \beta_0 = 0$. Thus, $P\{\beta \in CR_x\} \leq P\{d^2(\beta) < \lambda_s(\hat{\Sigma})\} \leq P\{o_p(1) \geq \delta\} \rightarrow 0$ as $n \rightarrow \infty$.

On the other hand, if $\beta \neq \beta_0$ and $\lim_{n \rightarrow \infty} E \tilde{Z}_{ni}(\beta) \neq 0$, then $l(\beta) \rightarrow +\infty$ a.s. by the standard proof of empirical likelihood method. Thus,

$$P\{\beta \in CR_x\} \leq P\{l(\beta) \leq C_x\} \rightarrow 0 \quad \text{as} \quad n \rightarrow \infty.$$

The proof of theorem 2 is complete.

Proof of Theorem 3

We first prove

$$\frac{1}{n} \sum_{i=1}^n \tilde{Z}_{ni}(\beta^*) - \frac{1}{n} \sum_{i=1}^n \tilde{Z}_{ni}(\beta_0) = n^{-1/2} \Omega^{1/2} \gamma + o_p(n^{-1/2}). \tag{18}$$

Note that

$$\left. \frac{\partial \tilde{Z}_{ni}(\beta)}{\partial \beta} \right|_{\beta=\beta_0} = -\tilde{X}_i \tilde{X}_i^\tau - \frac{2(\tilde{Y}_i - \tilde{X}_i^\tau \beta_0) \beta_0 \tilde{X}_i^\tau}{1 + \|\beta_0\|^2} + \frac{(\tilde{Y}_i - \tilde{X}_i^\tau \beta_0)^2}{1 + \|\beta_0\|^2} \left(I_p - \frac{2\beta_0 \beta_0^\tau}{1 + \|\beta_0\|^2} \right),$$

then it follows from lemma 3 and the proof of lemma 4 that $\frac{1}{n} \sum_{i=1}^n \partial \tilde{Z}_{ni}(\beta) / \partial \beta|_{\beta=\beta_0} = -\Sigma_v + o(1)$ a.s. Employing Taylor expansion, we have that

$$\frac{1}{n} \sum_{i=1}^n \tilde{Z}_{ni}(\beta^*) - \frac{1}{n} \sum_{i=1}^n \tilde{Z}_{ni}(\beta_0) = \frac{1}{n} \sum_{i=1}^n \left. \frac{\partial \tilde{Z}_{ni}(\beta)}{\partial \beta} \right|_{\beta=\beta_0} (\beta^* - \beta_0) + o_p(n^{-1/2})$$

and (18) is true from $\beta^* - \beta_0 = -n^{-1/2} \Sigma_v^{-1} \Omega^{1/2} \gamma$. Using formulae (16), lemma 3 and the proof of lemma 4, we obtain

$$\max_{1 \leq i \leq n} \|\tilde{Z}_{ni}(\beta^*)\| = o(n^{1/2}) \text{ a.s.}, \quad \frac{1}{n} \sum_{i=1}^n \|\tilde{Z}_{ni}(\beta^*)\|^3 = o(n^{1/2}) \text{ a.s.}$$

Therefore from the standard proof of Owen (1990), we get

$$\begin{aligned} l(\beta^*) &= n \left[\frac{1}{n} \sum_{i=1}^n \tilde{Z}_{ni}(\beta^*) \right]^\tau \left[\frac{1}{n} \sum_{i=1}^n \tilde{Z}_{ni}(\beta^*) \tilde{Z}_{ni}^\tau(\beta^*) \right]^{-1} \left[\frac{1}{n} \sum_{i=1}^n \tilde{Z}_{ni}(\beta^*) \right] + o_p(1) \\ &= n \left[\frac{1}{n} \sum_{i=1}^n \tilde{Z}_{ni}(\beta_0) - n^{-1/2} \Omega^{1/2} \gamma \right]^\tau \left[\frac{1}{n} \sum_{i=1}^n \tilde{Z}_{ni}(\beta^*) \tilde{Z}_{ni}^\tau(\beta^*) \right]^{-1} \left[\frac{1}{n} \sum_{i=1}^n \tilde{Z}_{ni}(\beta_0) - n^{-1/2} \Omega^{1/2} \gamma \right] \\ &\quad + o_p(1). \end{aligned}$$

Since $\frac{1}{n} \sum_{i=1}^n \tilde{Z}_{ni}(\beta^*) \tilde{Z}_{ni}^\tau(\beta^*) - \Omega = o(1)$ a.s., $\frac{1}{n} \sum_{i=1}^n \tilde{Z}_{ni}(\beta_0) \tilde{Z}_{ni}^\tau(\beta_0) - \Omega = o(1)$ a.s. and

$$n \left[\frac{1}{n} \sum_{i=1}^n \tilde{Z}_{ni}(\beta_0) \right]^\tau \Omega^{-1} \left[\frac{1}{n} \sum_{i=1}^n \tilde{Z}_{ni}(\beta_0) \right] \xrightarrow{d.} \chi_p^2,$$

we have $l(\beta^*) \xrightarrow{d.} \chi_p^2(\|\gamma\|^2)$ as $n \rightarrow \infty$. This ends the proof of theorem 3.