Available at WWW.MATHEMATICSWEB.ORG
POWERED BY SCIENCE DIRECT•

Journal of Multivariate Analysis 84 (2003) 101–115

Journal of Multivariate Analysis

http://www.elsevier.com/locate/jmva

Empirical likelihood confidence region for parameter in the errors-in-variables models

Hengjian Cui[a,*] and Song Xi Chen[b]

[a] *Department of Mathematics, Beijing Normal University, Beijing 100875, China*
[b] *Department of Statistics and Applied Probability, National University of Singapore, 117543, Singapore*

Received 18 January 2001

Abstract

This paper proposes a constrained empirical likelihood confidence region for a parameter β_0 in the linear errors-in-variables model: $Y_i = x_i^\tau \beta_0 + \varepsilon_i, X_i = x_i + u_i, (1 \leqslant i \leqslant n)$, which is constructed by combining the score function corresponding to the squared orthogonal distance with a constrained region of β_0. It is shown that the coverage error of the confidence region is of order n^{-1}, and Bartlett corrections can reduce the coverage errors to n^{-2}. An empirical Bartlett correction is given for practical implementation. Simulations show that the proposed confidence region has satisfactory coverage not only for large samples, but also for small to medium samples.

AMS 1991 subject classifications: 62G05; 62F12

Keywords: Bartlett correction; Confidence region; Coverage error; Empirical likelihood; Errors-in-variables; Linear regression

1. Introduction

Let (X, Y) be a pair of random variables from the following linear errors-in-variables (EV) regression model:

$$Y = x^\tau \beta_0 + \varepsilon, \quad X = x + u, \tag{1}$$

where β_0 is a $p \times 1$ vector of unknown parameters, x and u are the $p \times 1$ unobservable covariates and measurement error vectors, respectively, Y is a scalar

*Corresponding author.
E-mail addresses: hjcui@bnu.edu.cn (H. Cui), stacsx@nus.edu.sg (S.X. Chen).

doi:10.1016/S0047-259X(02)00017-9

response and ε is the model error. It is assumed that x and $(\varepsilon, u^\tau)^\tau$ are independent. Let $\Sigma_x =: Cov(x)$ and $\Sigma_u =: Cov(u)$ be the covariance matrices of the covariates and the measurement error. In order to identify model (1), we assume Σ_x is a positive definite matrix (PDM) and $\Sigma_1 =: \Sigma_u/var(\varepsilon)$ is a known $p \times p$ PDM. Without losing generality (otherwise, transform X by $\Sigma_1^{-1/2}X$), we assume

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

which means ε and u have the same dispersion parameter $\sigma^2 > 0$. This is the standard framework taken by Cui and Li [5], Liang et al. [9, p. 1523], and He and Liang [8]. Another way to identify model (1) is to assume that Σ_u is a known $p \times p$ PDM [7].

The last two decades have seen an increasing number of applications of the linear EV model due to its simple form and wide applicability. Comprehensive reviews on the research and development of the EV model can be found in [1,7] and the references therein. An important problem in the EV regression model (1) is how to construct confidence regions for β_0 when the distributions of ε and u are unknown. In this nonparametric setting, the standard method is to construct confidence regions based on the asymptotic normal distribution of an estimator of β_0 by estimating the asymptotic covariance matrix of the estimator. While the covariance matrix is easily estimated in a standard linear model without measurement errors, this is not the case for the EV model as the covariance has a complicated form due to the error u in the observed covariate X. Any direct estimation may be subject to large error which may lead to larger coverage errors for the confidence region. Also, there is no guarantee that the estimated covariance is positive definite for a finite sample size. The bootstrap has been used to construct confidence regions for β_0. However, like all the multi-dimensional bootstrap confidence regions, subjective instructions on the shapes and orientations of the regions have to be given.

The empirical likelihood as an alternative to the bootstrap for constructing nonparametric confidence regions were introduced by Owen [11,12]. Instead of resampling with equal probability weights like the bootstrap, the empirical likelihood profiles a multinomial likelihood under a set of constraints which reflect the characteristics of the quantity of interests. Important features of the empirical likelihood are its automatic determination of the shape and orientation of a confidence region by the data, and its implicit studentizing carried out in its internal optimization without the need to estimate the covariance explicitly. Furthermore, it has been shown in many cases that the empirical likelihood confidence regions are Bartlett correctable, meaning that a simple mean adjustment reduces coverage error by one order of magnitude, see [6] for smoothed functions of means, Chen and Hall [4] for quantiles and others. The empirical likelihood has also been used for parameters defined by estimation equations in [10]. Owen [13] proposed empirical likelihood confidence regions for β_0 in the ordinary linear model without measurement error by deriving a nonparametric version of the Wilks' theorem. The coverage accuracy and Bartlett correction of the confidence regions are evaluated in [2,3].

The aims of the paper are to construct empirical likelihood confidence regions for β_0 in the EV model and to evaluate the coverage accuracy and Bartlett correctability of the confidence regions. In the linear EV model (1)–(2), the estimation of β_0 is obtained by solving a score equation that adds up squared orthogonal distances from each data point to a hyperplane in R^{p+1} (see (3) in Section 2). The score function has more than two solutions and obviously only one of them is genuine. This is a quite different situation from an ordinary linear model. The empirical likelihood has to be constrained in order to eliminate those unwanted solutions by restricting the parameter space in a region such that the score equation has an unique solution that converges to β_0.

The paper is organized as follows. We formulate the constrained empirical likelihood and confidence regions for β_0 in Section 2. The coverage accuracy and Bartlett correction are studied in Section 3. Empirical results from a simulation study is presented in Section 4. Section 5 provides extensions to other EV models which are parallel to the linear EV model (1)–(2). All the technical proofs are put in Section 6.

2. Empirical likelihood confidence regions

Suppose $\{(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)\}$ is an independent and identically distributed random sample from the model (1)–(2). An estimator of β_0 given by the generalized least square (GLS) method is

$$\hat{\beta}_n = \arg \min_{\beta \in R^{p+1}} \sum_{i=1}^{n} \frac{(Y_i - X_i^\tau \beta)^2}{1 + ||\beta||^2}, \tag{3}$$

where $(Y_i - X_i^\tau \beta)^2/(1 + ||\beta||^2)$ is the squared orthogonal norm from a point (X_i, Y_i) to the plane $L_\beta = \{z : z \in R^{p+1}, (\beta^\tau, -1)z = 0\}$. Here $||\cdot||$ denotes the Euclidean norm. The estimator $\hat{\beta}_n$ is strongly consistent and asymptotically normal [5]. Its asymptotic variance of $\sqrt{n}(\hat{\beta}_n - \beta_0)$ is $\Sigma = \Sigma_x^{-1}\Sigma_0\Sigma_x^{-1}$ where

$$\Sigma_0 = \sigma^2(1 + ||\beta_0||^2)\Sigma_x + Cov\left[(\varepsilon - u^\tau\beta_0)u + \frac{(\varepsilon - u^\tau\beta_0)^2}{(1 + ||\beta_0||^2)}\beta_0\right] \tag{4}$$

is $p \times p$ PDM provided the following moment condition is satisfied:

$$E[|\varepsilon|^4 + ||u||^4] < +\infty. \tag{5}$$

Eq. (3) implies that $\hat{\beta}_n$ is a root of the following score equation:

$$\frac{1}{n}\sum_{i=1}^{n} Z_i(\beta) = 0, \tag{6}$$

where

$$Z_i(\beta) = X_i(Y_i - X_i^\tau\beta) + \frac{(Y_i - X_i^\tau\beta)^2}{1 + ||\beta||^2}\beta \quad \text{and} \quad E[Z_i(\beta_0)] = 0.$$

Usually, we estimate Σ by the sandwich estimator $\hat{\Sigma}_x^{-1}\hat{\Sigma}_0\hat{\Sigma}_x^{-1}$ in which $\hat{\Sigma}_0 = \frac{1}{n}\sum_{i=1}^n Z_i(\hat{\beta}_n)Z_i^\tau(\hat{\beta}_n)$ where an estimator for Σ_x given in (13). But we could not guarantee the $\hat{\Sigma}_x$ is a PDM, and the finite sample estimation error of $\hat{\Sigma}$ can be substantial as shown in our simulation.

Let $p_1, \dots, p_n$ be non-negative numbers adding to unity. The log empirical likelihood ratio evaluated at β, a candidate value of β_0, is

$$\ell(\beta) = -2 \min_{\sum p_i Z_i(\beta)=0} \sum_{i=1}^n \log(np_i). \tag{7}$$

By introducing a Lagrange multiplier $\lambda \in R^{p+1}$, standard derivations in the empirical likelihood lead to

$$\ell(\beta) = 2\sum_{i=1}^n \log\{1 + \lambda^\tau Z_i(\beta)\},$$

where λ satisfies

$$\sum_{i=1}^n \frac{Z_i(\beta)}{1 + \lambda^\tau Z_i(\beta)} = 0.$$

We have the following nonparametric version of Wilks' theorem.

Theorem 2.1. *Under the moment condition* (5), *as* $n \to \infty$ $\ell(\beta_0) \xrightarrow{\mathrm{d}} \chi_p^2$.

In a standard situation, an empirical likelihood confidence region with nominal level of α would be

$$CR_{\alpha,0} = \{\beta_0 | \ell(\beta_0) < c_\alpha\} \tag{8}$$

by contouring $\ell(\beta)$ where c_α satisfies $P\{\chi_p^2 < c_\alpha\} = \alpha$. However, for the EV model the above confidence region is not appropriate. This is because, as mentioned in the introduction, the equation $E[Z_i(\beta)] = 0$ has at least two solutions when $\beta_0 \neq 0$ (see the lemma in Section 6). As the empirical likelihood ratio $\ell(\beta)$ achieves its minimum value 0 when all $p_i = 1/n$, $\ell(\beta)$ will be zero at all the roots of the score equation. So, the empirical likelihood surface is multi-modal. As a result, the confidence regions given in (8) will be disconnected and inconsistent.

To overcome this problem, we restrict β in a sub-parameter space

$$\Omega = \left\{\beta : E\left[\frac{(Y - X^\tau\beta)^2}{1 + ||\beta||^2}\right] < t_1[E(XX^\tau)]\right\},$$

where $t_1(B)$ is the minimum eigenvalue of a matrix B. Note that

$$\Omega = \{\beta : (\beta - \beta_0)^\tau \Sigma_x(\beta - \beta_0)/(1 + ||\beta||^2) < t_1(\Sigma_x)\},$$

and Ω is a open convex space and $\beta_0 \in \Omega$. If $E(Z_i(\beta)) = 0$ and $\beta \neq \beta_0$, then β should satisfy $\beta_0^\tau \beta = -1$ (see the proof of the lemma in Section 6). Hence,

$$\frac{(\beta-\beta_0)^\tau \Sigma_x (\beta-\beta_0)}{(1+||\beta||^2)} \geqslant \frac{t_1(\Sigma_x)||\beta-\beta_0||^2}{(1+||\beta||^2)} = t_1(\Sigma_x)\left[1+\frac{(1+||\beta_0||^2)}{(1+||\beta||^2)}\right] \geqslant t_1(\Sigma_x).$$

Thus, $\beta \bar{\in} \Omega$. This means that $E[Z_i(\beta)] = 0$ with $\beta \in \Omega$ is necessary and sufficient for $\beta = \beta_0$.

Define

$$\Omega_n = \left\{\beta : \frac{1}{n}\sum_{i=1}^n \frac{(Y_i - X_i^\tau \beta)^2}{1+||\beta||^2} < t_1\left[\frac{1}{n}\sum_{i=1}^n X_i X_i^\tau\right]\right\}$$

as an estimator of Ω. The lemma in Section 6 shows that Ω_n is also open convex and $P\{\beta_0 \in \Omega_n\} \to 1$, as $n \to \infty$. Therefore, constraining $\ell(\beta)$ for $\beta \in \Omega_n$ will eliminate the multiple solution problems. A proper confidence region for β_0 is then

$$CR_{\alpha,\mathrm{el}} = \{\beta | \beta \in \Omega_n, \ell(\beta) < c_\alpha\} \tag{9}$$

by restricting the naive confidence region $CR_{\alpha,0}$ in Ω_n. Theorem 2.1 shows that $CR_{\alpha,\mathrm{el}}$ has correct asymptotic coverage level.

The empirical likelihood confidence region is not convex. However, it is "asymptotically convex" which means that there is a convex region such that the gap between $CR_{\alpha,\mathrm{el}}$ and the convex region attract zero probability as $n \to \infty$. To appreciate this point, let $lsel(\beta) = n(\beta-\beta_0)^\tau \Sigma_0^{-1}(\beta-\beta_0)$ where $\Sigma_0 = Cov\{Z_i(\beta_0)\}$. The expansion given in the proof of Theorem 2.1 shows that $\ell(\beta) = lsel(\beta) + \Delta(\beta)$ where $\Delta(\beta) = O_p(n^{-1/2})$ for $\beta = \beta_0 + O_p(n^{-1/2})$. Define $CR_{\alpha+n^{-1/2+\eta}} = \{\beta|\beta \in \Omega_n, lsel(\beta) < c_{\alpha+n^{-1/2+\eta}}\}$ where $\eta \in (0, 1/2)$. Clearly, $CR_{\alpha+n^{-1/2+\eta}}$ is convex for any n. Let $A = \{CR_{\alpha,\mathrm{el}} - CR_{\alpha+n^{-1/2+\eta}}\} \cup \{CR_{\alpha+n^{-1/2+\eta}} - CR_{\alpha,\mathrm{el}}\}$ be the "gap" between $CR_{\alpha,\mathrm{el}}$ and $CR_{\alpha+n^{-1/2+\eta}}$. Clearly, $\beta \in A$ if and only if $|\Delta(\beta)| > |c_\alpha - c_{\alpha+n^{-1/2+\eta}}| \sim c_0 n^{-1/2+\eta}$ for a positive constant c_0 which is related to the derivative of the χ_p^2 quantile function at c_α. As $\Delta(\beta) = O_p(n^{-1/2})$,

$$P(\beta \in A) = P\{|\Delta(\beta)| > c_0 n^{-1/2+\eta}\} \to 0 \quad \text{as } n \to \infty.$$

3. Coverage accuracy and Bartlett correction

Let $W_i = \Sigma_0^{-1/2} Z_i(\beta_0)$, we have $E(W_i) = 0$ and $Cov(W_i) = I_p$.

Theorem 3.1. *Assume that* $E[||x||^{15} + (||u||^2 + \varepsilon^2)^{15}] < +\infty$ *and the characteristic function of* $Z_1(\beta_0)$ *satisfies the Cramér's condition* $\sup_{||t||>b} |g(t)| < 1$ *for every positive b. Then*

$$P\{\beta_0 \in CR_{\alpha,\mathrm{el}}\} = \alpha - a c_\alpha \psi_p(c_\alpha) n^{-1} + O(n^{-3/2}),$$

where $\psi_p(\cdot)$ *is the density of* χ_p^2 *distribution and*

$$a = p^{-1}[\tfrac{1}{2}E(W_1^\tau W_1)^2 - \tfrac{1}{3}E(W_1^\tau W_2)^3]. \tag{10}$$

Theorem 3.1 reveals that the coverage error of $CR_{\alpha,\mathrm{el}}$ is of order $O(n^{-1})$. The global and local type II errors of $CR_{\alpha,\mathrm{el}}$ is considered in the following Theorem 3.2.

Theorem 3.2. *Under the conditions of Theorem* 3.1, *we have*

$$\lim_{n\to\infty} P\{\tilde{\beta}\in CR_{\alpha,\mathrm{el}}\} = 0 \quad \textit{for any fixed } \tilde{\beta}\neq\beta_0,$$

$$\lim_{n\to\infty} P\{\tilde{\beta}_n\in CR_{\alpha,\mathrm{el}}\} = P\{\chi_p^2(||\gamma||^2)<c_\alpha\} \quad \textit{for } \tilde{\beta}_n = \beta_0 + \frac{1}{\sqrt{n}}\Sigma_x^{-1}\Sigma_0^{1/2}\gamma,$$

where $\chi_p^2(||\gamma||^2)$ *stands for the noncentral* χ^2 *random variable with p degree of freedoms and noncentral parameter* $||\gamma||^2$ *for a fixed* $\gamma\in R^p$.

From expansion (15) in Section 6, we may obtain the following expansion for $E[l(\beta_0)]$:

$$E[l(\beta_0)] = p(1+an^{-1}) + O(n^{-2}),$$

where a is given by (10). It is known that part of the coverage error of $CR_{\alpha,\mathrm{el}}$ is due to the fact that the mean of $l(\beta_0)$ is not p, the mean of χ_p^2. Bartlett correction is a procedure that reduces the coverage error by re-adjusting the mean of $\ell(\beta)$. We demonstrate in the following theorem that $CR_{\alpha,\mathrm{el}}$ is Bartlett correctable.

Theorem 3.3. *Assume that the conditions of Theorem* 3.1 *hold. Then*

$$P\{l(\beta_0)<c_\alpha(1+\zeta n^{-1})\} = \alpha + O(n^{-2}),$$

where $P\{\chi_p^2<c_\alpha\}=\alpha$ *and* ζ *is either a or an* $n^{1/2}$*-consistent estimate of a.*

The theorem means that a simple mean adjustment to $\ell(\beta)$ reduces the size of coverage error from the order n^{-1} to the order n^{-2}. For practical implementation, we give an $n^{1/2}$-consistent estimate of a. Let $\hat{\beta}_n$ be an $n^{1/2}$-consistent estimator of β_0, for instance that given in (3) or the orthogonal L_1 estimator given by He and Liang [8]. Let $\hat{Z}_i = Z_i(\hat{\beta}_n)$ and $\hat{\Sigma}_0 = \frac{1}{n}\sum_{i=1}^n \hat{Z}_i\hat{Z}_i^\tau$ be an estimator of Σ_0. Then, an estimator of a is

$$\hat{a}_n = p^{-1}\left(\frac{1}{2}n^{-1}\sum_{i=1}^n [\hat{Z}_i^\tau\hat{\Sigma}_0^{-1}\hat{Z}_i]^2 - \frac{1}{3}[n(n-1)]^{-1}\sum_{i\neq j}[\hat{Z}_i^\tau\hat{\Sigma}_0^{-1}\hat{Z}_j]^3\right). \tag{11}$$

A Bartlett corrected confidence region for β_0:

$$CR_{\alpha,\mathrm{bcel}} = \{\beta | \beta \in \Omega_n, \ell(\beta) < c_\alpha(1 + \hat{a}n^{-1})\}. \tag{12}$$

4. Simulation study

We report results from a simulation study designed to evaluate the performance of the proposed empirical likelihood confidence region and compare it with the confidence region based on the asymptotic normal distribution of the generalized linear square estimator $\hat{\beta}_n$ given in (3). Our simulation shows that the empirical likelihood confidence region has better coverage and shorter length than that based on the normal approximation.

Cui and Li [5] proved that

$$\sqrt{n}\hat{\Sigma}_0^{-1/2}\hat{\Sigma}_x(\hat{\beta}_n - \beta_0) \xrightarrow{\mathrm{d}} N(0, I_p) \quad \text{as } n \to \infty,$$

where

$$\hat{\Sigma}_x = \frac{1}{n}\sum_{i=1}^n X_i X_i^\tau - \hat{\sigma}_n^2 I_p, \quad \hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n \frac{(Y_i - X_i^\tau \hat{\beta}_n)^2}{1 + ||\hat{\beta}_n||^2}. \tag{13}$$

Therefore, we can formulate the following normal approximation based confidence region with nominal confidence level α:

$$CR_{\alpha,\mathrm{nor}} = \{\beta : n(\hat{\beta}_n - \beta)^\tau \hat{\Sigma}(\hat{\beta}_n - \beta) \leqslant c_\alpha\},$$

where $\hat{\Sigma} = \hat{\Sigma}_x \hat{\Sigma}_0^{-1} \hat{\Sigma}_x$.

We start with the following framework of simulation designed to compare the coverage of $CR_{\alpha,\mathrm{el}}$ with that of $CR_{\alpha,\mathrm{nor}}$. We choose $p = 2$ and two sets of distributions for x, u and ε: (i) $x \sim N(0, I_2)$, $u \sim N(0, \sigma^2 I_2)$ and $\varepsilon \sim N(0, \sigma^2)$; and (ii) x and u are the uniform distributions on D_1 and D_2, respectively, where $D_1 = [-\sqrt{3}, \sqrt{3}] \times [-\sqrt{3}, \sqrt{3}]$ and $D_2 = [-\sqrt{3}\sigma, \sqrt{3}\sigma] \times [-\sqrt{3}\sigma, \sqrt{3}\sigma]$, and $\varepsilon \sim N(0, \sigma^2)$. The true value of β_0 is fixed as $(1, 0)^\tau$, the sample size $n = 20, 30, 50$ and 100, the standard dispersion parameter σ are 0.9 and 0.3, and nominal coverage level α are 0.90 and 0.95, respectively. For each simulation, the empirical coverage is evaluated by contouring the proportion of $\ell(\beta_0) \leqslant c_\alpha$ for the empirical likelihood regions and similarly done for the asymptotic regions.

Tables 1 and 2 contains the empirical coverage of the two types confidence regions based on 1000 simulation. It is fairly clear that the empirical likelihood confidence region has much better coverage than its normal approximation counterpart consistently over the entire range of sample size considered and both sets of distributions for x, u and ε. The improvement in coverage by using the empirical likelihood is substantial for n ranging from 20 to 50.

We then proceed comparing the size of the two types of confidence regions, which in turns requires determination of the boundary of the confidence regions. The boundary can be located by computing $\ell(\beta)$ over a densely spaced lattice around $\hat{\beta}_n$.

Table 1
The coverage probability comparisons of confidence regions when $x \sim N(0, I_2)$, $u \sim N(0, \sigma^2 I_2)$ and $\varepsilon \sim N(0, \sigma^2)$.

	$\sigma = 0.3$				$\sigma = 0.9$			
	$\alpha = 0.90$		$\alpha = 0.95$		$\alpha = 0.90$		$\alpha = 0.95$	
n	GLS	ELR	GLS	ELR	GLS	ELR	GLS	ELR
20	0.764	0.915	0.825	0.970	0.762	0.831	0.828	0.869
30	0.807	0.908	0.871	0.961	0.815	0.872	0.879	0.920
50	0.856	0.905	0.913	0.958	0.828	0.894	0.915	0.948
100	0.865	0.898	0.952	0.951	0.870	0.902	0.913	0.953

Table 2
The coverage probability comparisons of confidence regions when $x \sim U(D_1)$, $u \sim U(D_2)$ and $\varepsilon \sim N(0, \sigma^2)$.

	$\sigma = 0.3$				$\sigma = 0.9$			
	$\alpha = 0.90$		$\alpha = 0.95$		$\alpha = 0.90$		$\alpha = 0.95$	
n	GLS	ELR	GLS	ELR	GLS	ELR	GLS	ELR
20	0.775	0.913	0.833	0.963	0.778	0.839	0.823	0.878
30	0.831	0.909	0.873	0.958	0.820	0.881	0.866	0.941
50	0.854	0.906	0.901	0.948	0.847	0.897	0.906	0.957
100	0.878	0.902	0.930	0.953	0.882	0.905	0.915	0.948

At a given point on the lattice, if $\ell(\beta) > (\leqslant) c_\alpha$ the β is outside (inside) $C_{\mathrm{el},\alpha}$, and the boundary corresponds to $\ell(\beta) = c_\alpha$. For simplicity of computation, we consider the case $p = 1$. The size of confidence region reduces to the length of confidence intervals. We set the distributions of x, u and ε as $x \sim N(0, 1)$, $u \sim N(0, \sigma^2)$ and $\varepsilon \sim N(0, \sigma^2)$, respectively. The true value $\beta_0 = 1$, standard dispersion parameter $\sigma = 0.3$, and nominal coverage level α are 0.90 and 0.95, respectively. The sample sizes are $n = 20, 30, 50, 100$. We draw a random sample with sample size n, compute $\ell(\beta)$, $\hat{\beta}_n$ and Ω_n. Then construct the confidence intervals $CR_{\alpha,\mathrm{el}}$ and $CR_{\alpha,\mathrm{nor}}$, and get their lengths (denoted as L_{el} and L_{nor}, respectively). The above procedure is repeated 100 times. We then obtain the average and standard deviation of their length difference $LD = L_{\mathrm{nor}} - L_{\mathrm{el}}$. The $\ell(\beta)$ and range of Ω_n are reported in Fig. 1 for $n = 100$. It shows that $l(\beta)$ has quadratic form near two local minimums, Ω_n is an interval which covers the true value $\beta_0 = 1$. Obviously, the $CR_{\alpha,\mathrm{el}}$ is a connected interval. The results are similar for the other values of n. Table 3 summarizes the length comparisons for various sample size and nominal coverage level α. It reveals that $CR_{\alpha,\mathrm{nor}}$ is much longer than $CR_{\alpha,\mathrm{el}}$ for all the cases considered. It is quite remarkable that the length of the empirical likelihood confidence interval is shorter than that of the normal approximation confidence interval over 90% of the times when $n \leqslant 30$. Although, the percentage decreases to 70% when n increases from 30 to 100, it is still remarkable by

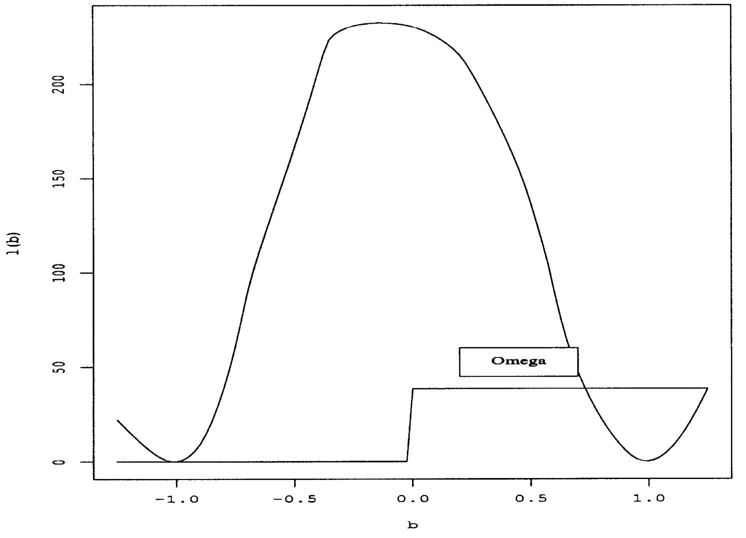

Fig. 1. Log empirical likelihood ratio $l(\beta)$ versus β and range of Ω_n.

Table 3
The LD comparisons of confidence intervals with $\alpha = 0.95$.

Sample size	α	Ave. of LD	Std. of LD	Percentage of LD>0
20	0.90	0.104	0.091	99
	0.95	0.159	0.114	98
30	0.90	0.072	0.077	92
	0.95	0.118	0.127	94
50	0.90	0.032	0.041	84
	0.95	0.043	0.045	85
100	0.90	0.007	0.014	74
	0.95	0.008	0.015	72

all measures. We also observe that the average difference of length increases as α is larger which is expected.

5. Extensions

In this section we consider possible extension of the empirical likelihood to other EV models. For a linear EV model: $Y_i = x_i^\tau \beta_0 + \varepsilon_i,\ X_i = x_i + u_i\ (1 \leqslant i \leqslant n),\ E[(\varepsilon_1, u_1^\tau)^\tau] = 0,\ Cov(u_1) = \Sigma_u > 0$ known, which is the key model in [7] and has a slight different variance assumption from ours. It is easier to construct confidence regions and there is no need to restrict the parameter space. The empirical likelihood $\ell(\beta)$ can be established based on a score function corresponding to the adjusted square norm $(Y_i - X_i^\tau\beta)^2 - \beta^\tau\Sigma_u\beta$ (except a constant factor):

$$Z_i(\beta) =: X_i(Y_i - X_i^\tau\beta) + \Sigma_u\beta.$$

Since $EZ_i(\beta) = 0 \Leftrightarrow \beta = \beta_0$, there is no need to place restriction on β. Theorems 3.1 and 3.3 of this paper remain true under some regular conditions.

For a nonlinear EV model: $Y_i = f(x_i, \beta_0) + \varepsilon_i$, $X_i = x_i + u_i$ $(1 \leqslant i \leqslant n)$, $\Sigma_x = Cov(x)$ is a PDM. Assume that the distribution of (x, u) is known and $E[(\varepsilon, u^\tau)^\tau] = 0$. In order to identify the nonlinear EV regression model, we need the additive assumption

$$E\left\{\frac{\partial E[f(x,\beta)|X]}{\partial\beta}E([f(x,\beta) - f(x,\beta_0)]|X)\right\} = 0 \quad \text{if and only if } \beta = \beta_0,$$

where $E[f(x,\beta)|X]$ is the conditional expectation of $f(x,\beta)$ given X. Let

$$Z_i(\beta) = \frac{\partial E[f(x,\beta)|X_i]}{\partial\beta}(Y_i - E[f(x,\beta)|X_i])$$

be the score function corresponding to the squared distance $(Y_i - E[f(x,\beta)|X_i])^2$. We have $E[Z_i(\beta_0)] = 0$ and $EZ_i(\beta) = 0$ is necessary and sufficient for $\beta = \beta_0$. The empirical likelihood can be established accordingly.

6. Proofs

Lemma. (i) *The equation $EZ_i(\beta) = 0$ has at least two solutions when $\beta_0 \neq 0$.*

(ii) *Ω (or Ω_n) is nonempty, open and convex set (a.s.).*

Proof. Note that

$$E[Z_i(\beta)] = \Sigma_x(\beta_0 - \beta) + \frac{(\beta_0 - \beta)^\tau\Sigma_x(\beta_0 - \beta)}{1 + ||\beta||^2}\beta. \tag{14}$$

It is obvious that $\beta = \beta_0$ is one solution. Suppose $\beta \neq \beta_0$ $(\beta_0 \neq 0)$, we only prove, without loss of generality, that (14) has at least one solutions for $\Sigma_x = diag(\sigma_1^2, \ldots, \sigma_p^2)$ (If Σ_x is not a diagonal matrix, we can make eigenvalue–eigenvector decomposition $\Sigma_x = U^\tau \Lambda U$, where Λ is a diagonal matrix, then corresponding β is $U\beta$). From (14) and $E[Z_i(\beta)] = 0$, it is easy to get $\beta_0^\tau\beta = -1$ by multiplying $(\beta - \beta_0)^\tau$ on the two sides of the equation $E[Z_i(\beta)] = 0$. Let $c > 0$ satisfy

$$\sum_{j \in J} \frac{\sigma_j^2}{c - \sigma_j^2}\beta_{0j}^2 = 1$$

and $\beta = (\Sigma_x - cI_p)^+\Sigma_x\beta_0$, where β_{0j} is the jth component of β_0, $J = \{j : \beta_{0j} \neq 0\}$ and "+" stands for the Moore–Penrose inverse.

We shall show that this $\beta \neq \beta_0$ and is a solution of the equation $E[Z_i(\beta)] = 0$. Since $\beta_0 \neq 0$, then $J \neq \emptyset$ and $c \neq \sigma_j^2$ $(j \in J)$ from the definition of c given above, thus $\beta \neq \beta_0$. Note that $\beta_0 - \beta = (I - (\Sigma_x - cI)^+\Sigma_x)\beta_0$, then by some calculations, we obtain that

$$(\beta_0-\beta)^\tau\Sigma_x(\beta_0-\beta)=\beta_0^\tau(I-(\Sigma_x-cI)^+\Sigma_x)\Sigma_x(I-(\Sigma_x-cI)^+\Sigma_x)\beta_0$$
$$=\sum_{j\in J}\sigma_j^2\left(1-\frac{\sigma_j^2}{\sigma_j^2-c}\right)^2\beta_{0j}^2=c^2\sum_{j\in J}\frac{\sigma_j^2}{(\sigma_j^2-c)^2}\beta_{0j}^2,$$

$$\beta^\tau(\beta-\beta_0)=\beta_0^\tau\Sigma_x(\Sigma_x-cI)^+((\Sigma_x-cI)^+\Sigma_x-I)\beta_0$$
$$=\sum_{j\in J}\frac{\sigma_j^2}{\sigma_j^2-c}\left(\frac{\sigma_j^2}{\sigma_j^2-c}-1\right)\beta_{0j}^2=c\sum_{j\in J}\frac{\sigma_j^2}{(\sigma_j^2-c)^2}\beta_{0j}^2.$$

Therefore,

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

Hence the β is a solution of $E(Z_i(\beta))=0$ from (14). This completes the proof of (i).

We now turn to the proof of (ii). By the definition of Ω,

$$\Omega=\{\beta:\beta^\tau[\Sigma_x-t_1(\Sigma_x)I_p]\beta-2\beta_0^\tau\Sigma_x(\beta-\beta_0)-\beta_0^\tau\Sigma_x\beta_0-t_1(\Sigma_x)<0\}.$$

Note that $\beta_0\in\Omega$, $\Sigma_x-t_1(\Sigma_x)I_p$ is a $p\times p$ nonnegative matrix, and the restriction in the Ω is a quadratic form, so we claim that Ω is nonempty open and convex set. Employing similar argument to the Ω_n, we complete the proof of (ii). □

Proof of Theorem 2.1. Since the analytic solution for λ is not obtainable, we have to resort to asymptotic expansions. Using an expansion given in [2], under some moment conditions, we have the following Taylor expansion for $\ell(\beta_0)$,

$$n^{-1}l(\beta_0)=A^jA^j-A^{jk}A^jA^k+(\tfrac{2}{3}\bar{\alpha}^{jkl}+\tfrac{2}{3}A^{jkl}-2\bar{\alpha}^{jkm}A^{lm})A^jA^kA^l$$
$$+(\bar{\alpha}^{jkq}\bar{\alpha}^{lmq}-\tfrac{1}{2}\bar{\alpha}^{jklm})A^jA^kA^lA^m+A^{jl}A^{kl}A^jA^k+O_p(n^{-5/2}),\tag{15}$$

where $\bar{\alpha}^{j_1\cdots j_k}=n^{-1}\Sigma E(W_{ij_1}\cdots W_{ij_k})$ and $A^{j_1\cdots j_k}=n^{-1}\Sigma(W_{ij_1}\cdots W_{ij_k}-\bar{\alpha}^{j_1\cdots j_k})$, W_i is defined in the beginning of Section 3. Here we use the summation convention according to which, if an index occurs more that once in an expression, summation over the index is understood.

It follows from condition (5) and the central limit theorem that $\sqrt{n}\bar{W}\xrightarrow{d}N(0,I_p)$, where $\bar{W}=\frac{1}{n}\sum_{i=1}^n W_i$. From expansion (15), we have

$$\ell(\beta_0)=nA^jA^j+o_p(1)=(\sqrt{n}\bar{W})^\tau(\sqrt{n}\bar{W})+o_p(1)\xrightarrow{d}\chi_p^2.$$

This completes the proof of Theorem 2.1. □

Before giving a proof to Theorem 3.1, we need some preparation as follows. We decompose $\ell(\beta_0)$ from (15) as

$$\ell(\beta_0)=(\sqrt{n}R^\tau)(\sqrt{n}R)+O_p(n^{-5/2}),\tag{16}$$

where $R = R_1 + R_2 + R_3$ is p-dimensional vector and $R_j = O_p(n^{-j/2})$ for $j = 1, 2, 3$. Comparing terms in (15) with those in (16) yields,

$$R_1^j = A^j, \quad R_2^j = -\tfrac{1}{2} A^{jk} A^k + \tfrac{1}{3} \bar{\alpha}^{jkm} A^k A^m,$$

$$\begin{aligned} R_3^j = {} & \tfrac{3}{8} A^{jm} A^{km} A^k + \tfrac{1}{3} A^{jkm} A^k A^l - \tfrac{5}{12} \bar{\alpha}^{jkm} A^{lm} A^k A^l \\ & - \tfrac{5}{12} \bar{\alpha}^{klm} A^{jm} A^k A^l + \tfrac{4}{9} \bar{\alpha}^{jkq} \bar{\alpha}^{lmq} A^m A^k A^l - \tfrac{1}{4} \bar{\alpha}^{jklm} A^m A^k A^l, \end{aligned}$$

where R_l^j is the jth component of R_l.

Proof of Theorem 3.1. Chen [2] handled auxiliary random vectors $Z_{i1}(\beta_0)$, $Z_{i2}(\beta_0), \ldots, Z_{in}(\beta_0)$ $(1 \leqslant i \leqslant n)$ for linear model which are independent but not necessarily identically distributed random variables. Under conditions (i)–(vi) in his Theorem 2.1, Chen established an Edgeworth expansion for $P\{l(\beta_0 < c_\alpha)\}$ with a remainder term of $O(n^{-3/2})$. Since our $Z_1(\beta_0), Z_2(\beta_0), \ldots, Z_n(\beta_0)$ are independent and identically distributed, we need only to check conditions (i)–(vi) in that Theorem 2.1 of Chen [2]. Conditions (i), (ii) and (vi) are satisfied automatically; the moment condition ensues that conditions (iii) and (iv) hold as well; condition (v) is just the Cramér condition assumed by us as well. Applying Theorem 2.1 of Chen [2] leads to the following Edgeworth expansion:

$$P\{l(\beta_0) < c_\alpha\} = \alpha - a c_\alpha \psi_p(c_\alpha) n^{-1} + O(n^{-3/2}), \tag{17}$$

where $a = p^{-1}[\frac{1}{2}\bar{\alpha}^{jjmm} - \frac{1}{3}\bar{\alpha}^{jkm}\bar{\alpha}^{jkm}] = p^{-1}[\frac{1}{2}E(W_1^\tau W_1)^2 - \frac{1}{3}E(W_1^\tau W_2)^3]$.

If $\beta_0 \bar{\in} \Omega_n$ then

$$\frac{1}{n} \sum_{i=1}^{n} \frac{(Y_i - X_i^\tau \beta_0)^2}{1 + ||\beta_0||^2} \geqslant t_1 \left[\frac{1}{n} \sum_{i=1}^{n} X_i X_i^\tau \right]. \tag{18}$$

Let $A_{1n} = \frac{1}{n}\sum_{i=1}^n X_i X_i^\tau$, then $A_1 = E(A_{1n}) = \Sigma_x + \sigma^2 I_p$; and let

$$A_{2n} = \begin{pmatrix} \frac{1}{n} \sum_{i=1}^{n} Y_i^2 & \frac{1}{n} \sum_{i=1}^{n} X_i^\tau Y_i \\ \frac{1}{n} \sum_{i=1}^{n} X_i Y_i & \frac{1}{n} \sum_{i=1}^{n} X_i X_i^\tau \end{pmatrix},$$

$$A_2 = E(A_{2n}) = \begin{pmatrix} \beta_0^\tau \Sigma_x \beta_0 + \sigma^2 & \beta_0^\tau \Sigma_x \\ \Sigma_x \beta_0 & \Sigma_x + \sigma^2 I_p \end{pmatrix},$$

Now (18) implies

$$\frac{(1, -\beta_0^\tau)[A_{2n} - A_2](1, -\beta_0^\tau)^\tau}{1 + ||\beta_0||^2} - [t_1(A_{1n}) - t_1(A_1)] \geqslant - \frac{(1, -\beta_0^\tau) A_2 (1, -\beta_0^\tau)^\tau}{1 + ||\beta_0||^2} + t_1(A_1).$$

Therefore,

$$||A_{2n} - A_2|| + ||A_{1n} - A_1|| \geqslant t_1(\Sigma_x).$$

By the Chebyshev's inequality

$$
\begin{aligned}
P\{\beta_0 \bar{\in} \Omega_n\} &\leqslant P\{||A_{2n} - A_2|| + ||A_{1n} - A_1|| \geqslant t_1(\Sigma_x)\} \\
&\leqslant \frac{E(||A_{2n} - A_2|| + ||A_{1n} - A_1||)^4}{t_1(\Sigma_x)^4} \\
&\leqslant C_2 E(||A_{2n} - A_2||^4 + ||A_{1n} - A_1||^4) = O(n^{-2}), \qquad (19)
\end{aligned}
$$

where C_2 is a positive constant independent of n. Combine (17) and (19),

$$
\begin{aligned}
P\{\beta_0 \in CR_{\alpha,\mathrm{el}}\} &= P\{\beta_0 \in \Omega_n, \ell(\beta_0) < c_\alpha\} \\
&= P\{\ell(\beta_0) < c_\alpha\} - P\{\beta_0 \bar{\in} \Omega_n, \ell(\beta_0) < c_\alpha\} \\
&= P\{l(\beta_0) < c_\alpha\} + O(n^{-2}) \\
&= \alpha - a c_\alpha \psi_p(c_\alpha) n^{-1} + O(n^{-3/2}). \qquad (20)
\end{aligned}
$$

This completes the proof of Theorem 3.1. □

Proof of Theorem 3.2. It is obvious that from (14) and $E[Z_i(\beta)] = 0$ that $\beta_0^\tau \beta = -1$. It can be shown that $||\beta - \beta_0||^2/(1 + ||\beta||^2) = 1 + (1 + ||\beta_0||^2)/(1 + ||\beta||^2) > 1$. Thus, if $||\tilde{\beta} - \beta_0||^2/(1 + ||\tilde{\beta}||^2) \leqslant 1$, $\tilde{\beta}$ is not any solution of $E[Z_i(\beta)] = 0$. From the standard empirical likelihood theory, we have $\ell(\tilde{\beta}) \to \infty$ a.s. as $n \to \infty$. Therefore, when $||\tilde{\beta} - \beta_0||^2/(1 + ||\tilde{\beta}||^2) > 1$, we get

$$
\begin{aligned}
P\{\tilde{\beta} \in CR_{\alpha,\mathrm{el}}\} &\leqslant P\{\tilde{\beta} \in \Omega_n\} \\
&\leqslant P\left\{ ||A_{2n} - A_2|| + ||A_{1n} - A_1|| \geqslant t_1(\Sigma_x) \left(\frac{||\tilde{\beta} - \beta_0||^2}{1 + ||\tilde{\beta}||^2} - 1 \right) \right\} \to 0. \qquad (21)
\end{aligned}
$$

When $||\tilde{\beta} - \beta_0||^2/(1 + ||\tilde{\beta}||^2) \leqslant 1$ and $\tilde{\beta} \neq \beta_0$, we get

$$
P\{\tilde{\beta} \in CR_{\alpha,\mathrm{el}}\} \leqslant P\{l(\tilde{\beta}) < c_\alpha\} \to 0. \qquad (22)
$$

So, the first part of the theorem is proved from (21) and (22).

For $\tilde{\beta}_n = \beta_0 + \frac{1}{\sqrt{n}} \Sigma_x^{-1} \Sigma_0^{1/2} \gamma$, it is obvious that

$$
P\{\tilde{\beta}_n \in \Omega_n\} \to 1 \quad \text{as } n \to \infty \qquad (23)
$$

and

$$
\begin{aligned}
\bar{Z}(\tilde{\beta}_n) &= \bar{Z}(\beta_0) - \Sigma_x \frac{1}{\sqrt{n}} \Sigma_x^{-1} \Sigma_0^{1/2} \gamma + o_p\left(\frac{1}{\sqrt{n}}\right) \\
&= \bar{Z}(\beta_0) - \frac{1}{\sqrt{n}} \Sigma_0^{1/2} \gamma + o_p\left(\frac{1}{\sqrt{n}}\right),
\end{aligned}
$$

where $\bar{Z}(\beta) = \frac{1}{n}\sum_{i=1}^{n} Z_i(\beta)$. Then, we have

$$\begin{aligned} l(\tilde{\beta}_n) &= n\bar{Z}(\tilde{\beta}_n)^\tau \Sigma_0^{-1}\bar{Z}(\tilde{\beta}_n) + o_p(1) \\ &= n\left[\bar{Z}(\beta_0) - \frac{1}{\sqrt{n}}\Sigma_0^{1/2}\gamma\right]^\tau \Sigma_0^{-1}\left[\bar{Z}(\beta_0) - \frac{1}{\sqrt{n}}\Sigma_0^{1/2}\gamma\right] + o_p(1) \\ &\xrightarrow{\mathrm{d}} \chi_p^2(||\gamma||^2). \end{aligned} \tag{24}$$

We obtain from (23) and (24) that

$$\lim_{n\to\infty} P\{\tilde{\beta}_n \in CR_{\alpha,\mathrm{el}}\} = \lim_{n\to\infty} P\{l(\tilde{\beta}_n) < c_\alpha\} = \lim_{n\to\infty} P\{\chi_p^2(||\gamma||^2) < c_\alpha\}.$$

Thus, Theorem 3.2 is proved. □

Proof of Theorem 3.3. Since our $Z_1(\beta_0), Z_2(\beta_0), \ldots, Z_n(\beta_0)$ are iid., we check that conditions (i)–(vi) of Theorem 3.1 in [2] are all satisfied, then we have that $P\{l(\beta_0) < c_\alpha(1+\zeta n^{-1})\} = \alpha + O(n^{-2})$. From (16), $P\{\beta_0 \bar{\in} \Omega_n\} = O(n^{-2})$, and it concludes that

$$\begin{aligned} &P\{\beta_0 \in \Omega_n, \ell(\beta_0) < c_\alpha(1+\zeta n^{-1})\} \\ &\quad = P\{l(\beta_0) < c_\alpha(1+\zeta n^{-1})\} - P\{\beta_0 \bar{\in} \Omega_n, \ell(\beta_0) < c_\alpha(1+\zeta n^{-1})\} \\ &\quad = \alpha + O(n^{-2}). \end{aligned}$$

Acknowledgments

The authors thanks two referees and an Associate Editor for their helpful suggestions which improve the presentation of the paper. The project is supported by the NSFC (Grant 19771011,10071009), the DPFIHE (20020027010), the EYTP of China and an Australian Research Council Grant.

References

[1] R.J. Carroll, D. Ruppert, L.A. Stefanski, Measurement Error in Nonlinear Models, Monographs on Statistics and Applied Probability, Vol. 63, Chapman & Hall, London, 1995.
[2] S.X. Chen, On the accuracy of empirical likelihood confidence regions for linear regression model, Ann. Inst. Statist. Math. 45 (4) (1993) 627–637.
[3] S.X. Chen, Empirical likelihood confidence intervals for linear regression coefficients, J. Multivariate Anal. 50 (4) (1994) 697–714.
[4] S.X. Chen, P. Hall, Smoothed empirical likelihood confidence intervals for quantiles, Ann. Statist. 21 (1993) 1166–1181.
[5] H.J. Cui, R.C. Li, On the parameter estimation for semilinear EV model, J. Multivariate Anal. 1 (1998) 1–64.
[6] T. DiCiccio, P. Hall, J. Romano, Empirical likelihood is Bartlett-correctable, Ann. Statist. 19 (1991) 1053–1061.
[7] W.A. Fuller, Measurement Error Models, Wiley, New York, 1987.

[8] X. He, H. Liang, Quantile regression estimates for a class of linear and partially linear errors-in-variables models, Statist. Sinica 10 (2000) 129–140.
[9] H. Liang, W. Hädle, R.J. Carroll, Estimation in a semiparametric partially linear errors-in-variables model, Ann. Statist. 27 (1999) 1519–1535.
[10] J. Qin, J. Lawless, Empirical likelihood and general estimating equations, Ann. Statist. 22 (1994) 300–325.
[11] A.B. Owen, Empirical likelihood ratio confidence intervals for a single functional, Biometrika 75 (1988) 237–249.
[12] A.B. Owen, Empirical likelihood confidence regions, Ann. Statist. 18 (1990) 90–120.
[13] A.B. Owen, Empirical likelihood for linear models, Ann. Statist. 19 (1991) 1725–1747.