Stat Papers (2012) 53:83–93
DOI 10.1007/s00362-010-0314-9

REGULAR ARTICLE

Two-sample empirical likelihood method for difference between coefficients in linear regression model

Xuemin Zi · Changliang Zou · Yukun Liu

Received: 10 November 2008 / Revised: 30 July 2009 / Published online: 4 February 2010

Abstract The empirical likelihood method is proposed to construct the confidence regions for the difference in value between coefficients of two-sample linear regression model. Unlike existing empirical likelihood procedures for one-sample linear regression models, as the empirical likelihood ratio function is not concave, the usual maximum empirical likelihood estimation cannot be obtained directly. To overcome this problem, we propose to incorporate a natural and well-explained restriction into likelihood function and obtain a restricted empirical likelihood ratio statistic (RELR). It is shown that RELR has an asymptotic chi-squared distribution. Furthermore, to improve the coverage accuracy of the confidence regions, a Bartlett correction is applied. The effectiveness of the proposed approach is demonstrated by a simulation study.

Keywords Bartlett correction · Coverage accuracy · Empirical likelihood · Linear regression model · Two-sample problem

Mathematics Subject Classification (2000) Primary 62G05 · Secondary 62E20

X. Zi
Department of Mathematics, Tianjin University of Technology and Education, Tianjin 300222, China

C. Zou (✉)
Department of Statistics, School of Mathematical Sciences, Nankai University, Tianjin 300071, China
e-mail: chlzou@yahoo.com.cn

Y. Liu
School of Finance and Statistics, East China Normal University, Shanghai, China

1 Introduction

Consider the following linear regression model,

$$y_k = \begin{cases} \boldsymbol{x}_k^\tau \boldsymbol{\beta} + \varepsilon_k, & 1 \leq k \leq n_1, \\ \boldsymbol{x}_k^\tau \boldsymbol{\beta}_1 + \varepsilon_k, & n_1 < k \leq n, \end{cases} \tag{1}$$

where $\boldsymbol{x}_k \in \mathbb{R}^d, 1 \leq k \leq n$ are fixed design points, $y_k \in \mathbb{R}, 1 \leq k \leq n$ are the corresponding responses and $\boldsymbol{\beta}$, $\boldsymbol{\beta}_1$ are unknown d-dimensional coefficients. Here we assume $\varepsilon_k \in \mathbb{R}, 1 \leq k \leq n$ are independent random variables with mean zero. Furthermore, $\{\varepsilon_i, 1 \leq i \leq n_1\}$ and $\{\varepsilon_j, n_1 \leq j \leq n\}$ come from two distributions F and G respectively.

Let $\boldsymbol{\delta} = \boldsymbol{\beta} - \boldsymbol{\beta}_1$. We are interested in constructing confidence intervals for $\boldsymbol{\delta}$ or testing $H_0 : \boldsymbol{\delta} = \boldsymbol{\delta}_0$. This problem has been well addressed in the literature if F and G are both normally distributed (see Sen and Srivastava 1997). In general, when both F and G are unknown, it seems that development of nonparametric approaches is highly desirable.

In this paper, we will apply the empirical likelihood method to this problem. Empirical likelihood that was first introduced by Owen (1988, 1990) is a nonparametric method of statistical inference. It allows the data analyst to use likelihood methods without having to assume that the data comes from a known family of distributions. It has sampling properties similar to bootstrap, but achieves them through profiling a multinomial with one parameter per data point instead of resampling. Many advantages of the empirical likelihood over the normal approximation-based method have been shown in literature. In particular, it does not impose prior constraints on the shape of region, it does not require the construction of a pivotal quantity and the region is range preserving and transformation respecting (see Hall and La Scala 1990). Specially, there are various authors to extend empirical likelihood methodology to regression model; see Owen (1991) and Chen (1993, 1994) for linear models, Kolaczyk (1994) for generalized linear models, Wang and Jing (1999), Shi and Lau (2000) and Zhu and Xue (2006) for partially linear models, Cui and Chen (2003) for linear EV models, *etc*.

Unlike common empirical likelihood procedures for one-sample linear regression models or the two-sample problem of univariate and multivariate mean models (Jing 1995; Liu et al. 2008), as the empirical likelihood ratio function is not concave, the usual maximum empirical likelihood estimation cannot be obtained directly. To attack this difficulty, we propose to incorporate a natural and well-explained restriction into likelihood function and obtain a restricted empirical likelihood ratio statistic (RELR). The RELR is proved to achieve asymptotic χ^2 distribution, which allows us to construct the confidence regions for the difference in value between the coefficients of two-sample linear regression model with the coverage error tending to zero at the rate of n^{-1}. Note that the problem of non-concavity in empirical likelihood has been mentioned in the existing literature, e.g., Chen et al. (2008) which suggested a novel method to overcome the difficulty posed by the nonexistence of solutions while computing the profile empirical likelihood.

Furthermore, we will give a Bartlett correction for the empirical likelihood confidence regions. As we know, an attractive advantage of empirical likelihood against bootstrap is that it admits Bartlett correction in many important situations such as the smooth function model (DiCiccio et al. 1991), linear regression model (Chen 1993, 1994) and two-sample mean problem (Jing 1995; Liu et al. 2008). With Bartlett correction the empirical likelihood ratio statistic has a limiting chi-squared distribution with error being reduced from $O(n^{-1})$ to $O(n^{-2})$. In this article, an explicit formula for the correction which can improve the coverage accuracy of confidence regions is derived. Also, an empirical Bartlett correction is given for practical implementation. Simulation shows that the proposed confidence region has satisfactory coverage accuracy.

This paper is organized as follows. We formulate the empirical likelihood and confidence regions for $\boldsymbol{\delta}$ in Sect. 2. The coverage accuracy and Bartlett correction are also studied in this section. The simulation results are presented in Sect. 3. All the technical proofs are put in the Appendix.

2 Methodology and main results

Recall the model (1) mentioned in Sect. 1. In this paper, it is assumed that $\theta = n_1/n \to \theta_0 \in (0, 1)$ as $n \to \infty$ and $\boldsymbol{\beta}_0$ is the true regression coefficient under H_0. Since then, unless otherwise stated, subscript i runs from 1 to $n\theta$, j from $n\theta + 1$ to n and k from 1 to n, while all superscripts run from 1 to d. Let $y_j^* = y_j + \boldsymbol{x}_j^\tau \boldsymbol{\delta}_0$, $\mathbf{V}_{n_1} = \frac{1}{n\theta} \sum_i \boldsymbol{x}_i \boldsymbol{x}_i^\tau$ and $\mathbf{V}_{n_2} = \frac{1}{n(1-\theta)} \sum_j \boldsymbol{x}_j \boldsymbol{x}_j^\tau$. Using the idea of Owen (1991), we introduce the auxiliary random vector of d components $\boldsymbol{z}_i \equiv \boldsymbol{z}_i(\boldsymbol{\beta}) = \boldsymbol{x}_i(y_i - \boldsymbol{x}_i^\tau \boldsymbol{\beta})$ and $\boldsymbol{z}_j \equiv \boldsymbol{z}_j(\boldsymbol{\beta}) = \boldsymbol{x}_j(y_j^* - \boldsymbol{x}_j^\tau \boldsymbol{\beta})$. Let $(p_1, p_2, \cdots, p_{n\theta})$ and $(q_{n\theta+1}, q_{n\theta+2}, \cdots, q_n)$ be probability vectors (that is, $\sum p_i = 1$, $\sum q_j = 1$ and $p_i \geq 0$, $q_j \geq 0$). The empirical likelihood for $\boldsymbol{\delta}$, evaluated at $\boldsymbol{\delta}_0$, is defined as

$$L(\boldsymbol{\delta}_0) = \sup\left\{ \prod_i p_i \prod_j q_j \Big| \sum_i p_i \boldsymbol{z}_i = \sum_j q_j \boldsymbol{z}_j = 0 \right\}$$

and the corresponding empirical log-likelihood ratio is defined as

$$\begin{aligned} l(\boldsymbol{\delta}_0) &= -2 \ln L(\boldsymbol{\delta}_0)/L(\hat{\boldsymbol{\delta}}) \\ &= -2 \sup \left\{ \sum_i \ln(n\theta p_i) + \sum_j \ln(n(1-\theta) q_j) \Big| \sum_i p_i \boldsymbol{z}_i = \sum_j q_j \boldsymbol{z}_j = 0 \right\} \end{aligned}$$

where $\hat{\boldsymbol{\delta}} = (\sum_i \boldsymbol{x}_i \boldsymbol{x}_i^\tau)^{-1} \sum_i \boldsymbol{x}_i^\tau y_i - (\sum_j \boldsymbol{x}_j \boldsymbol{x}_j^\tau)^{-1} \sum_j \boldsymbol{x}_j^\tau y_j$ and we have used the fact that $L(\hat{\boldsymbol{\delta}}) = (n\theta)^{-n\theta}[n(1-\theta)]^{-n(1-\theta)}$.

The lagrange multiplier method leads to

$$p_i = \frac{1}{n\theta} \frac{1}{1 + \theta^{-1} \boldsymbol{\lambda}_1^\tau \boldsymbol{z}_i} \quad q_j = \frac{1}{n(1-\theta)} \frac{1}{1 - (1-\theta)^{-1} \boldsymbol{\lambda}_2^\tau \boldsymbol{z}_j},$$

where the multipliers $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2 \in \mathbb{R}^d$. Consequently, the maximum log-likelihood ratio is

$$l(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\beta}) = 2\Big[\sum_i \ln\{1+\theta^{-1}\boldsymbol{\lambda}_1^\tau z_i\} + \sum_j \ln\{1-(1-\theta)^{-1}\boldsymbol{\lambda}_2^\tau z_j\}\Big]. \quad (2)$$

In general, the empirical likelihood method is to seek the maximum of (2) under the following inequality constraints

$$1+\theta^{-1}\boldsymbol{\lambda}_1^\tau z_i > \frac{1}{n\theta}, \quad 1-(1-\theta)^{-1}\boldsymbol{\lambda}_2^\tau z_j > \frac{1}{n(1-\theta)}.$$

However, such maximum mechanism does't work in this case because the function $l(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\beta})$ is not concave, which is easy to be verified. To overcome this problem, consider one of the score functions of (2)

$$\frac{\partial l(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\beta})}{-2\partial\boldsymbol{\beta}} = \sum_i p_i \boldsymbol{x}_i \boldsymbol{x}_i^\tau \boldsymbol{\lambda}_1 - \sum_j q_j \boldsymbol{x}_j \boldsymbol{x}_j^\tau \boldsymbol{\lambda}_2 = 0.$$

It can be shown that the solution to the above score equation satisfies $\boldsymbol{\lambda}_1 = O_p(n^{-1/2})$ and $\boldsymbol{\lambda}_2 = O_p(n^{-1/2})$. Therefore the resulting probability weights p_is and q_js are approximately $\frac{1}{n\theta}$ and $\frac{1}{n(1-\theta)}$, respectively. This motivates us to set $p_i = \frac{1}{n\theta}$ and $q_j = \frac{1}{n(1-\theta)}$ in the above equation, and restrict $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ in

$$\Omega = \Big\{(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) : \mathbf{V}_{n_1}\boldsymbol{\lambda}_1 = \mathbf{V}_{n_2}\boldsymbol{\lambda}_2\Big\}.$$

It follows that a restricted empirical likelihood ratio (RELR) function can be defined as

$$\begin{aligned} l_R(\boldsymbol{\lambda}, \boldsymbol{\beta}) &\equiv l(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\beta} | (\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) \in \Omega) \\ &= 2\Big[\sum_i \ln\{1+\theta^{-1}\boldsymbol{\lambda}^\tau z_i\} + \sum_j \ln\{1-(1-\theta)^{-1}\boldsymbol{\lambda}^\tau \mathbf{V}_{n_1}\mathbf{V}_{n_2}^{-1} z_j\}\Big]. \end{aligned}$$

Let $(\widehat{\boldsymbol{\lambda}}, \widehat{\boldsymbol{\beta}})$ be the solution to the score equations:

$$\begin{cases} \sum_i \dfrac{z_i}{\theta + \lambda^\tau z_i} - \sum_j \dfrac{\mathbf{V}_{n_1}\mathbf{V}_{n_2}^{-1} z_j}{1-\theta-\boldsymbol{\lambda}^\tau \mathbf{V}_{n_1}\mathbf{V}_{n_2}^{-1} z_j} = 0, \\ \sum_i \dfrac{\boldsymbol{x}_i \boldsymbol{x}_i^\tau \boldsymbol{\lambda}}{\theta + \boldsymbol{\lambda}^\tau z_i} - \sum_j \dfrac{\boldsymbol{x}_j \boldsymbol{x}_j^\tau \mathbf{V}_{n_2}^{-1}\mathbf{V}_{n_1}\boldsymbol{\lambda}}{1-\theta-\boldsymbol{\lambda}^\tau \mathbf{V}_{n_1}\mathbf{V}_{n_2}^{-1} z_j} = 0 \end{cases} \quad (3)$$

and define $l_1(\boldsymbol{\delta}_0) = l_R(\widehat{\boldsymbol{\lambda}}, \widehat{\boldsymbol{\beta}})$. It can be shown that the function $l_R(\boldsymbol{\lambda}, \boldsymbol{\beta})$ is at least concave in the $n^{-\frac{1}{2}}$ neighborhood of $(0, \boldsymbol{\beta}_0)$ and asymptotic concave in the neighborhood of $(\widehat{\boldsymbol{\lambda}}, \widehat{\boldsymbol{\beta}})$, which means that the maximum value can be attained as $n \to \infty$.

In the following theorem, we shall establish a nonparametric version of Wilk's theorem for RELR. To this end, we make the following assumptions. Let $v_{dn}^{(1)}$, $v_{1n}^{(1)}$ and $v_{dn}^{(2)}$, $v_{1n}^{(2)}$ be the smallest and largest eigenvalues of $\mathbf{V}_{n_1}$ and $\mathbf{V}_{n_2}$, respectively.

C1. There exist positive constants m and M such that for all $n, m < v_{dn}^{(1)}, v_{dn}^{(2)} \leq v_{1n}^{(1)}, v_{1n}^{(2)} < M$.

C2. $\frac{E\varepsilon_1^4}{(n\theta)^2}\sum_i(\boldsymbol{x}_i^\tau \boldsymbol{x}_i)^2 + \frac{E\varepsilon_n^4}{(n(1-\theta))^2}\sum_j(\boldsymbol{x}_j^\tau \boldsymbol{x}_j)^2 \to 0$ when $n \to \infty$.

We have the following theorem.

Theorem 1 *Let $\boldsymbol{\delta}_0$ be the true value of $\boldsymbol{\delta}$. Then under the conditions* C1 *and* C2*, $l_1(\boldsymbol{\delta}_0)$ has an asymptotic χ_d^2 distribution. Furthermore, $P(l_1(\boldsymbol{\delta}_0) < c_\alpha) = \alpha + O(n^{-1})$, where c_α is the α percentile of the χ_d^2 distribution.*

From Theorem 1, one can construct an α-level confidence region for $\boldsymbol{\delta}$ by

$$I_\alpha = \{\boldsymbol{\delta} : l_1(\boldsymbol{\delta}) < c_\alpha\}.$$

Then I_α will have correct asymptotic coverage accuracy in the sense that its coverage error tends to zero at the rate of n^{-1}, *i.e.*

$$P(\boldsymbol{\delta} \in I_\alpha) = \alpha + O(n^{-1}).$$

For one-sample linear regression model, the order $O(n^{-1})$ was pointed out by Chen (1993). Besides, Chen (1993) also showed that using Bartlett correction could reduce the coverage error of empirical likelihood confidence regions from order $O(n^{-1})$ to order $O(n^{-2})$. Here we show the proposed $l_1(\boldsymbol{\delta}_0)$ for the two-sample linear regression problem is also Bartlett correctable.

Let $\boldsymbol{U} = (G^1, \ldots, G^t, \ldots, G^{ddd})$ (see the Appendix for the definitions of G's). Some additional regularity conditions are needed.

C3. The smallest eigenvalue of $n\mathrm{Cov}(\boldsymbol{U})$ is bounded away from 0.

C4. There exist $M_1 > m_1 > 0$, such that $m_1 < \inf\|\boldsymbol{x}_k\| \leq \sup\|\boldsymbol{x}_k\| < M_1$ uniformly in n.

C5. $E|\varepsilon_i|^{15} < \infty$ and $E|\varepsilon_j|^{15} < \infty$.

C6. For all $b > 0$, $\max\limits_{1\leq l\leq 2}\sup\limits_{\|t\|>b}\|g_l(t)\| < 1$, where $g_1(t) = Ee^{i\varepsilon_1 t}$ and $g_2(t) = Ee^{i\varepsilon_n t}$.

Remark 1 Under condition C3, the vector $\boldsymbol{U}$ is of full rank. Condition C4 assures that matrices such as $\mathbf{V}_{n1}, \mathbf{V}_{n2}$ are uniformly nonsingular and bounded for n larger than some integer. Condition C5 is similar to that of Chen (1993) and assures the validity of the Edgeworth expansion of $l_1(\boldsymbol{\delta}_0)$. Condition C6 is just common Cramer's condition.

For notation convenience, define $\mathbf{H} = [\theta\sigma_2^2\mathbf{V}_{n_1} + (1-\theta)\sigma_1^2\mathbf{V}_{n_2}]^{-1}$, $\mathbf{M}_n = \theta(1-\theta)\mathbf{V}_{n_1}\mathbf{H}\mathbf{V}_{n_2}$, $\boldsymbol{u}_i = \mathbf{M}_n^{\frac{1}{2}}\mathbf{V}_{n_1}^{-1}\boldsymbol{x}_i/\theta$ and $\boldsymbol{u}_j = -\mathbf{M}_n^{\frac{1}{2}}\mathbf{V}_{n_2}^{-1}\boldsymbol{x}_j/(1-\theta)$, where $\sigma_1^2 = E\varepsilon_1^2$ and $\sigma_2^2 = E\varepsilon_n^2$ are the second moments of F and G respectively. For a vector $\boldsymbol{u}$ we use $\boldsymbol{u}^r$ to denote its r-th component. Further define

$$\eta = \sum_{r,k,l} \Big(\frac{1}{n} \sum_i \boldsymbol{u}_i^r \boldsymbol{u}_i^k \boldsymbol{u}_i^l E\varepsilon_1^3 + \frac{1}{n} \sum_j \boldsymbol{u}_j^r \boldsymbol{u}_j^k \boldsymbol{u}_j^l E\varepsilon_n^3 \Big)^2,$$

$$\zeta = \sum_{r,k} \Big(\frac{1}{n} \sum_i \boldsymbol{u}_i^r \boldsymbol{u}_i^r \boldsymbol{u}_i^k \boldsymbol{u}_i^k E\varepsilon_1^4 + \frac{1}{n} \sum_j \boldsymbol{u}_j^r \boldsymbol{u}_j^r \boldsymbol{u}_j^k \boldsymbol{u}_j^k E\varepsilon_n^4 \Big),$$

$$\xi = \sum_{r,s,t} \frac{1}{n} \sum_k \boldsymbol{u}_k^r \boldsymbol{u}_k^r \boldsymbol{x}_k^s \boldsymbol{x}_k^t \mathbf{H}^{st},$$

where $\mathbf{H}^{st}$ denotes the (s,t)-th element of $\mathbf{H}$. 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. Then we have the following theorem.

Theorem 2 *Assume that conditions* C1 − C6 *hold. Let* $\boldsymbol{\delta}_0$ *be the true value of* $\boldsymbol{\delta}$*, then we have*

$$P\big(l_1(\boldsymbol{\delta}_0) \le c_\alpha(1 + an^{-1})\big) = \alpha + O(n^{-2}). \tag{4}$$

where $a = -\eta/3d + \zeta/2d + \xi/d$.

In practical application, the moments of ε_i and ε_j are unknown. Consider the least square estimates $\boldsymbol{\beta}_{L1}$ and $\boldsymbol{\beta}_{L2}$ for $\{(\boldsymbol{x}_i, y_i)\}$ and $\{(\boldsymbol{x}_j, y_j)\}$ respectively, then the estimates of the kth moments are defined as $\widehat{\mu}_{k1} = \frac{1}{n\theta} \sum_i (y_i - \boldsymbol{x}_i^\tau \boldsymbol{\beta}_{L1})^k$ and $\widehat{\mu}_{k2} = \frac{1}{n(1-\theta)} \sum_j (y_j - \boldsymbol{x}_j^\tau \boldsymbol{\beta}_{L2})^k$. We can substitute these moment estimates for the corresponding unknown ones in (4) to obtain $\sqrt{n}$-consistent estimates $\widehat{\eta}$, $\widehat{\zeta}$, $\widehat{\xi}$ and correspondingly obtain an estimate $\widehat{a} = -\widehat{\eta}/3d + \widehat{\zeta}/2d + \widehat{\xi}/d$ for a. Replacing a by $\widehat{a}$ in (4) does't affect the theorem (by utilizing the parity property of the polynomials in Edgeworth expansion; see Hall and La Scala 1990, Sect. 3.3). So we can construct an α-level confidence region for $\boldsymbol{\delta}$ based on (4), which is

$$J_\alpha = \{\boldsymbol{\delta} : l_1(\boldsymbol{\delta}) \le c_\alpha(1 + \widehat{a}n^{-1})\}.$$

The corresponding coverage error is reduced from $O(n^{-1})$ to $O(n^{-2})$.

3 A simulation study

In this section we report a simulation study designed to evaluate the performance of the proposed empirical likelihood confidence region. Under consideration is the following simple linear regression model:

$$y_k = 1 + x_k + \varepsilon_k, \quad 1 \le k \le n. \tag{5}$$

The distributions of x_i and x_j are both standard Normal. In this case, $d = 2$, $\boldsymbol{\beta}_0 = (1, 1)$ and $\boldsymbol{x}_k = (1, x_k)$ for $1 \le k \le n$. Three error patterns are considered: (i) $\varepsilon_i = N(0, 1)$ and $\varepsilon_j = N(0, 1)$; (ii) $\varepsilon_i = t(4)$ and $\varepsilon_j = N(0, 1)$; (iii) $\varepsilon_i = \chi_4^2 - 4.0$, $\varepsilon_j = \exp(1.0) - 1.0$, where $N(0, 1)$, $\exp(1.0)$, $\chi^2(4)$ and $t(4)$ are random variables

Table 1 The coverage probability comparisons of confidence regions with three error patterns for $\alpha = 0.95$

Error	θ	Bootstrap	Uncorrected I_α	J_α with a	J_α with $\widehat{a}$
(i)	0.250	0.923	0.922	0.939	0.935
	0.375	0.930	0.933	0.942	0.939
	0.500	0.938	0.938	0.945	0.943
(ii)	0.250	0.908	0.906	0.931	0.923
	0.375	0.917	0.920	0.940	0.936
	0.500	0.932	0.931	0.941	0.938
(iii)	0.250	0.898	0.896	0.920	0.911
	0.375	0.908	0.911	0.931	0.925
	0.500	0.927	0.928	0.937	0.934

Table 2 Simulated results on the lengths of confidence intervals and two one-sided errors for slope parameter at $\alpha = 0.95$

Error	θ	Bootstrap			Uncorrected I_α			J_α with $\widehat{a}$		
		LCI	LE	UE	LCI	LE	UE	LCI	LE	UE
(i)	0.250	0.90	0.44	0.31	0.90	0.49	0.18	0.95	0.42	0.20
	0.375	0.68	0.38	0.25	0.66	0.30	0.26	0.69	0.40	0.11
	0.500	0.51	0.15	0.17	0.51	0.21	0.14	0.52	0.15	0.05
(ii)	0.250	1.63	0.60	0.49	1.32	0.83	0.47	2.14	0.43	0.28
	0.375	1.24	0.50	0.34	1.10	0.56	0.31	1.52	0.38	0.18
	0.500	0.76	0.26	0.21	0.89	0.34	0.12	1.17	0.21	0.06
(iii)	0.250	2.40	0.03	2.75	2.54	0.42	1.99	2.82	0.45	1.21
	0.375	1.86	0.20	1.66	1.92	0.14	1.17	2.14	0.23	0.97
	0.500	1.45	0.06	0.65	1.47	0.12	0.42	1.51	0.02	0.28

LCI length of confidence interval, *LE* (lower-sided error)×100, *UE* (upper-sided error)×100

with standard Normal distribution, exponential distribution with unit mean, chi-square distribution with four degrees of freedom and student-t distribution with four degrees of freedom. For each of these three error patterns, we choose the sample size $n = 200$, $\theta = 0.25, 0.375, 0.5$ and nominal coverage level $\alpha = 0.95$. Table 1 shows the simulated coverage percentage comparisons for confidence regions through ten thousand replications. The four values in each row are the coverages of the confidence regions obtained by studentized bootstrap approach (see Davison and Hinkley 1997), the uncorrected regions I_α and two corrected regions J_α which use a and $\widehat{a}$ respectively. As we can expect, the unadjusted empirical likelihood has similar performance to the bootstrap approach because they both have theoretical coverage errors of order $O(n^{-1})$. The corrected empirical confidence regions perform similarly to their theoretical counterparts. The empirical likelihood with Bartlett corrections work uniformly better than the uncorrected one and in come cases the improvements are quite remarkable.

To further check the efficiency of the proposed RELR approach, we consider the case that the model (5) only contains the slope term, i.e., $d = 1$. This simple model allows us to compute the length of confidence interval which is also a good benchmark for evaluating the relative performance of various methods. We only present the results at nominal coverage level $\alpha = 0.95$ because similar conclusion holds for other levels. The studentized bootstrap approach, the uncorrected regions I_α and corrected regions J_α with $\widehat{a}$ are considered. Besides the lengths of confidence intervals, the two one-sided errors are also presented (DiCiccio and Romano 1989). The simulated results over ten thousand replications under the same settings of Table 1 are summarized in Table 2. Similar to the findings of Table 1, the unadjusted RELR has comparable performance to the bootstrap approach in terms of both lengths of confidence intervals and one-sided errors. In comparison, the corrected RELR usually has wider intervals which results in smaller sum of one-sided errors.

Acknowledgements The authors thank the editor, and an anonymous referee for many constructive comments that greatly improved the article. This work was supported by the NNSF of China grant No. 10771107, 10711120448, 10901092, 10926063.

Appendix: Proofs

Throughout the Appendix, we use the following additional notation:

$$\boldsymbol{w}_i \overset{\triangle}{=} \boldsymbol{w}_i(\boldsymbol{\beta}) = \mathbf{M}_n^{\frac{1}{2}} \mathbf{V}_{n_1}^{-1} z_i, \quad \boldsymbol{w}_j \overset{\triangle}{=} \boldsymbol{w}_j(\boldsymbol{\beta}) = \mathbf{M}_n^{\frac{1}{2}} \mathbf{V}_{n_2}^{-1} \boldsymbol{z}_j,$$

where $\mathbf{M}_n$ is defined in Sect. 2. In this appendix, all the expressions in Sect. 2 are rewritten by using $\boldsymbol{w}_i$ and $\boldsymbol{w}_j$ instead of z_i and z_j, which makes the arguments more concise.

First we rewrite $l_R(\boldsymbol{\lambda}, \boldsymbol{\beta})$ in terms of $\boldsymbol{w}_i$ and $\boldsymbol{w}_j$,

$$l(\boldsymbol{\lambda}, \boldsymbol{\beta}) = 2 \sum_i \ln \left(1 + \frac{1}{\theta} \boldsymbol{\lambda}^\tau \boldsymbol{w}_i \right) + 2 \sum_j \ln \left(1 - \frac{1}{1-\theta} \boldsymbol{\lambda}^\tau \boldsymbol{w}_j \right).$$

Meanwhile Eq. 3 become

$$\begin{cases} \sum\limits_i \frac{\boldsymbol{w}_i}{\theta + \boldsymbol{\lambda}^\tau \boldsymbol{w}_i} - \sum\limits_j \frac{\boldsymbol{w}_j}{1-\theta-\boldsymbol{\lambda}^\tau \boldsymbol{w}_j} = 0, \\ \mathbf{M}_n^{\frac{1}{2}} \mathbf{V}_{n_1}^{-1} \sum\limits_i \frac{\boldsymbol{x}_i \boldsymbol{x}_i^\tau \boldsymbol{\lambda}}{\theta + \boldsymbol{\lambda}^\tau \boldsymbol{w}_i} - \mathbf{M}_n^{\frac{1}{2}} \mathbf{V}_{n_2}^{-1} \sum\limits_j \frac{\boldsymbol{x}_j \boldsymbol{x}_j^\tau \boldsymbol{\lambda}}{1-\theta-\boldsymbol{\lambda}^\tau \boldsymbol{w}_j} = 0. \end{cases} \tag{A.1}$$

To prove Theorem 1, we need the following lemma.

Lemma 1 *Under the conditions of Theorem 1, let $(\widehat{\boldsymbol{\lambda}}, \widehat{\boldsymbol{\beta}})$ be the solution to Eq. A.1, we have*

$$\widehat{\boldsymbol{\lambda}} = O_p(n^{-\frac{1}{2}}), \quad \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = o_p(1).$$

Proof Let $l(\boldsymbol{\beta}) = l_R(\boldsymbol{\lambda}(\boldsymbol{\beta}), \boldsymbol{\beta})$, where $\boldsymbol{\lambda}(\boldsymbol{\beta})$ is determined by the first equation of (A.1). For $\boldsymbol{\beta} = \boldsymbol{\beta}_0 + \boldsymbol{\gamma}$, where $\| \boldsymbol{\gamma} \| = \varepsilon > 0$, expanding the first equation of (A.1) and following similar arguments to Owen (2001, p. 220), we can obtain

$$\boldsymbol{\lambda} = \mathbf{S}(\boldsymbol{\beta})^{-1}\boldsymbol{\psi} + o_p(n^{-\frac{1}{2}}),$$

where

$$\boldsymbol{\psi} = \frac{1}{n\theta}\sum_i \boldsymbol{w}_i(\boldsymbol{\beta}) - \frac{1}{n(1-\theta)}\sum_j \boldsymbol{w}_j(\boldsymbol{\beta})$$

and

$$\mathbf{S}(\boldsymbol{\beta}) = \frac{1}{n\theta^2}\sum_i \boldsymbol{w}_i(\boldsymbol{\beta})\boldsymbol{w}_i(\boldsymbol{\beta})^{\tau} + \frac{1}{n(1-\theta)^2}\sum_j \boldsymbol{w}_j(\boldsymbol{\beta})\boldsymbol{w}_j(\boldsymbol{\beta})^{\tau}.$$

Note that here $\boldsymbol{\psi}$ is independent of $\boldsymbol{\beta}$.

Then expanding $l(\boldsymbol{\beta})$ leads to

$$l(\boldsymbol{\beta}) = \boldsymbol{\psi}\mathbf{S}^{-1}(\boldsymbol{\beta})\boldsymbol{\psi} + o_p(1) = \boldsymbol{\psi}(\mathbf{S}(\boldsymbol{\beta}_0) + \boldsymbol{\phi})^{-1}\boldsymbol{\psi} + o_p(1),$$

where

$$\begin{aligned}\boldsymbol{\phi} = &\frac{1}{n\theta^2}\sum_i \mathbf{M}_n^{\frac{1}{2}}\mathbf{V}_{n_1}^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^{\tau}\boldsymbol{\gamma}\boldsymbol{\gamma}^{\tau}\boldsymbol{x}_i\boldsymbol{x}_i^{\tau}\mathbf{V}_{n_1}^{-1}\mathbf{M}_n^{\frac{1}{2}}\\ &+\frac{1}{n(1-\theta)^2}\sum_j \mathbf{M}_n^{\frac{1}{2}}\mathbf{V}_{n_2}^{-1}\boldsymbol{x}_j\boldsymbol{x}_j^{\tau}\boldsymbol{\gamma}\boldsymbol{\gamma}^{\tau}\boldsymbol{x}_j\boldsymbol{x}_j^{\tau}\mathbf{V}_{n_2}^{-1}\mathbf{M}_n^{\frac{1}{2}}.\end{aligned}$$

We can see that when $\boldsymbol{\gamma} = 0$, that is $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, $\boldsymbol{\psi}(\mathbf{S}(\boldsymbol{\beta_0}) + \phi)^{\mathbf{-1}}\boldsymbol{\psi}$ achieves its maximum in the neighborhood of $\boldsymbol{\beta}_0$. This means when n tends to infinity in any ε-neighborhood of $\boldsymbol{\beta}_0$ the function $l_R(\boldsymbol{\lambda}, \boldsymbol{\beta})$ has one local maximum, which indicates $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = o_p(1)$. Therefore the above approximation for $\boldsymbol{\lambda}$ is valid. In fact, since $\mathbf{S}(\boldsymbol{\beta}_0) = \mathbf{I}_d + O_p(n^{-\frac{1}{2}})$,

$$\widehat{\boldsymbol{\lambda}} = \mathbf{S}^{-1}(\boldsymbol{\beta}_0)\boldsymbol{\psi} + o_p(n^{-\frac{1}{2}}) = \boldsymbol{\psi} + o_p(n^{-\frac{1}{2}}) = O_p(n^{-\frac{1}{2}}),$$

which completes the proof. □

Proof of Theorem 1 From the proof of Lemma 1 we have

$$\widehat{\boldsymbol{\lambda}} = \boldsymbol{\psi} + o_p(n^{-\frac{1}{2}}) = \frac{1}{n\theta}\sum_i \mathbf{M}_n^{\frac{1}{2}}\mathbf{V}_{n_1}^{-1}\boldsymbol{x}_i y_i - \frac{1}{n(1-\theta)}\sum_j \mathbf{M}_n^{\frac{1}{2}}\mathbf{V}_{n_2}^{-1}\boldsymbol{x}_j y_j + o_p(n^{-\frac{1}{2}}).$$

By Linderberg-Feller Theorem, we know that $\sqrt{n}\widehat{\boldsymbol{\lambda}} \xrightarrow{L} N(0, \mathbf{I}_d)$. Using Taylor expansion of $l_R(\widehat{\boldsymbol{\lambda}}, \widehat{\boldsymbol{\beta}})$ and treating high-order terms similarly as in the standard empirical likelihood method (Owen 2001), we can get

$$l_R(\widehat{\boldsymbol{\lambda}}, \widehat{\boldsymbol{\beta}}) = n\widehat{\boldsymbol{\lambda}}^\tau \widehat{\boldsymbol{\lambda}} + o_p(1) \xrightarrow{L} \chi^2_d,$$

which completes the proof. □

Before giving the proof of Theorem 2, we need some additional notation. Define

$$\begin{aligned}
g^{t_1 t_2 \cdots t_l} &= \frac{1}{n\theta^l} \sum_i E(\boldsymbol{w}_{i0}^{t_1} \boldsymbol{w}_{i0}^{t_2} \cdots \boldsymbol{w}_{i0}^{t_l}) + \frac{1}{n(\theta-1)^l} \sum_j E(\boldsymbol{w}_{j0}^{t_1} \boldsymbol{w}_{j0}^{t_2} \cdots \boldsymbol{w}_{j0}^{t_l}), \\
G^{t_1 t_2 \cdots t_l} &= \frac{1}{n\theta^l} \sum_i \boldsymbol{w}_{i0}^{t_1} \boldsymbol{w}_{i0}^{t_2} \cdots \boldsymbol{w}_{i0}^{t_l} + \frac{1}{n(\theta-1)^l} \sum_j \boldsymbol{w}_{j0}^{t_1} \boldsymbol{w}_{j0}^{t_2} \cdots \boldsymbol{w}_{j0}^{t_l} - g^{t_1 t_2 \cdots t_l}, \\
G_1^{t_1 t_2 \cdots t_l} &= \frac{1}{n\theta^l} \sum_i \boldsymbol{w}_{i1}^{t_1} \boldsymbol{w}_{i1}^{t_2} \cdots \boldsymbol{w}_{i1}^{t_l} + \frac{1}{n(\theta-1)^l} \sum_j \boldsymbol{w}_{j1}^{t_1} \boldsymbol{w}_{j1}^{t_2} \cdots \boldsymbol{w}_{j1}^{t_l} - g^{t_1 t_2 \cdots t_l},
\end{aligned}$$

where $\boldsymbol{w}_{k0} = \boldsymbol{w}_k(\boldsymbol{\beta}_0)$ and $\boldsymbol{w}_{k1} = \boldsymbol{w}_k(\boldsymbol{\beta}_{(1)})$ and $\boldsymbol{\beta}_{(1)}$ is defined in the following proof of Theorem 2.

Proof of Theorem 2 Based on Lemma 1, by lengthy algebra we can obtain $\boldsymbol{\beta}_{(r)}$ and $\boldsymbol{\xi}_r, r = 1, 2, 3$ which satisfy $\boldsymbol{\beta}_{(r)} - \widehat{\boldsymbol{\beta}} = O_p(n^{-\frac{1+r}{2}})$ and $\boldsymbol{\xi}_r - \widehat{\boldsymbol{\lambda}} = O_p(n^{-\frac{1+r}{2}})$. The details are omitted here and a similar manipulation can be found in Liu et al. (2008).

Similar to the methodology given in DiCiccio et al. (1991), by using $\boldsymbol{\beta}_{(r)}$ and $\boldsymbol{\xi}_r, r = 1, 2, 3$, we have the following Taylor expansion for $l_1(\boldsymbol{\delta}_0)$,

$$n^{-1} l_1(\boldsymbol{\delta}_0) = \Delta_1 + \Delta_2 + O_p\left(n^{-\frac{5}{2}}\right), \tag{A.2}$$

where

$$\begin{aligned}
\Delta_1 &= G^r G^r - G^{rs} G^r G^s + \frac{2}{3} g^{ruv} G^r G^u G^v + G^{rs} G^{st} G^r G^t + \frac{2}{3} G^{ruv} G^r G^u G^v \\
&\quad - 2 g^{rst} G^{ru} G^s G^u G^t + g^{rst} g^{ruv} G^s G^t G^u G^v - \frac{1}{2} g^{rstu} G^r G^s G^t G^u, \\
\Delta_2 &= (G^{rs} - G_1^{rs}) G_r G_s + \frac{2}{3} (G_1^{ruv} - G^{ruv}) G^r G^u G^v.
\end{aligned}$$

Note that Δ_1 is similar to the expansion of one-sample empirical log-likelihood ratio, compared with the terms given in Chen (1993). However, Δ_2 is a new and unique term for the two-sample case. The existence of this term is because under H_0 we know nothing but the difference about the two coefficients, therefore we have to estimate them. For a related but more detailed discussion on two-sample multivariate mean problem, please refer to Liu et al. (2008).

Next, we decompose $l_1(\boldsymbol{\delta}_0)$ from (A.2) as

$$l_1(\boldsymbol{\delta}_0) = (\sqrt{n} \boldsymbol{R}^\tau)(\sqrt{n} \boldsymbol{R}) + O_p(n^{-\frac{3}{2}}),$$

where $\boldsymbol{R} = \boldsymbol{R}_1 + \boldsymbol{R}_2 + \boldsymbol{R}_3$ is d-dimensional vector and $\boldsymbol{R}_i = O_p(n^{-\frac{i}{2}})$ for $i = 1, 2, 3$.

$$\begin{aligned}
\boldsymbol{R}_1^r &= G^r, \quad \boldsymbol{R}_2^r = -\frac{1}{2}G_1^{rk}G^k + \frac{1}{3}g^{rkl}G^kG^l,\\
\boldsymbol{R}_3^r &= \frac{4}{9}g^{ukl}g^{urv}G^kG^lG^v + \frac{3}{8}G^{rk}G^{kl}G^l - \frac{5}{12}g^{krl}G^{ku}G^uG^l\\
&\quad -\frac{5}{12}g^{url}G^{ur}G^kG^l + \frac{1}{3}G_1^{rkl}G^kG^l - \frac{1}{4}g^{rklm}G^kG^lG^m.
\end{aligned}$$

Checking the conditions and applying Theorem 2.1 of Chen (1993) lead to the following Edgeworth expansion:

$$P(l_1(\boldsymbol{\delta}_0) < c_\alpha) = \alpha - ac_\alpha h_d(c_\alpha)n^{-1} + O(n^{-\frac{3}{2}})$$

where h_d is the density of χ_d^2 distribution, $P(\chi_d^2 < c_\alpha) = \alpha$ and a is defined in Sect. 2. This completes the proof of Theorem 2. □

References

Chen SX (1993) On the accuracy of empirical likelihood confidence regions for linear regression model. Ann Inst Stat Math 45:621–637

Chen SX (1994) Empirical likelihood confidence intervals for linear regression coefficients. J Multivar Anal 49:24–40

Chen JH, Variyath AM, Abraham B (2008) Adjusted empirical likelihood and its properties. J Comput Graph Stat 17:426–443

Cui HJ, Chen SX (2003) Empirical likelihood confidence region for parameter in the errors-in-variables models. J Multivar Anal 84:101–115

Davison AC, Hinkley DV (1997) Bootstrap methods and their applications. Cambridge University Press, Cambridge, UK

DiCiccio TJ, Romano JP (1989) On adjustment based on the signed root of the empirical likelihood ratio statistic. Biometrika 76:447–456

DiCiccio TJ, Hall P, Romano JP (1991) Bartlett adjustment for empirical likelihood. Ann Stat 19:1053–1061

Hall P, La Scala B (1990) Methodology and algorithms of empirical likelihood. Int Stat Rev 58:109–127

Jing B (1995) Two-sample empirical likelihood method. Stat Probab Lett 24:315–319

Kolaczyk ED (1994) Empirical likelihood for generalized linear models. Stat Sin 4:199–218

Liu Y, Zou C, Zhang R (2008) The Bartlett correction for two-sample empirical likelihood. Stat Probab Lett 78:548–556

Owen AB (1988) Empirical likelihood ratio confidence intervals for a single functional. Biometrika 75:237–249

Owen AB (1990) Empirical likelihood ratio confidence regions. Ann Stat 18:90–120

Owen AB (1991) Empirical likelihood for linear regression model. Ann Stat 19:1725–1747

Sen A, Srivastava M (1997) Regression analysis: theory, methods, and applications. Springer, Verlag

Shi J, Lau TS (2000) Empirical likelihood for partially linear models. J Multivar Anal 72:132–148

Wang QH, Jing BY (1999) Empirical likelihood for partial linear models with fixed designs. Stat Probab Lett 41:425–433

Zhu LX, Xue LG (2006) Empirical likelihood confidence regions in a partially linear single-index model. J R Stat Soc B 68:549–570