

COMPARISON STUDY ON THE CONFIDENCE INTERVALS IN LINEAR REGRESSION MODELS WITH RESTRICTED PARAMETER SPACE

XIUQIN BAI

Department of Mathematics, Eastern Washington University, Cheney, WA 99004, USA
Email: xbai@ewu.edu

WEIXING SONG*

Department of Statistics, Kansas State University, Manhattan, KS 66503, USA
Email: weixing@ksu.edu

SUMMARY

This paper proposes an empirical likelihood confidence region for the regression coefficients in linear regression models when the regression coefficients are subjected to some equality constraints. The shape of the confidence set does not depend on the reparametrization of the regression model induced by the equality constraint. It is shown that the asymptotic coverage rate attains the nominal confidence level and the Bartlett correction can successfully reduce the coverage error rate from $O(n^{-1})$ to $O(n^{-2})$, where n denotes the sample size. Simulation studies are conducted to evaluate the finite sample performance of the proposed empirical likelihood empirical confidence estimation procedure. Finally, a comparison study is conducted to compare the finite sample performance of the proposed and the classical ellipsoidal confidence sets based on normal theory.

Keywords and phrases: Linear Regression Model; Empirical Likelihood; Bartlett Correction; Edgeworth Expansion

AMS Classification: 62G08; secondary 62G10

1 Introduction

A classical statistical inference in linear regression models is to construct confidence regions for the regression coefficients. In a parametric setup, the random error is often assumed to obey a normal law. As a result, a t -type or F -type confidence region can be built from the distributions of least square estimate or MLE of the regression coefficients. If the distribution of the random error is unknown, we may seek some large sample confidence regions based on the asymptotic distributions of some estimates of the regression coefficient, or construct some bootstrap confidence regions. But a main drawback of the large sample confidence regions and the bootstrap ones is that the shapes and orientations of the confidence region are predetermined by the large sample theory or the bootstrap procedure itself.

* Corresponding author

© Institute of Statistical Research and Training (ISRT), University of Dhaka, Dhaka 1000, Bangladesh.

As a nonparametric alternative to the bootstrap procedure for constructing the confidence regions, the empirical likelihood (EL) procedure, introduced by Owen (1988, 1990, 1991), have gained much popularity during the past two decades. Unlike the bootstrap procedure, the EL does not re-sample the data with equal probability, instead, a set of constraints on the resampling probabilities enable the EL procedure to be able to collect the important features of the data, which in turn implies that the EL uses only the data to determine the shape and orientation of the confidence regions. Compared to the large sample confidence intervals, the EL procedure does not need to explicitly estimate the asymptotic covariance of the estimate of the regression coefficients, since an embedded studentization will be carried out in the optimization process. More importantly, under some regularity conditions, it has been already shown that in many statistical models, the EL confidence interval is Bartlett correctable, meaning that the coverage error can be reduced from $O(1/n)$ to $O(1/n^2)$ by a simple adjustment on the χ^2 -critical value used in the EL procedure, but in general, the bootstrap confidence intervals are not Bartlett correctable. The application of the EL procedure on the linear regression model was started with Owen (1991), then inspired by the Bartlett correction theory in DiCiccio et al. (1991), Chen (1993, 1994) further investigated the higher order property of the EL procedure in the linear regression models, and established the formula for the correction factor. Recent years also see the application of the EL methodology to other important statistical models where the data are either randomly censored, measured with errors, or missing. See Wang (2000), Wang and Rao (2002), Cui and Chen (2002), and the references therein.

Usually, we do not impose any restrictions on the regression coefficients in the linear regression models, they are free parameters. But in real applications, sometimes we have to build such a linear regression model in which the regression coefficients are subject to some restriction, often expressed by an equality or inequality. For example, in ANOVA or ANCOVA models, the fixed or interaction effects are often assumed to have sum 0 for the sake of model identifiability.

To be specific, the following linear regression model would be used to fit the data

$$Y = X'\beta + \varepsilon, \quad R\beta = b, \quad (1.1)$$

where Y is a scalar response, X is a $p \times 1$ observable design covariate, either fixed or random, β is the $p \times 1$ unknown regression coefficients, and the random error ε has mean 0 and variance σ^2 . The equality restriction on β is $R\beta = b$, where R is an $r \times p$ known constant matrix, $r \leq p$, and b is an $r \times 1$ known constant vector. In fact, without loss of generality, b is often assumed to be 0 by a reparametrization, but we will not do so since the reparametrization does not bring much simplification in the subsequent discussion.

The purpose of this paper is to develop confidence regions for β in model 1.1 using three different procedures, the classical Wald-type confidence region, the bootstrap confidence region and the empirical likelihood confidence region, with focus on the latter. The paper will be organized as follows. Section 2 provides the formula of Wald-type and the algorithms of bootstrap confidence intervals; the empirical likelihood confidence interval will be constructed in Section 4, together with the theory on Bartlett correction. Comparison studies on the finite sample performance among these procedures will be made through simulation studies in Section 5. Finally, the proofs of the main results will be presented in Section 6.

2 Wald-type and Bootstrap Confidence Regions

Let (X'_i, Y_i) , $i = 1, 2, \dots, n$ be a sample from (1.1), $EXX' > 0$, ε_i 's are independent and identically distributed. Denote \mathbf{X} the $n \times p$ matrix with i -th row being X'_i , $\mathbf{Y} = (Y_1, \dots, Y_n)$. Further assume that $\text{rank}(\mathbf{X}) = p$, $\text{rank}(R) = r$, and $\hat{\beta}_{LS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$, the LS estimate of β . It is well known that the restricted LS estimate of β has the form of

$$\hat{\beta}_{RLS} = \hat{\beta}_{LS} - (\mathbf{X}'\mathbf{X})^{-1}R'(R(\mathbf{X}'\mathbf{X})^{-1}R')^{-1}(R\hat{\beta}_{LS} - b).$$

If we assume that $\varepsilon \sim N(0, \sigma^2)$, ε and X are independent when X is random, $T = \mathbf{X}'\mathbf{X} + R'R$, then one can show that given \mathbf{X} ,

$$\hat{\beta}_{RLS} \sim N_p(\beta_0, \sigma^2\Omega), \quad \Omega = T^{-1} - T^{-1}R'(RT^{-1}R')^{-1}RT^{-1}.$$

Although feasible, it would be a little tedious to construct a confidence region of β based on the above normality result, since the covariance matrix Ω is a singular $p \times p$ matrix, and $\text{rank}(\Omega) = p - r$. Instead, we will proceed by rewriting model (1.1) with restricted parameter space as a model without any restriction. For this purpose, let $R = [R_1, R_2]$, where R_1 is an $r \times r$ and R_2 is an $r \times (p - r)$ matrix. Without loss of generality, assume that R_1 is nonsingular. Accordingly, decompose β into two blocks $[\beta'_1, \beta'_2]'$ with β_1 being $r \times 1$. Then from $R\beta = b$ we obtain $R_1\beta_1 + R_2\beta_2 = b$ and $\beta_1 = R_1^{-1}(b - R_2\beta_2)$. Thus, model (1.1) can be written as

$$Y_i = X'_i \begin{pmatrix} R_1^{-1}b - R_1^{-1}R_2\beta_2 \\ \beta_2 \end{pmatrix} + e_i = X'_i \begin{pmatrix} R_1^{-1}b \\ 0 \end{pmatrix} + X'_i \begin{pmatrix} -R_1^{-1}R_2 \\ I \end{pmatrix} \beta_2 + e_i. \quad (2.1)$$

Define

$$A = \begin{pmatrix} R_1^{-1}b \\ 0 \end{pmatrix}, \quad B = \begin{pmatrix} -R_1^{-1}R_2 \\ I \end{pmatrix}, \quad \tilde{Y}_i = Y_i - X'_i A, \quad \tilde{X}'_i = X'_i B.$$

One can see that $(\tilde{X}'_i, \tilde{Y}_i)'$, $i = 1, 2, \dots, n$ are i.i.d. random vectors. With this definition, (2.1) can be concisely written as a typical linear regression model

$$\tilde{Y}_i = \tilde{X}'_i \beta_2 + \varepsilon_i. \quad (2.2)$$

Therefore, to construct a confidence region of β , we can construct one for β_2 based on model (2.2), then combine the relationship $\beta_1 = R_1^{-1}(b - R_2\beta_2)$, we can build a confidence region of β .

2.1 Wald-type Confidence Region

Note that $\text{rank}(\mathbf{X}B) = p - r$, so $B'\mathbf{X}'\mathbf{X}B$ is nonsingular, and the least square estimate of β_2 in model (2.2) can be written as

$$\hat{\beta}_2 = (B'\mathbf{X}'\mathbf{X}B)^{-1}B'\mathbf{X}'(\mathbf{Y} - \mathbf{X}A).$$

If ε_i i.i.d. $\sim N(0, \sigma^2)$, then it is easily seen that, given \mathbf{X} ,

$$\hat{\beta}_2 \sim N(\beta_2, \sigma^2(B' \mathbf{X}' \mathbf{X} B)^{-1}).$$

Denote $\hat{\beta}_1 = R_1^{-1}(b - R_2 \hat{\beta}_2)$, and $\hat{\beta} = (\hat{\beta}_1', \hat{\beta}_2')$, an unbiased estimate of σ^2 can be taken as

$$\hat{\sigma}^2 = \frac{(\mathbf{Y} - \mathbf{X}\hat{\beta})'(\mathbf{Y} - \mathbf{X}\hat{\beta})}{n - (p - r)}, \quad (2.3)$$

which, together with the facts that $[n - (p - r)]\hat{\sigma}^2/\sigma^2 \sim \chi_{n-(p-r)}^2$ and $\hat{\beta}$ is independent of $\hat{\sigma}^2$, imply

$$\frac{(\hat{\beta}_2 - \beta_2)' B' \mathbf{X}' \mathbf{X} B (\hat{\beta}_2 - \beta_2)}{(p - r)\hat{\sigma}^2} \sim F_{p-r, n-(p-r)}.$$

Therefore, an F -type confidence region of β_2 with confidence level $1 - \alpha$ can be constructed as

$$C_{F, \beta_2} = \left\{ \beta_2 : \frac{(\hat{\beta}_2 - \beta_2)' B' \mathbf{X}' \mathbf{X} B (\hat{\beta}_2 - \beta_2)}{(p - r)\hat{\sigma}^2} \leq F_{1-\alpha, p-r, n-(p-r)} \right\},$$

which is clearly an ellipsoid in \mathbb{R}^{p-r} . Accordingly, a confidence region of β with confidence level $1 - \alpha$ can be defined as

$$C_{F, \beta} = \{ \beta = (\beta_1', \beta_2')' : \beta_1 = R^{-1}(b - R_2 \beta_2), \beta_2 \in C_{F, \beta_2} \}.$$

If ε is not normally distributed, then one can find a confidence interval of β either by large sample theory or by bootstrap algorithm. Under some mild regularity conditions, central limit theorem implies that

$$(B' \mathbf{X}' \mathbf{X} B)^{1/2}(\hat{\beta}_2 - \beta_2) \implies N(0, \sigma^2 I)$$

and law of large number implies that $\hat{\sigma}^2$ defined in (2.3) is a consistent estimator of σ^2 . Therefore,

$$\frac{(\hat{\beta}_2 - \beta_2)' B' \mathbf{X}' \mathbf{X} B (\hat{\beta}_2 - \beta_2)}{\hat{\sigma}^2} \implies \chi_{p-r}^2.$$

Hence, a χ^2 -type confidence region of β_2 can be constructed as

$$C_{\chi^2, \beta_2} = \left\{ \beta_2 : \frac{(\hat{\beta}_2 - \beta_2)' B' \mathbf{X}' \mathbf{X} B (\hat{\beta}_2 - \beta_2)}{\hat{\sigma}^2} \leq \chi_{1-\alpha, p-r}^2 \right\},$$

which is also an ellipsoid in \mathbb{R}^{p-r} . Accordingly, a confidence region of β with confidence level $1 - \alpha$ can be defined as

$$C_{\chi^2, \beta} = \{ \beta = (\beta_1', \beta_2')' : \beta_1 = R^{-1}(b - R_2 \beta_2), \beta_2 \in C_{\chi^2, \beta_2} \}.$$

2.2 Bootstrap Confidence Region

Generally there are two bootstrap algorithms in regression models. The case resampling treats the predictors as random, and model based resampling views the predictor as fixed. Sometimes, the predictor will be treated as fixed even if they are generated randomly. An attractive advantage of doing so is that the standard error of the bootstrap estimate of the regression coefficients reflects the precision associated with the sample of predictors actually observed. See Efron and Tibshirani (1993) for more discussion on this issue. The case bootstrap procedure for constructing the confidence interval of β consists of the following steps.

- (1) Estimate β_2 from the sample $(\tilde{X}'_i, \tilde{Y}'_i)'$, $i = 1, 2, \dots, n$, using least square procedure, denote the estimate as $\hat{\beta}_2$, and calculate the fitted value and residual for each observation $\hat{Y}_i = \tilde{X}'_i \hat{\beta}_2$, $e_i = \tilde{Y}_i - \hat{Y}_i$.
- (2) Select b bootstrap samples of the residuals, $e^*_{j,i}$, $i = 1, 2, \dots, n$, and calculate the bootstrapped Y -values using $Y^*_{j,i} = \hat{Y}_i + e^*_{j,i}$, $j = 1, 2, \dots, b$.
- (3) For each bootstrap sample, regress $Y^*_{j,i}$ on \tilde{X}_i , $i = 1, 2, \dots, n$ to obtain the bootstrap estimates of $\hat{\beta}^*_{j,2}$.
- (4) For each j , calculate

$$T_j^* = \frac{(\hat{\beta}^*_{j,2} - \hat{\beta}_2)' B' \mathbf{X}' \mathbf{X} B (\hat{\beta}^*_{j,2} - \hat{\beta}_2)}{\hat{\sigma}_j^{2*}},$$

where

$$\hat{\sigma}_j^{2*} = \frac{\sum_{i=1}^n (Y^*_{j,i} - \tilde{X}_i \hat{\beta}^*_{j,2})' (Y^*_{j,i} - \tilde{X}_i \hat{\beta}^*_{j,2})}{n - (p - r)},$$

and the $100(1 - \alpha)$ th percentile, denoted by $T_{[(1-\alpha)b]}^*$, of T_j^* , $j = 1, 2, \dots, b$.

- (5) The $1 - \alpha$ confidence region of β_2 can be constructed as

$$C_{B,\beta_2} = \left\{ \beta_2 : \frac{(\hat{\beta}_2 - \beta_2)' B' \mathbf{X}' \mathbf{X} B (\hat{\beta}_2 - \beta_2)}{\hat{\sigma}^2} \leq T_{[(1-\alpha)b]}^* \right\},$$

which is an ellipsoid in \mathbb{R}^{p-r} . Accordingly, a bootstrap confidence region of β with confidence level $1 - \alpha$ can be defined as

$$C_{B,\beta} = \{ \beta = (\beta'_1, \beta'_2)' : \beta_1 = R^{-1}(b - R_2 \beta_2), \beta_2 \in C_{B,\beta_2} \}.$$

If the model-based bootstrap procedure is preferred, then steps (1), (2) and (3) should be modified as follows:

- (1) Estimate β_2 from the original sample $(\tilde{X}'_i, \tilde{Y}'_i)'$, $i = 1, 2, \dots, n$, using least square procedure, denote the estimate as $\hat{\beta}_2$.
- (2) Select b bootstrap samples from $(\tilde{X}'_i, \tilde{Y}'_i)'$, $i = 1, 2, \dots, n$, denote them as $(X^*_{j,i}, Y_{j,i})'$, $j = 1, 2, \dots, b$.
- (3) For each bootstrap sample, regress $Y^*_{j,i}$ on $X_{j,i}$, $i = 1, 2, \dots, n$, to obtain the bootstrap estimates of $\hat{\beta}^*_{j,2}$.

3 Empirical Likelihood Confidence Interval

The motivation, theoretical development and computational consideration of the empirical likelihood methodology can be found in Owen (2001). Denote $Z_i(\beta_2) = \tilde{X}_i(\tilde{Y}_i - \tilde{X}_i'\beta_2)$. The empirical likelihood ratio function of β_2 is defined as

$$\mathcal{R}_n(\beta_2) = \max \left\{ \prod_{i=1}^n n w_i : \sum_{i=1}^n w_i Z_i(\beta_2) = 0, w_i \geq 0, \sum_{i=1}^n w_i = 1 \right\}.$$

Based on the theory of empirical likelihood inference, the empirical likelihood confidence interval for β_2 with confidence level $1 - \alpha$ should have the form

$$C_{EL,\beta_2} = \{\beta_2 : -2 \log \mathcal{R}_n(\beta_2) \leq c_{1-\alpha}\},$$

where $c_{1-\alpha}$ is the $1 - \alpha$ percentile of the asymptotic distribution of $-2 \log \mathcal{R}_n(\beta_{20})$ assuming that β_{20} is the true value of the regression parameter β_2 . Therefore, the empirical likelihood confidence interval for β with confidence level $1 - \alpha$ can be defined as

$$C_{EL,\beta} = \{\beta = (\beta_1', \beta_2')' : \beta_1 = R_1^{-1}(b - R_2\beta_2), -2 \log \mathcal{R}_n(\beta_2) \leq c_{1-\alpha}\}. \quad (3.1)$$

The following result shows that asymptotically, the confidence region defined in (3.1) has the right coverage rate.

Theorem 3.1. Suppose $E\|X\|^2 + E\varepsilon^2 < \infty$, then

$$\lim_{n \rightarrow \infty} P(\beta_0 \in C_{EL,\beta}) = 1 - \alpha.$$

In the above discussion, we have assumed that R_1 , the matrix consisting of the first r columns of R , is nonsingular. It is likely that there are other r columns of R form a nonsingular matrix, we denote this matrix as Q_1 and the corresponding parameters as γ_1 , the rest $p - r$ columns of R as Q_2 , and the corresponding parameters as γ_2 , then we will have another linear model and can construct another empirical confidence region using exactly the same argument. This brings up an interesting question, are these two confidence regions same?

To be specific, let T be the elementary matrix which permute the locations of the elements in β so that the after the permutation, the first r elements forms the vector γ_1 , and the rest $p - r$ elements forms the vector γ_2 . Then the new linear regression model with restriction can be written as

$$Y_i = X_i' T' T \beta + \varepsilon_i, \quad R T' T \beta = b, \quad (3.2)$$

Thus

$$T \beta = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix}, \quad R T' = [Q_1, Q_2].$$

With these new notations, we have

$$\begin{aligned} Y_i &= X_i' T' \begin{pmatrix} Q_1^{-1} b - Q_1^{-1} Q_2 \gamma_2 \\ \gamma_2 \end{pmatrix} + e_i \\ &= X_i' T' \begin{pmatrix} Q_1^{-1} b \\ 0 \end{pmatrix} + X_i' T' \begin{pmatrix} -Q_1^{-1} Q_2 \\ I \end{pmatrix} \gamma_2 + e_i. \end{aligned} \quad (3.3)$$

Define

$$\hat{Y}_i = Y_i - X_i' T' \begin{pmatrix} Q_1^{-1} b \\ 0 \end{pmatrix}, \quad \hat{X}_i' = X_i' T' \begin{pmatrix} -Q_1^{-1} Q_2 \\ I \end{pmatrix},$$

then, (3.2) can be concisely written as a typical linear regression model

$$\hat{Y}_i = \hat{X}_i' \gamma_2 + \varepsilon_i. \quad (3.4)$$

Denote $W_i = \hat{X}_i' (\hat{Y}_i - \hat{X}_i' \gamma_2)$. The empirical likelihood ratio function of γ_2 is defined as

$$\mathcal{R}_n(\gamma_2) = \max \left\{ \prod_{i=1}^n n u_i : \sum_{i=1}^n u_i W_i(\beta_2) = 0, u_i \geq 0, \sum_{i=1}^n u_i = 1 \right\}.$$

and the empirical confidence interval for γ , or β with confidence level $1 - \alpha$ can be defined as

$$C_{EL,\gamma}(\gamma) = \{\gamma = (\gamma_1', \gamma_2')' : \gamma_1 = Q_1^{-1}(b - Q_2 \gamma_2), -2 \log \mathcal{R}_n(\gamma_2) \leq c_{1-\alpha}\}.$$

The following result claims that $C_{EL,\beta}$ and $C_{EL,\gamma}$ are just the permutation of each other. More precisely, we have

Theorem 3.2. Assume the conditions in Theorem 3.1 holds. Then

$$T(C_{EL,\beta}) = C_{EL,\gamma},$$

where $T(C_{EL,\beta})$ denotes $\{T\beta : \beta \in C_{EL,\beta}\}$.

Imposing some stronger conditions on the distribution of X and ε , we can derive an asymptotic expansion of the coverage probability of the proposed empirical confidence region. This expansion also makes it possible to increase the coverage accuracy by a Bartlett correction on the original empirical confidence region.

A simple computation shows that $\text{Cov}(\tilde{Z}_i) = \text{Cov}(\tilde{X}_i'(\tilde{Y}_i - \tilde{X}_i' \beta)) = \sigma^2 E X X' = \Sigma > 0$. Define $V_i = \Sigma^{-1/2} \tilde{Z}_i$, $i = 1, 2, \dots, n$. Then we have the following result

Theorem 3.3. Assume that $E|\varepsilon|^{15} + E\|X\|^{15} < \infty$, and the characteristic function $\psi(t)$ of $X_i \varepsilon_i$ satisfies the Cramér condition $\sup_{\|t\|>b} |\psi(t)| < 1$, then

$$P[\beta_2 \in C_{EL,\beta_2}] = 1 - \alpha - \frac{ac_{1-\alpha} g_{p-r}(c_{1-\alpha})}{n} + O(n^{-3/2}),$$

where g_{p-r} is the density function of χ_{p-r}^2 , and

$$a = \frac{1}{p-r} \left[\frac{1}{2} E(V_1' V_1)^2 - \frac{1}{3} E(V_1' V_2)^3 \right].$$

Theorem 3.3 implies that the true coverage rate and the nominal coverage rate differs in an order of $O(n^{-1})$. Similar arguments as in the proof of Theorem 3.1 in Chen (1993), we can obtain the following result

Corollary 3.1. Assume the conditions in Theorem 3.3 hold. Then

$$P(-2 \log \mathcal{R}(\beta_{20}) \leq (1 + \xi/n)\chi_{1-\alpha}^2) = 1 - \alpha + O(n^{-2}),$$

where ξ is either a or a \sqrt{n} -consistent estimator of a .

Now let $\hat{\beta}_{2n}$ be any \sqrt{n} -consistent estimate of β_2 , and let $\hat{Z}_i = \tilde{X}_i(\tilde{Y}_i - \tilde{X}_i\hat{\beta}_{2n})$ and $\hat{\Sigma}_n = S_n^2$, the sample covariance matrix of $\hat{Z}_i, i = 1, 2, \dots, n$, then it is easy to check that

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

is a \sqrt{n} -consistent estimator of a . Thus, a Bartlett confidence region of β_2 can be defined as

$$C_{BCEL, \beta_2} = \{\beta_2 : -2 \log \mathcal{R}(\beta_2) < c_{1-\alpha}(1 + \hat{a}/n)\},$$

and the corresponding Bartlett confidence region of β will be

$$C_{BCEL, \beta} = \{\beta = (\beta_1', \beta_2')' : \beta_1 = R_1^{-1}(b - R_2\beta_2), -2 \log \mathcal{R}(\beta_2) < c_{1-\alpha}(1 + \hat{a}/n)\}.$$

4 Simulation Studies

In this section, we will evaluate the performance of the proposed empirical likelihood procedure through a simulation study. The comparison will also be made among all the procedures discussed in the previous sections. The linear model used in the simulation is $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$, with true parameter values $\beta_0 = 1, \beta_1 = 2\beta_2$ and $\beta_2 = 1$, the restriction imposed on the regression coefficients is $\beta_1 = 2\beta_2$. The random design variable X follows standard normal distribution, while the random error ε is chosen to have standard normal distribution and uniform distribution on $[-1, 1]$. Two sample sizes, $n = 100, 200$, will be used in the simulation. The empirical coverage rates based on 500 replications of each simulation will be report to see if the proposed confidence regions maintain the nominal confidence level 0.95, and a plot will display the volumes of each confidence regions.

If ε follows the standard normal distribution, the coverage rate of the F -Type confidence region is exactly the nominal confidence level. Table 1 reports the empirical coverage rates from all the procedures when $\varepsilon \sim N(0, 1)$. It is not surprising to notice that the empirical coverage rates from the F -Type procedure for both sample sizes are very close to 0.95. The empirical coverage rates from empirical likelihood and Bartlett corrected empirical likelihood procedures are less than the nominal level when the sample size is small, but they approach to the nominal level when sample size gets bigger. Other methods also work well.

The simulation results for $\varepsilon \sim U(-1, 1)$ are reported in Table 2. One may expect that the F -Type confidence region have a poor coverage rate. To our surprise, the F -Type confidence region

Table 1: $\varepsilon \sim N(0, 1)$

n	F-Type	χ^2 -Type	EL	BCEL	Boot(data)	Boot(residual)
100	0.940	0.930	0.922	0.928	0.948	0.932
200	0.956	0.950	0.950	0.952	0.956	0.952

maintains the nominal level very well. Similar for the χ^2 -Type, the empirical likelihood and the Bartlett corrected empirical likelihood procedures, but the empirical coverage rates seem bigger than the nominal coverage rate. Both bootstrap procedures work very well.

Table 2: $\varepsilon \sim U(-1, 1)$

n	F-Type	χ^2 -Type	EL	BCEL	Boot(data)	Boot(residual)
100	0.940	0.946	0.948	0.954	0.948	0.946
200	0.956	0.952	0.958	0.960	0.954	0.948

5 Proof of Main Results

For the sake of brevity, $Z_i(\beta_2)$ will be denoted as Z_i .

Proof of Theorem 3.1: A straightforward argument based on Lagrangian multiplier shows that

$$\mathcal{R}(\beta_2) = \prod_{i=1}^n \frac{1}{1 + \lambda' Z_i}, \tag{5.1}$$

and λ is uniquely determined by

$$\sum_{i=1}^n \frac{Z_i}{n[1 + \lambda' Z_i]} = 0. \tag{5.2}$$

In the following, we would like to argue that $\lambda = O_p(n^{-1/2})$. For this purpose, let $\bar{Z}_n = \sum_{i=1}^n Z_i$, $Z^* = \max_{1 \leq i \leq n} \|Z_i\|$, and $\lambda = \rho\theta$ with $\rho > 0$, $\|\theta\| = 1$. Here $\|\cdot\|$ denotes the Euclidean norm of a vector. From (5.1), we have

$$0 = \frac{1}{n} \theta' \sum_{i=1}^n \frac{Z_i}{1 + \rho\theta' Z_i} = \frac{1}{n} \left[\theta' \sum_{i=1}^n Z_i - \rho\theta' \sum_{i=1}^n \frac{Z_i Z_i' \theta'}{1 + \rho\theta' Z_i} \right] = \theta' \bar{Z} - \rho\theta' \tilde{V}_n \theta, \tag{5.3}$$

where $\tilde{V}_n = n^{-1} \sum_{i=1}^n Z_i' Z_i / (1 + \rho\theta' Z_i)$. If we denote $V_n = n^{-1} \sum_{i=1}^n Z_i' Z_i$, then from $0 \leq 1 + \rho\theta' Z_i \leq 1 + \rho \max_{1 \leq i \leq n} \|Z_i\|$, we have

$$\rho\theta' V_n \theta = \rho\theta' \left(\frac{1}{n} \sum_{i=1}^n Z_i' Z_i \right) \theta = \rho\theta' \left(\frac{1}{n} \sum_{i=1}^n \frac{Z_i' Z_i}{1 + \rho\theta' Z_i} (1 + \rho\theta' Z_i) \right) \theta \leq \rho\theta' \tilde{V}_n \theta [1 + \rho \max_{1 \leq i \leq n} \|Z_i\|].$$

By (5.3), we have $\rho\theta'V_n\theta \leq \theta'\bar{Z}_n[1 + \rho]$, and

$$\rho[\theta'V_n\theta - \theta'\bar{Z}_n \max_{1 \leq i \leq n} \|Z_i\|] \leq \theta'\bar{Z}_n.$$

Denote η_1 and η_{p-r} are the largest and the smallest eigenvalues of $\text{Cov}(Z)$, and

$$\Sigma_0 = E(XX'), \quad B = \begin{pmatrix} -R_1^{-1}R_2 \\ I \end{pmatrix}.$$

Note that $\Sigma_0 > 0$, $\text{rank}(B) = p - r$, so $\text{Cov}(Z) = \sigma^2 B' \Sigma_0 B > 0$, therefore, $\eta_1 \geq \eta_{p-r} > 0$. Law of large numbers implies that $\eta_{p-r} + o_p(1) \leq \theta'V_n\theta \leq \eta_1 + o_p(1)$. Central limit theorem implies that $\sqrt{n}\bar{Z}_n \implies N(0, \text{Cov}(Z))$. The finiteness of $E\|Z\|^2$ implies that $\max_{1 \leq i \leq n} \|Z_i\| = o_p(n^{1/2})$. It follows that

$$\rho[\theta'V_n\theta + o_p(1)] = O_p(n^{-1/2})$$

and $\rho = O_p(n^{-1/2})$. This also implies

$$\max_{1 \leq i \leq n} |\rho\theta'Z_i| \leq \rho \max_{1 \leq i \leq n} \|Z_i\| = O_p(n^{-1/2})o_p(n^{1/2}) = o_p(1). \quad (5.4)$$

Note that

$$\begin{aligned} 0 &= \frac{1}{n} \sum_{i=1}^n \frac{Z_i}{1 + \lambda'Z_i} = \frac{1}{n} \sum_{i=1}^n Z_i \left[1 - \lambda'Z_i + \frac{\lambda'Z_i Z_i' \lambda}{1 + \lambda'Z_i} \right] \\ &= \bar{Z}_n - V_n \lambda + \frac{1}{n} \sum_{i=1}^n \frac{(\lambda'Z_i)^2 Z_i}{1 + \lambda'Z_i}. \end{aligned} \quad (5.5)$$

Note that $E|\varepsilon|^2 < \infty$, $E\|X\|^2 < \infty$ implies

$$\frac{1}{n} \sum_{i=1}^n \|\tilde{X}_i(\tilde{Y}_i - \tilde{X}_i' \beta_2)\|^3 = o_p(n^{1/2}),$$

the third term is bounded by

$$\frac{\|\lambda\|^2}{n} \sum_{i=1}^n \|\tilde{X}_i(\tilde{Y}_i - \tilde{X}_i' \beta_2)\|^3 [1 + \lambda' \tilde{X}_i(\tilde{Y}_i - \tilde{X}_i' \beta_2)]^{-1} = O_p(n^{-1})o_p(n^{1/2})O_p(1) = o_p(n^{-1/2}).$$

Therefore, we have

$$\lambda = V_n^{-1} \bar{Z}_n + \delta_n, \quad (5.6)$$

and $\delta_n = o_p(n^{-1/2})$. By (5.4), we have the following Taylor expansion

$$\log[1 + \lambda'Z_i] = \lambda'Z_i - \frac{1}{2}(\lambda'Z_i)^2 + \mu_i$$

where, for some $c > 0$, $P(|\mu_i| \leq c\|\lambda'Z_i\|^3, 1 \leq i \leq n) \rightarrow 1$ as $n \rightarrow \infty$. From (5.6), we can write

$$\begin{aligned} -2 \log \mathcal{R}(\beta_2) &= -2 \log \prod_{i=1}^n n w_i = 2 \sum_{i=1}^n \log[1 + \lambda'Z_i] \\ &= 2n\lambda'\bar{Z}_n - n\lambda'V_n\lambda + 2 \sum_{i=1}^n \mu_i \\ &= 2n[V_n^{-1}\bar{Z}_n + \delta_n]' \bar{Z}_n - n[V_n^{-1}\bar{Z}_n + \delta_n]' V_n [V_n^{-1}\bar{Z}_n + \delta_n] + 2 \sum_{i=1}^n \mu_i \\ &= n\bar{Z}_n' V_n^{-1} \bar{Z}_n - n\delta_n' V_n^{-1} \delta_n + 2 \sum_{i=1}^n \mu_i. \end{aligned}$$

Central limit theorem and Slutsky theorem imply that $n\bar{Z}_n' V_n^{-1} \bar{Z}_n \Rightarrow \chi_{p-r}^2$ in distribution, and the fact $\delta_n = o_p(n^{-1/2})$ implies that $n\delta_n' V_n^{-1} \delta_n = o_p(1)$. Also

$$\left| \sum_{i=1}^n \mu_i \right| = c \sum_{i=1}^n |\lambda'Z_i|^3 \leq c\|\lambda\|^3 \cdot \sum_{i=1}^n \|Z_i\|^2 \cdot \max_{1 \leq i \leq n} \|Z_i\| = O_p(n^{-3/2})O_p(n)o_p(n^{1/2}) = o_p(1).$$

Finally, we obtain $-2 \log \mathcal{R}(\beta_2) \Rightarrow \chi_{p-r}^2$. \square

Proof of Theorem 3.2: For any $\beta \in C_{EL,\beta}$, let $\gamma = T\beta$, we will show that $\gamma \in C_{n,1-\alpha}(\gamma)$.

Note that $\beta_1 = R_1^{-1}(b - R_2\beta_2)$, equivalently, $R\beta = b$, so $RT'T\beta = Q\gamma = b$, which is equivalent to $\gamma_1 = Q_1^{-1}(b - Q_2\gamma_2)$.

Let

$$C = T' \begin{pmatrix} -Q_1^{-1}Q_2 \\ I \end{pmatrix}, \quad B = \begin{pmatrix} -R_1^{-1}R_2 \\ I \end{pmatrix}.$$

It is easily seen that C and B are full rank, that is, $\text{rank}(C) = \text{rank}(B) = p - r$. Now suppose β_2 satisfies $-2 \log \mathcal{R}(\beta_2) < c_{1-\alpha}$. Then there exists a λ such that

$$-2 \log \mathcal{R}(\beta_2) = -2 \log \prod_{i=1}^n \frac{1}{1 + \lambda'Z_i} = -2 \log \prod_{i=1}^n \frac{1}{1 + (B\lambda)'X_i\varepsilon_i} < c_{1-\alpha}.$$

and λ is a solution to the following equation

$$\sum_{i=1}^n \frac{Z_i}{n[1 + \lambda'Z_i]} = 0 \quad \iff \quad \sum_{i=1}^n \frac{B'X_i\varepsilon_i}{n[1 + (B\lambda)'X_i\varepsilon_i]} = 0.$$

Since C and B are full rank, so for the above λ , there exist a vector τ and a matrix G such that $C\tau = B\lambda$, and $C = BG$. For such a τ and G , we have

$$\begin{aligned} \sum_{i=1}^n \frac{W_i}{n[1 + \tau'W_i]} &= \sum_{i=1}^n \frac{C'X_i\varepsilon_i}{n[1 + (C\tau)'X_i\varepsilon_i]} = \sum_{i=1}^n \frac{G'B'X_i\varepsilon_i}{n[1 + (B\lambda)'X_i\varepsilon_i]} \\ &= G' \sum_{i=1}^n \frac{B'X_i\varepsilon_i}{n[1 + (B\lambda)'X_i\varepsilon_i]} = 0 \end{aligned}$$

and

$$\begin{aligned} -2 \log \mathcal{R}(\gamma_2) &= -2 \log \prod_{i=1}^n \frac{1}{1 + \tau' W_i} = -2 \log \prod_{i=1}^n \frac{1}{1 + (C\tau)' X_i \varepsilon_i} \\ &= -2 \log \prod_{i=1}^n \frac{1}{1 + (B\lambda)' X_i \varepsilon_i} = -2 \log \mathcal{R}(\beta_2) < c_{1-\alpha}. \end{aligned}$$

Therefore, we have shown that $\gamma \in C_{n,1-\alpha}(\gamma)$ or $TC_{n,1-\alpha}(\beta) \subset C_{n,1-\alpha}(\gamma)$.

Using similar arguments, we can also show that $TC_{n,1-\alpha}(\beta) \supset C_{n,1-\alpha}(\gamma)$. This completes the proof. \square

Proof of Theorem 3.3: Note that $\text{Cov}(\tilde{X}_i(\tilde{Y}_i - \tilde{X}_i' \beta)) = \sigma^2 E X X' = \Sigma > 0$, and $V_i = \Sigma^{-1/2} \tilde{Z}_i$, we have $E(V_i) = 0$ and $\text{Cov}(V_i) = I$. It is easy to see that

$$\mathcal{R}(\beta_2) = \max \left\{ \prod_{i=1}^n w_i V_i : \sum_{i=1}^n w_i V_i = 0, w_i \geq 0, \sum_{i=1}^n w_i = 1 \right\}.$$

Denote V_{ij} the j -th component of V_i . Under the moment condition in the theorem, we have the following Taylor expansion for $-2 \log \mathcal{R}(\beta_2)$,

$$\begin{aligned} -2 \log \mathcal{R}(\beta_2) &= A^j A^j - A^{jk} A^j A^k + \left(\frac{2}{3} \bar{\alpha}^{jkl} + \frac{2}{3} A^{jkl} - 2 \bar{\alpha}^{jkm} A^{lm} \right) A^j A^k A^l \\ &\quad + (\bar{\alpha}^{jkq} \bar{\alpha}^{lmq} - \frac{1}{2} \bar{\alpha}^{jklm}) A^j A^k A^l A^m + A^{jl} A^{kl} A^j A^k + O_p(n^{-5/2}), \end{aligned}$$

where

$$\bar{\alpha}^{j_1 \dots j_k} = \frac{1}{n} \sum_{i=1}^n E[V_{ij_1} \dots V_{ij_k}], \quad A^{j_1 \dots j_k} = \frac{1}{n} \sum_{i=1}^n [V_{ij_1} \dots V_{ij_k} - \bar{\alpha}^{j_1 \dots j_k}].$$

Here we use the convention that terms with repeated super indices are to be summed over. We further denote $R = R_1 + R_2 + R_3$, the j -th components of R_1 , R_2 and R_3 are defined as

$$R_{1j} = A^j, \quad R_{2j} = -\frac{1}{2} A_{jk} A^k + \frac{1}{3} \bar{\alpha}^{jkm} A^k A^m,$$

and

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

respectively. Then we can show that $R_j = O_p(n^{-j/2})$, $j = 1, 2, 3$, and

$$-2 \log \mathcal{R}(\beta_2) = -(\sqrt{n}R)'(\sqrt{n}R) + O_p(n^{-5/2}).$$

Applying Theorem 20.6 of Bhattacharya and Rao (1976) and Theorem 3.2, Remarks 3.3 and 3.4 of Skovgaard (1981), and using a similar argument as in the proof of Theorem 2.1 of Chen (1993), we can develop an Edgeworth expansion for the distribution function of $\sqrt{n}R$, which in turn enables us to obtain the probability statement in Theorem 3.3. \square

References

- Bhattacharya, R. and Rao, R. (1976), "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society Series B*.
- Chen, S. (1993), "On the accuracy of empirical likelihood confidence regions for linear regression model," *Annals of the Institute of Statistical Mathematics*, 621–637.
- (1994), "Empirical likelihood confidence intervals for linear regression coefficients," *Journal of Multivariate Analysis*, 24–40.
- Cui, H. and Chen, S. (2002), "Empirical likelihood confidence regions for parameters in errors-in-variables models," *Journal of Multivariate Analysis*, 101–115.
- DiCiccio, T., Hall, P., and Romano, J. (1991), "Empirical likelihood is bartlett-correctable," *Ann. Statist.*, 1053–1061.
- Efron, B. and Tibshirani, R. (1993), "An Introduction to the Bootstrap," .
- Owen, A. (1988), "Empirical likelihood ratio confidence intervals for a single functional," *Biometrika*, 237–249.
- (1990), "Empirical likelihood confidence regions," *Ann. Statist.*, 90–120.
- (1991), "Empirical likelihood for linear models," *Ann. Statist.*, 1725–1747.
- (2001), "Empirical likelihood," .
- Skovgaard, I. (1981), "Transformation of an Edgeworth expansion by a sequence of smooth functions," *Scand. J. Statist.*, 207–217.
- Wang, Q. (2000), "Estimation of linear error-in-covariables models with validation data under random censorship," *Journal of Multivariate Analysis*, 245–266.
- Wang, Q. and Rao, J. (2002), "Empirical likelihood-based inference in linear models with missing data," *Scand J. Statist.*, 563–576.

Received: December 9, 2018

Accepted: February 24, 2019