Communications in Statistics - Simulation and Computation

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/lssp20

The information domain confidence intervals in univariate linear calibration

Guimei Zhao & Xingzhong Xu

To cite this article: Guimei Zhao & Xingzhong Xu (2020): The information domain confidence intervals in univariate linear calibration, Communications in Statistics - Simulation and Computation, DOI: 10.1080/03610918.2020.1777302

To link to this article: https://doi.org/10.1080/03610918.2020.1777302

Published online: 07 Jul 2020.

Submit your article to this journal

Article views: 7

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=lssp20

COMMUNICATIONS IN STATISTICS - SIMULATION AND COMPUTATION®
https://doi.org/10.1080/03610918.2020.1777302

Taylor & Francis
Taylor & Francis Group

Check for updates

The information domain confidence intervals in univariate linear calibration

Guimei Zhao[a] and Xingzhong Xu[b,c]

[a]Department of Statistics, North China University of Technology, Beijing, China; [b]School of Mathematics and Statistics, Beijing Institute of Technology, Beijing, China; [c]Beijing Key Laboratory on MCAACI, Beijing Institute of Technology, Beijing, China

ABSTRACT

We consider the confidence interval for the univariate linear calibration, where a response variable is related to an explanatory variable by a simple linear model, and the observations of the response variable and known values of the explanatory variable are used to make inferences on a single unknown value of the explanatory variable. Since the univariate linear calibration suffers from a problem of local unidentifiability, which results in the confidence coefficient of every confidence interval with finite length being zero, we propose new confidence intervals in terms of information domain, which are verified to be $1-\alpha$ confidence intervals for a specified range of the interesting parameter. The proposed intervals are numerically compared with two existing methods, and simulations show that our confidence intervals have good behavior in the coverage probability and the expected length. We also illustrate the results using an example.

ARTICLE HISTORY
Received 4 July 2019
Accepted 28 May 2020

KEYWORDS
Coverage probability; Expected length; The information domain confidence interval; Univariate linear calibration

1. Introduction

This paper is concerned with the confidence interval for the univariate linear calibration. Assume that there are a response variable and an explanatory variable, which is nonrandom. Let $y_1, y_2, \dots, y_n$ be n independent observations of the response variable corresponding to the values $x_1, x_2, \dots, x_n$ of the explanatory variable, respectively. The univariate linear calibration model is

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, i = 1, 2, \dots, n, \tag{1}$$

where $\varepsilon_i \sim N(0, \sigma^2), i = 1, 2, \dots, n$ are independent, and β_0, β_1 and σ^2 are unknown parameters. Further, let y_0 be an observation of the response variable corresponding to an unknown value θ of the explanatory variable. Suppose that

$$y_0 = \beta_0 + \beta_1 \theta + \varepsilon_0, \tag{2}$$

with y_0 and ε_0 being independent of y_i and $\varepsilon_i, i = 1, 2, \dots, n,$ respectively.

The calibration problem is to make inferences on the interesting parameter θ. A classical solution to the univariate calibration was derived by Eisenhart (1939). Since then there have been many papers devoted to this problem. Krutchkoff (1967) proposed the inverse regression procedure. Kang, Koo, and Roh (2017) introduced reversed inverse regression. Muhammad, Riaz, and Dawood (2019) proposed an estimator which make both the classical and the inverse estimators become its special cases. The Bayesian method was used to the univariate calibration by

CONTACT Xingzhong Xu xuxz@bit.edu.cn School of Mathematics and Statistics, Beijing Institute of Technology, Beijing 100081, China.

© 2020 Taylor & Francis Group, LLC

Hoadley (1970), Hunter and Lamboy (1981), Kubokawa and Robert (1994) and Zhao and Xu (2017a), etc. Dunne (1995), Krishnamoorthy, Kulkarni, and Mathew (2001) and Zhao and Xu (2017b) provided solutions to testing problems for univariate calibration. Dahiya and Mckeon (1991) developed a method to obtain the confidence interval for parameter θ based on Naszodi's (1978) estimation. Schechtman and Spiegelman (2002) proposed a nonlinear approach to the linear calibration interval. Bonate (1993), Jones and Rocke (1999) and Ng and Pooi (2008) explored the use of the bootstrap in linear calibration. A likelihood-based approach was developed by Bellio (2003). More details on the calibration problem see Muhammad and Riaz (2016); Marciano, Blas Achic, and Cysneiros (2016); Lin and Huang (2010); Chvosteková(2019), etc.

The univariate linear calibration problem can be reduced to making inference on the ratio of two normal means, known as Fieller's problem. It is challenging since there is a problem of local unidentifiability, referred to as G-H problem. The G-H problem has been discussed in detail by Gleser and Hwang (1987) and Liseo (2003), and which we will illustrate in Sec. 2. For a G-H problem, every confidence interval with a given confidence level $1-\alpha$ $(0<\alpha<1)$ satisfies that the probability of having an unbounded confidence interval is greater than zero, that is, the confidence coefficient for the confidence interval with finite length is zero. Hence, no frequentist procedures so far can provide completely satisfactory solutions. In this paper, we shall propose the information domain confidence intervals, which have the given confidence level on the information domain. Also, a comparison of our proposed intervals to two existing approaches is presented by simulation.

The paper is organized as follows. In Sec. 2, we give the information domain confidence intervals for the interesting parameter θ, and demonstrate theoretically the coverage probability behavior. In Sec. 3, the proposed confidence intervals are numerically compared with two existing methods. An example appears in Sec. 4. Some conclusions are given in Sec. 5.

2. The information domain confidence intervals

The G-H problem (Gleser and Hwang 1987) is illustrated as follows. The parameter space is denoted by $\Omega_1\times\Omega_2$. Let $\xi=(\omega,\eta)\in\Omega_1\times\Omega_2$ and $\gamma(\omega)$ be a function of ω. Assume that there exists a subset $\Omega_1^*\subset\Omega_1$ and a point $\eta^*\in\bar{\Omega}_2$ (the closure of Ω_2), such that $\gamma(\omega)$ is unbounded over Ω_1^*, and for each $\omega\in\Omega_1^*$ and any sample z from the sampling distribution $p(z|\xi)$, it holds that

$$\lim_{\eta\to\eta^*} p(z|\xi)=p(z|\eta^*),$$

with $p(z|\eta^*)$ being independent of ω. Then, every confidence interval $(\hat{\gamma}_L,\hat{\gamma}_U)$ for $\gamma(\omega)$ with finite length satisfies

$$\inf_{\xi\in\Omega_1\times\Omega_2} P_\xi(\hat{\gamma}_L\le\gamma\le\hat{\gamma}_U)=0.$$

That is, the confidence coefficient for $(\hat{\gamma}_L,\hat{\gamma}_U)$ is zero.

It can be seen that, for the univariate linear calibration, no matter what the confidence level $1-\alpha$, no confidence intervals with finite length exist for θ by taking $\omega=\gamma(\omega)=\theta,\eta=(\beta_0,\beta_1,\sigma^2),\Omega_1^*=(k,+\infty)$ for some constant k, and $\eta^*=(\beta_0,0,\sigma^2)$. Namely, if the interesting parameter θ has an unbounded range, and as β_1 tends to zero, the parameter θ is unidentifiable. Thus, the traditional frequency property of confidence intervals is not applicable for univariate linear calibration.

Note that, in carefully designed experiments, the interesting parameter θ will be within the range of $x_i, i=1,2,\dots,n$ (see Benton, Krishnamoorthy, and Mathew 2003). Therefore, for a confidence interval for θ, it is enough to have confidence level $1-\alpha$ for all $\theta\in\Omega_{1B_n}\subset\Omega_1$, where the set Ω_{1B_n} is chosen to contain the values $x_i, i=1,2,\dots,n$, referred to as the information domain. It is required that the confidence interval $(\hat{\theta}_L, \hat{\theta}_U)$, named the information domain confidence interval, satisfy

$$\inf_{(\theta,\eta)\in\Omega_{B_n}} P_{(\theta,\eta)}(\hat{\theta}_L \le \theta \le \hat{\theta}_U) \ge 1-\alpha,$$

with $\Omega_{B_n} = \{(\theta,\eta) : \theta \in \Omega_{1B_n}, \eta \in \Omega_2\}$.

Now, we consider the information domain confidence interval on the interesting parameter θ in model (2), based on $(y_i, x_i), i = 1, 2, ..., n$, and y_0, in models (1) and (2), respectively. Let $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$, and $\bar{y} = \frac{1}{n}\sum_{i=1}^n y_i$. Denote by $\hat{\beta}_0$ and $\hat{\beta}_1$ the least squares estimators of β_0 and β_1, that is, $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$ and

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

Define $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ and $\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^n (y_i - \hat{y}_i)^2$.

As mentioned above, the range of the $x_i, i = 1, 2, ..., n$ will be designed to cover the interesting parameter θ. At least we can assume that $|\theta - \bar{x}| \le 3\sqrt{\frac{\sum_{i=1}^n (x_i-\bar{x})^2}{n-1}}$ (see Benton, Krishnamoorthy, and Mathew 2003). Since $\frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2$ generally tends to a constant when $n \to \infty$, we define the information domain as $\Omega_{1B_n} = \{\theta : |\theta - \bar{x}| \le B_n\}$, with $B_n = o(n^{1/2})$. Let $y = (y_0, y_1, y_2, ..., y_n)$. The information domain confidence interval is constructed as

$$C(y) = \left\{\theta : |y_0 - \bar{y} - \hat{\beta}_1(\theta - \bar{x})| \le t_{1-\alpha/2}(n-2)\hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{B_n^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}\right\},$$

where $t_{1-\alpha/2}(n-2)$ denotes the $(1-\alpha/2)th$ quantile of the t-distribution with n – 2 degrees of freedom. It can be verified that $C(y)$ is a $1-\alpha$ confidence interval when $|\theta - \bar{x}| \le B_n$, and for all θ the coverage probability of the confidence interval is approximately equal to the given level $1 - \alpha$ when the sample size n is large enough.

Theorem Suppose that $\bar{x} \to c_1$ and $\frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2 \to c_2 > 0$ as $n \to \infty$, with c_1 and c_2 being two constants. The information domain confidence interval satisfies

$$\inf_{|\theta-\bar{x}|\le B_n} P(\theta \in C(Y)) \ge 1-\alpha,$$

and for all θ,

$$\lim_{n\to\infty} P(\theta \in C(Y)) = 1-\alpha.$$

Proof. A ccording to models (1) and (2), $\bar{y}, \hat{\beta}_1$, y_0 and $\hat{\sigma}^2$ are mutually independently distributed with

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x}_1)^2}\right), y_0 - \bar{y} \sim N\left(\beta_1(\theta - \bar{x}), \left(1 + \frac{1}{n}\right)\sigma^2\right), \frac{(n-2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-2).$$

Then we have

$$\frac{y_0 - \bar{y} - \hat{\beta}_1(\theta - \bar{x})}{\hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(\bar{x}-\theta)^2}{\sum_{i=1}^n (x_i-\bar{x})^2}}} \sim t(n-2).$$

Denote by $F(\cdot)$ the cumulative distribution function of the t-distribution, it holds that

$$P\left(|y_0-\bar{y}-\hat{\beta}_1(\theta-\bar{x})| \le t_{1-\alpha/2}(n-2)\hat{\sigma}\sqrt{1+\frac{1}{n}+\frac{B_n^2}{\sum_{i=1}^n (x_i-\bar{x})^2}}\right) \tag{3}$$

$$= P\left(\frac{|y_0-\bar{y}-\hat{\beta}_1(\theta-\bar{x})|}{\hat{\sigma}\sqrt{1+\frac{1}{n}+\frac{(\bar{x}-\theta)^2}{\sum_{i=1}^n (x_i-\bar{x})^2}}} \le t_{1-\alpha/2}(n-2)\frac{\sqrt{1+\frac{1}{n}+B_n^2/\sum_{i=1}^n (x_i-\bar{x})^2}}{\sqrt{1+\frac{1}{n}+\frac{(\bar{x}-\theta)^2}{\sum_{i=1}^n (x_i-\bar{x})^2}}}\right) \tag{4}$$

$$= 2F\left(t_{1-\alpha/2}(n-2)\frac{\sqrt{1+\frac{1}{n}+B_n^2/\sum_{i=1}^n (x_i-\bar{x})^2}}{\sqrt{1+\frac{1}{n}+\frac{(\bar{x}-\theta)^2}{\sum_{i=1}^n (x_i-\bar{x})^2}}}\right)-1. \tag{5}$$

For $|\theta-\bar{x}| \le B_n$, it follows from above expressions that

$$\inf_{|\theta-\bar{x}|\le B_n} P\left(|y_0-\bar{y}-\hat{\beta}_1(\theta-\bar{x})| \le t_{1-\alpha/2}(n-2)\hat{\sigma}\sqrt{1+\frac{1}{n}+\frac{B_n^2}{\sum_{i=1}^n (x_i-\bar{x})^2}}\right)$$
$$\ge 2F\big(t_{1-\alpha/2}(n-2)\big)-1 = 1-\alpha.$$

Since $B_n^2/\sum_{i=1}^n (x_i-\bar{x})^2 \to 0$ and $(\bar{x}-\theta)^2/\sum_{i=1}^n (x_i-\bar{x})^2 \to 0$ when $n\to\infty$, adopting expressions in (3) and (5), we get, for all θ,

$$\lim_{n\to\infty} P\left(|y_0-\bar{y}-\hat{\beta}_1(\theta-\bar{x})| \le t_{1-\alpha/2}(n-2)\hat{\sigma}\sqrt{1+\frac{1}{n}+\frac{B_n^2}{\sum_{i=1}^n (x_i-\bar{x})^2}}\right) = 1-\alpha.$$

Since, in carefully designed experiments, the interesting parameter will satisfy $|\theta-\bar{x}| \le B_n$, the information domain confidence interval can be reconstructed as

$$C^*(y) = \left(max\left\{\bar{x}+\frac{y_0-\bar{y}-sgn(\hat{\beta}_1)t_{1-\alpha/2}(n-2)\hat{\sigma}\sqrt{1+\frac{1}{n}+\frac{B_n^2}{\sum_{i=1}^n (x_i-\bar{x})^2}}}{\hat{\beta}_1}, \bar{x}-B_n\right\},\right.$$
$$\left. min\left\{\bar{x}+\frac{y_0-\bar{y}+sgn(\hat{\beta}_1)t_{1-\alpha/2}(n-2)\hat{\sigma}\sqrt{1+\frac{1}{n}+\frac{B_n^2}{\sum_{i=1}^n (x_i-\bar{x})^2}}}{\hat{\beta}_1}, \bar{x}+B_n\right\}\right),$$

where $sgn(\hat{\beta}_1)$ denotes the sign of $\hat{\beta}_1$, with $sgn(\hat{\beta}_1)=-1$ if $\hat{\beta}_1<0$ and $sgn(\hat{\beta}_1)$=1 if $\hat{\beta}_1>0$. According to Theorem, it is easy to verify that, for $|\theta-\bar{x}| \le B_n$,

$$P(\theta\in C^*(Y)) = P(\theta\in C(Y)),$$

that is, the coverage probability for the confidence interval $C^*(y)$ is the same as that for the confidence interval $C(y)$. So the probability behavior for $C^*(y)$ is similar to that for the confidence interval $C(y)$.

Corollary The information domain confidence interval $C^*(y)$ satisfies

$$\inf_{|\theta-\bar{x}|\le B_n} P(\theta \in C^*(Y)) \ge 1-\alpha,$$

and when $|\theta-\bar{x}| \le B_n$,

$$\lim_{n\to\infty} P(\theta \in C^*(Y)) = 1-\alpha.$$

3. Numerical simulation

In this section, for the univariate linear calibration, the proposed confidence intervals are numerically compared with two existing methods which are illustrated as follows.

The $1-\alpha$ confidence interval for θ due to Dahiya and Mckeon (1991) is

$$D(y) = \left\{\theta : |\bar{x}-\theta+\hat{\gamma}(y_0-\bar{y})| \le z_{1-\alpha/2} \times \sqrt{\frac{\hat{\sigma}^2\hat{\gamma}^2}{1+\frac{\hat{\sigma}^2\hat{\gamma}^2}{\sum_{i=1}^n (x_i-\bar{x})^2}}\left(1+\frac{1}{n}+\frac{\hat{\gamma}^2(y_0-\bar{y})^2}{\sum_{i=1}^n (x_i-\bar{x})^2}+\frac{\hat{\sigma}^2\hat{\gamma}^2}{\sum_{i=1}^n (x_i-\bar{x})^2}\right)}\right\},$$

where $\hat{\gamma} = \frac{\hat{\beta}_1 \sum_{i=1}^n (x_i-\bar{x})^2}{\hat{\beta}_1^{\,2}\sum_{i=1}^n (x_i-\bar{x})^2+\hat{\sigma}^2}$ and $z_{1-\alpha/2}$ denotes the $(1-\alpha/2)th$ quantile of the standard normal distribution.

The $1-\alpha$ confidence interval based on nonlinear approach is given by Schechtman and Spiegelman (2002), which is constructed as

$$S(y) = \left\{\theta : |y_0-\bar{y}-\hat{\beta}_1(\theta-\bar{x})| \le t_{1-\alpha/2}(n-2)\hat{\sigma}\sqrt{1+\frac{1}{n}+\frac{(y_0-\bar{y})^2}{\hat{\beta}_1^{\,2}\sum_{i=1}^n (x_i-\bar{x})^2}}\right\}.$$

Now, we display the coverage probability and expected length for confidence intervals using Monte Carlo approach. In simulation studies, the values of the explanatory variable x were assumed that $\bar{x}=0.5$, $\frac{1}{n-1}\sum_{i=1}^n (x_i-\bar{x})^2 = 1$ with $n = 5, 10, 15, 20, 30$. The interesting parameter θ was taken as $-2, 1.5, 3.5$ with $|\theta-\bar{x}| \le B_n$. We chose $\beta_0 = 2$, $\sigma = 1$ and $\beta_1 = 0.1, 0.25, 0.5, 1$. In the proposed intervals simulation, we took $B_n = 3n^{0.05}\sqrt{\frac{1}{n-1}\sum_{i=1}^n (x_i-\bar{x})^2}$ and $B_n = 2n^{0.25}\sqrt{\frac{1}{n-1}\sum_{i=1}^n (x_i-\bar{x})^2}$. The simulation results appear in Tables 1 and 2.

It can be seen from Table 1 that, in the coverage probability, the confidence interval $D(y)$ performs rather badly when the sample size is small, which especially grows worse as β_1 is small. The confidence interval $S(y)$ and confidence intervals we proposed have always good coverage behavior for all parameters configurations and sample sizes $n = 5, 10, 15, 20, 30$.

According to Table 2, we see that the interval lengths for $S(y)$ are too long. Although confidence interval $D(y)$ seems to perform well in the expected length, Table 1 indicates that confidence interval $D(y)$ has bad coverage behavior. The information domain confidence intervals $C^*(y)$ have favorable behavior in the expected length. Especially, confidence interval $C^*(y)$ with $B_n = 3n^{0.05}\sqrt{\frac{1}{n-1}\sum_{i=1}^n (x_i-\bar{x})^2}$ is the shortest for the most parameters configurations.

Figure 1 shows Simulated expected lengths and coverage probabilities for confidence intervals $D(y)$, $S(y)$ and $C^*(y)$ when β_1 is taken from 1 to 5. It is clear that the confidence interval $S(y)$ is too long when β_1 is small. Although the the interval lengths for $D(y)$ are more shorter than proposed intervals $C^*(y)$, Figure 1 also shows that the coverage probability for $D(y)$ is less than the assumed confidence level 95%. Therefore, The information domain confidence intervals $C^*(y)$ have good behavior in both the expected length and the coverage probability.

Table 1. Simulated coverage probabilities of 95% confidence intervals.

β_1	θ	confidence intervals	n = 5	10	15	20	30
0.1	−2	DM	0.2346	0.4385	0.5678	0.6446	0.7398
		NL	0.9376	0.9639	0.9773	0.9842	0.9901
		ID1	0.9683	0.9725	0.9672	0.9707	0.9674
		ID2	0.9625	0.9753	0.9761	0.9805	0.9817
	1.5	DM	0.6740	0.7961	0.8447	0.8744	0.9012
		NL	0.9836	0.9924	0.9957	0.9975	0.9979
		ID1	0.9871	0.9880	0.9852	0.9809	0.9759
		ID2	0.9846	0.9895	0.9898	0.9880	0.9866
	3.5	DM	0.1639	0.3396	0.4705	0.5555	0.6736
		NL	0.9150	0.9452	0.9638	0.9745	0.9845
		ID1	0.9576	0.9577	0.9617	0.9592	0.9596
		ID2	0.9517	0.9619	0.9706	0.9713	0.9750
0.25	−2	DM	0.3101	0.5532	0.6810	0.7714	0.8531
		NL	0.9356	0.9630	0.9684	0.9784	0.9803
		ID1	0.9662	0.9726	0.9696	0.9716	0.9657
		ID2	0.9617	0.9761	0.9776	0.9813	0.9797
	1.5	DM	0.6857	0.8313	0.8849	0.9133	0.9427
		NL	0.9829	0.9921	0.9938	0.9950	0.9942
		ID1	0.9864	0.9869	0.9845	0.9826	0.9760
		ID2	0.9843	0.9886	0.9878	0.9895	0.9871
	3.5	DM	0.2376	0.4721	0.6252	0.7158	0.8235
		NL	0.9202	0.9411	0.9549	0.9647	0.9721
		ID1	0.9585	0.9589	0.9581	0.9623	0.9613
		ID2	0.9528	0.9620	0.9681	0.9748	0.9755
0.5	−2	DM	0.7557	0.7557	0.8594	0.912	0.9413
		NL	0.9352	0.9481	0.9597	0.9654	0.9659
		ID1	0.9636	0.9709	0.9717	0.9706	0.9658
		ID2	0.9587	0.9752	0.9795	0.9806	0.9808
	1.5	DM	0.7530	0.8953	0.9472	0.9589	0.9710
		NL	0.9788	0.9839	0.9864	0.9861	0.9848
		ID1	0.9858	0.9873	0.9847	0.9807	0.9797
		ID2	0.9827	0.9888	0.9885	0.9865	0.9882
	3.5	DM	0.4370	0.7301	0.8374	0.8972	0.9278
		NL	0.9287	0.9376	0.9489	0.9618	0.9595
		ID1	0.9587	0.9651	0.9627	0.9597	0.9580
		ID2	0.9518	0.9676	0.9702	0.9735	0.9740
1	−2	DM	0.7751	0.9004	0.9270	0.9318	0.9441
		NL	0.9415	0.9502	0.9559	0.9562	0.9583
		ID1	0.9641	0.9703	0.9664	0.9657	0.9664
		ID2	0.9598	0.9744	0.9767	0.9773	0.9799
	1.5	DM	0.8520	0.9320	0.9520	0.9494	0.9510
		NL	0.9693	0.9685	0.9704	0.9637	0.9631
		ID1	0.9862	0.9858	0.9853	0.9806	0.9767
		ID2	0.9842	0.9880	0.9897	0.9881	0.9863
	3.5	DM	0.7592	0.8912	0.9199	0.9301	0.9464
		NL	0.9390	0.9471	0.9495	0.9539	0.9595
		ID1	0.9565	0.9595	0.9589	0.9597	0.9587
		ID2	0.9507	0.9641	0.9679	0.9727	0.9724

DM: The confidence interval $D(y)$ due to Dahiya and Mckeon (1991); NL: The confidence interval $S(y)$ due to Schechtman and Spiegelman (2002); ID1: The information domain confidence interval $C^*(y)$ we proposed with $B_n = 3n^{0.05}\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}$; ID2: The information domain confidence interval $C^*(y)$ we proposed with $B_n = 2n^{0.25}\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}$.

4. An example

We now consider a real data set which is from the carbon dating example discussed in Kromer et al. (1986) and Schechtman and Spiegelman (2002). In this example, y represents the radio-carbon age of

Table 2. Simulated expected lengths of 95% confidence intervals.

β_1	θ	confidence intervals	n = 5	10	15	20	30
0.1	−2	DM	3.4927	4.8968	6.0527	7.0798	8.7202
		NL	96133.1045	111664.5029	728658.9073	129621.1655	345768.7626
		ID1	6.3710	6.5560	6.6091	6.7457	6.7786
		ID2	5.8341	6.9391	7.6403	8.2662	9.0574
	1.5	DM	3.4879	4.9231	6.0413	7.0393	8.6546
		NL	776799.4394	22683.6323	5155972.1671	79624392.0130	1022027.1919
		ID1	6.3574	6.5279	6.6873	6.3219	6.6932
		ID2	5.8145	6.9177	7.6967	8.1704	9.0208
	3.5	DM	3.4931	4.9822	6.0330	7.1052	8.7313
		NL	108642.6266	17064.0103	183141439.0158	112316423.0079	256145.3729
		ID1	6.3731	5.9577	6.5126	6.7275	6.7448
		ID2	5.8398	6.3713	7.5705	8.2675	9.0369
0.25	−2	DM	3.5711	5.1640	6.3767	7.5008	9.1642
		NL	172478.2582	2433180.4212	3991577.07972	1754392.5611	256611.2734
		ID1	6.2997	6.4305	6.4394	6.5243	6.4745
		ID2	5.7701	6.8175	7.4535	8.0203	8.6642
	1.5	DM	3.4913	5.0527	6.2691	7.3211	8.9531
		NL	713530.7096	29323.0886	129172.5983	1218062.2557	31529.5645
		ID1	6.3436	6.3080	6.5500	6.6556	6.1557
		ID2	5.8104	6.7141	7.5457	8.1604	8.4800
	3.5	DM	3.6438	5.2160	6.4883	7.5331	9.2381
		NL	3430801.6966	2130089.8144	104455.1642	29721.0891	424868.7293
		ID1	6.2567	6.2970	6.3542	6.3718	6.1750
		ID2	5.7099	6.6964	7.3745	7.8969	8.4049
0.5	−2	DM	3.8543	5.4461	6.5665	7.2908	8.0053
		NL	50933.5927	37585.6077	22899.3506	23181.0640	1021.4760
		ID1	6.1037	5.7945	5.7447	5.4666	5.5189
		ID2	5.5749	6.1713	6.6727	6.8895	7.2807
	1.5	DM	3.5596	5.1003	6.1378	6.8622	7.5351
		NL	29431.3935	39805.9801	32129.4858	428.4412	354.5173
		ID1	6.2361	6.2852	6.2773	6.2869	6.2306
		ID2	5.7128	6.6449	7.1808	7.5729	8.0008
	3.5	DM	4.0228	5.6935	6.7418	7.5731	8.1778
		NL	123719.0788	8697.2055	2728.1873	1762.4918	131.0772
		ID1	5.8197	5.4920	5.3820	5.1780	5.1255
		ID2	5.2463	5.8837	6.3315	6.5282	6.8768
1	−2	DM	4.1083	4.7349	4.6696	4.5796	4.3722
		NL	2999.4681	100.3660	6.9411	5.5118	4.8445
		ID1	5.2588	4.2462	3.8385	3.6574	3.4565
		ID2	4.7395	4.5457	4.5113	4.5628	4.6293
	1.5	DM	3.3274	4.0120	4.0881	4.1207	4.0618
		NL	3744.4962	46.8362	6.7725	4.8359	4.4557
		ID1	5.9576	5.3536	5.0259	4.8190	4.5481
		ID2	5.4773	5.6352	5.6226	5.5675	5.3791
	3.5	DM	4.4444	5.0486	4.9663	4.7694	4.5314
		NL	318939.6857	32.5805	11.1868	5.7857	5.0570
		ID1	5.0263	3.7746	3.3466	3.1555	2.9771
		ID2	4.4907	4.0789	4.0390	4.0964	4.2397

DM: The confidence interval $D(y)$ due to Dahiya and Mckeon (1991); NL: The confidence interval $S(y)$ due to Schechtman and Spiegelman (2002); ID1: The information domain confidence interval $C^*(y)$ we proposed with $B_n = 3n^{0.05}\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}$; ID2: The information domain confidence interval $C^*(y)$ we proposed with $B_n = 2n^{0.25}\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}$.

an artifact, and x is an age obtained by more accurate methods such as by counting tree rings. The data is given in Table 3.

In the analysis, we deleted a data point from the data set, and use the remaining 20 data values to estimate the parameters of the linear calibration model. Then, we adopted the proposed and

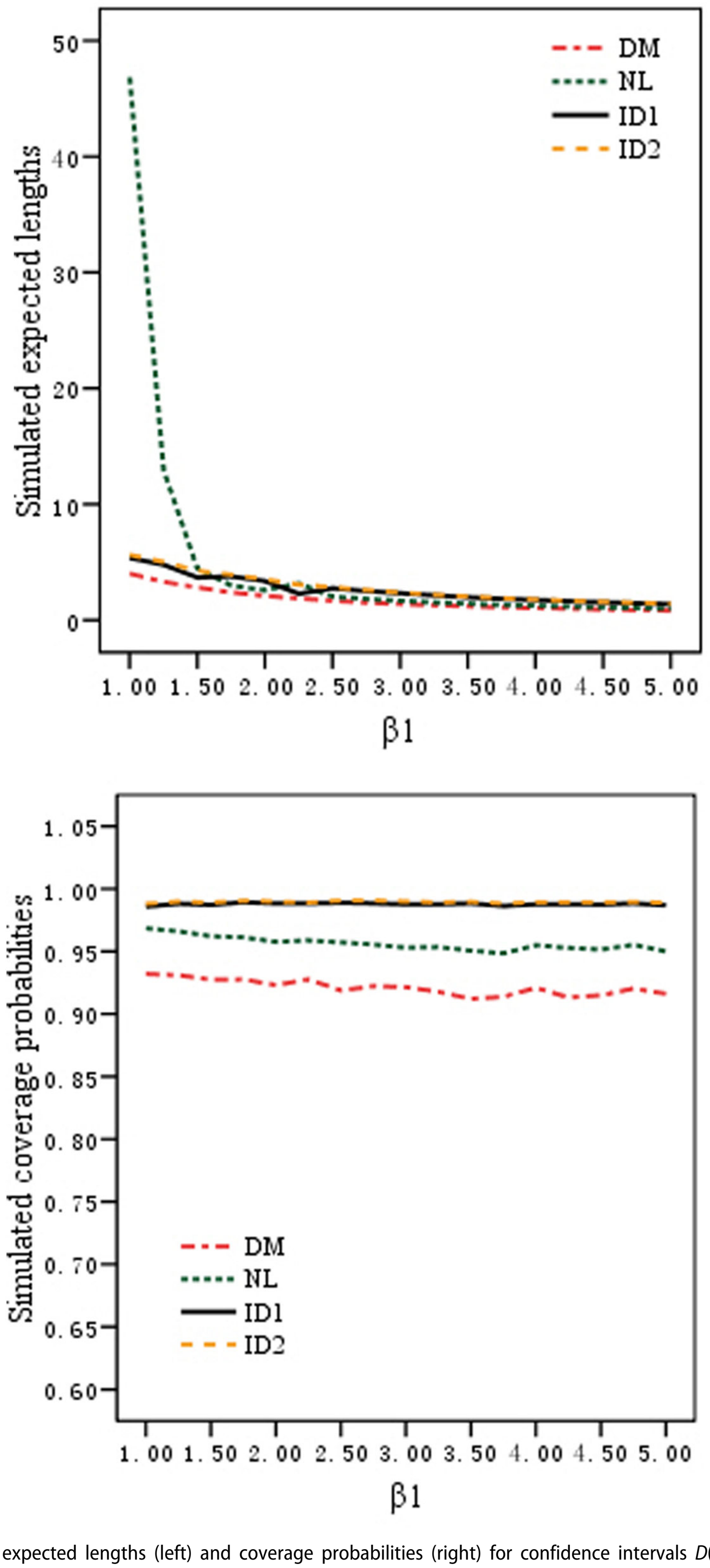

Figure 1. Simulated expected lengths (left) and coverage probabilities (right) for confidence intervals $D(y)$ (DM), $S(y)$(NL), and $C^*(y)$(ID1 and ID2) when $n = 10$, $\beta_0 = 2$, $\sigma = 1$ and $\theta = 1.5$.

Table 3. The carbon dating data.

x	7207	7194	7178	7173	7166	7133	7129	7107	7098	7088	7087
y	8199	8271	8212	8211	8198	8141	8166	8249	8263	8161	8163
x	7085	7077	7074	7072	7069	7064	7062	7060	7058	7035	
y	8158	8152	8157	8081	8000	8150	8166	8083	8019	7913	

Table 4. 95% confidence intervals and interval lengths for the carbon dating example.

The removed x value	$\hat{\beta}_1$	methods	95% confidence intervals	interval lengths
7098	1.1677	DM	(7092.4490,7314.5611)	222.1121
		NL	(7081.5258,7336.6673)	255.1415
		ID1	(7061.2842,7286.7152)	225.4310
		ID2	(7048.5988,7325.3571)	276.7583
7069	1.0668	DM	(6840.4974,7102.5540)	262.0566
		NL	(6806.1755,7116.7140)	310.5385
		ID1	(6928.9210,7127.4696)	198.5486
		ID2	(6890.7891,7141.7181)	250.9290
7058	1.0723	DM	(6860.4048,7121.2496)	260.8448
		NL	(6826.3305,7136.3800)	310.0495
		ID1	(6931.2182,7152.9997)	221.7815
		ID2	(6893.4596,7167.7304)	274.2708

DM: The confidence interval $D(y)$ due to Dahiya and Mckeon (1991); NL: The confidence interval $S(y)$ due to Schechtman and Spiegelman (2002); ID1: The information domain confidence interval $C^*(y)$ we proposed with $B_n = 3n^{0.05}\sqrt{\frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2}$; ID2: The information domain confidence interval $C^*(y)$ we proposed with $B_n = 2n^{0.25}\sqrt{\frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2}$.

the two existing methods to obtain the 95% confidence intervals for the removed x value. The results are summarized in Table 4.

It is clear from Table 4 that the information domain confidence intervals $C^*(y)$ have favorable behavior in the interval length. Especially, the confidence interval $C^*(y)$ with $B_n = 3n^{0.05}\sqrt{\frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2}$ is the shortest as $\hat{\beta}_1$ is small.

5. Conclusions

In this paper, we deal with the problem of the confidence interval for the univariate linear calibration. Since the interesting parameter θ is unidentifiable as β_1 tends to zero, no frequentist procedures so far can provide completely satisfactory solutions. We propose the information domain confidence interval for the interesting parameter θ. It is verified that the formation domain confidence interval has the confidence level $1-\alpha$ on the given information domain, and for all θ the coverage probability of the confidence interval is approximately equal to $1-\alpha$ when the sample size n is large enough. We also numerically compare the proposed confidence intervals with two existing intervals. Numerical results suggest that two existing intervals perform bad in either the coverage probability or the expected length. Especially, both intervals grow worse as β_1 is small. The information domain confidence intervals have the coverage probability above the nominal level, and have satisfactory behavior in the expected length, even if the value of β_1 is small. In simulation studies, we choose B_n as $3n^{0.05}\sqrt{\frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2}$ and $2n^{0.25}\sqrt{\frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2}$. It can be also taken as $3\sqrt{\frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2}$ in many practical situation, which makes the expected lengths of proposed intervals shorter.

Funding

We are grateful to the reviewer for the valuable comments and suggestions. This work was supported by the National Natural Science Foundation of China under Grant (No. 11471035) and Beijing Education Committee Science And Technology Plan General Project under Grant (No. KM201810009013).

ORCID

Xingzhong Xu http://orcid.org/0000-0003-2665-6211

References

Bellio, R. 2003. Likelihood methods for controlled calibration. *Scandinavian Journal of Statistics* 30 (2):339–53. doi:10.1111/1467-9469.00334.

Benton, D., K. Krishnamoorthy, and T. Mathew. 2003. Inferences in multivariate-univariate calibration problems. *Journal of the Royal Statistical Society: Series D (the Statistician)* 52 (1):15–39. doi:10.1111/1467-9884.00339.

Bonate, P. L. 1993. Approximate confidence intervals in calibration using the bootstrap. *Analytical Chemistry* 65 (10):1367–72. doi:10.1021/ac00058a012.

Chvosteková, M. 2019. Multiple use confidence intervals for a univariate statistical calibration. *Measurement Science Review* 19 (6):264–70. doi:10.2478/msr-2019-0034.

Dahiya, R. C., and J. J. Mckeon. 1991. Modified classical and inverse estimators in calibration. *Sankhyā* 53:48–55.

Dunne, A. 1995. Decision and detection limits for linear homoscedastic assays. *Statistics in Medicine* 14 (18): 1949–59. doi:10.1002/sim.4780141802.

Eisenhart, C. 1939. The interpretation of certain regression methods and their use in biological and industrial research. *The Annals of Mathematical Statistics* 10 (2):162–86. doi:10.1214/aoms/1177732214.

Gleser, L. J., and J. T. Hwang. 1987. The nonexistence of $100(1-\alpha)$ confidence set of finite expected diameter in errors-in-variable and related models. *The Annals of Statistics* 15 (4):1351–62. doi:10.1214/aos/1176350597.

Hoadley, B. A. 1970. Bayesian look at inverse linear regression. *Journal of the American Statistical Association* 65 (329):356–69. doi:10.1080/01621459.1970.10481085.

Hunter, W., and W. A. Lamboy. 1981. Bayesian analysis of the linear calibration problem. *Technometrics* 23 (4): 323–28. doi:10.1080/00401706.1981.10487672.

Jones, G., and D. M. Rocke. 1999. Bootstrapping in controlled calibration experiments. *Technometrics* 41 (3): 224–33. doi:10.1080/00401706.1999.10485671.

Kang, P., C. Koo, and H. Roh. 2017. Reversed inverse regression for the univariate linear calibration and its statistical properties derived using a new methodology. *International Journal of Metrology and Quality Engineering* 8: 28. doi:10.1051/ijmqe/2017021.

Krishnamoorthy, K.,. P. M. Kulkarni, and T. Mathew. 2001. Multiple use one-sided hypotheses testing in univariate linear calibration. *Journal of Statistical Planning and Inference* 93 (1–2):211–23. doi:10.1016/S0378-3758(00)00164-6.

Kromer, B., M. Rhein, M. Bruns, H. Schoch-Fischer, K. O. Munnich, M. Stuiver, and B. Becker. 1986. Radiocarbon calibration data for the 6th to the 8th Millennia BC. *Radiocarbon* 28 (2B):954–60. doi:10.1017/S0033822200060239.

Krutchkoff, R. G. 1967. Classical and inverse regression methods of calibration. *Technometrics* 9 (3):425–39. doi:10.1080/00401706.1967.10490486.

Kubokawa, T., and C. P. Robert. 1994. New prespective in linear Calibration. *Journal of Multivariate Analysis* 51 (1):178–200. doi:10.1006/jmva.1994.1056.

Lin, C. S., and M. Huang. 2010. Optimal designs for estimating the control values in multi-univariate regression models. *Journal of Multivariate Analysis* 101 (5):1055–66. doi:10.1016/j.jmva.2009.12.021.

Liseo, B. 2003. Bayesian and conditional frequentist analyses of the Fieller's problem: A critical review. *International Journal of Statistics* LXI:133–52.

Marciano, F. W. P., B. G. Blas Achic, and F. Cysneiros. 2016. Symmetrical linear calibration model to replicated data. *Chilean Journal of Statistics* 7:3–16.

Muhammad, F., and M. Riaz. 2016. An improved approach to multivariate linear calibration. *Scientia Iranica* 23 (3):1355–69. doi:10.24200/sci.2016.3903.

Muhammad, F., M. Riaz, and H. Dawood. 2019. An effective approach to linear calibration estimation with its applications. *Communications in Statistics - Theory and Methods*. Advance online publication. doi:10.1080/03610926.2019.1615092.

Naszodi, L. J. 1978. Elimination of the bias in the course of calibration. *Technometrics* 20:201–205.

Ng, K. H., and A. H. Pooi. 2008. Calibration intervals in linear regression models. *Communications in Statistics - Theory and Methods* 37 (11):1688–96. doi:10.1080/03610920701826120.

Schechtman, E., and C. Spiegelman. 2002. A nonlinear approach to linear calibration intervals. *Journal of Quality Technology* 34 (1):71–79. doi:10.1080/00224065.2002.11980130.

Zhao, G. M., and X. Z. Xu. 2017a. The p-values for one-sided hypothesis testing in univariate linear calibration. *Communications in Statistics - Simulation and Computation* 46 (6):4828–40. doi:10.1080/03610918.2015.1134569.

Zhao, G. M., and X. Z. Xu. 2017b. Uniformly most powerful unbiased test in univariate linear calibration. *Statistics* 51 (3):609–14. doi:10.1080/02331888.2016.1265968.