International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 13, Number 12 (2020), pp. 4118-4124

Application of Resampling Techniques in Orthogonal Regression

Anwar Fitrianto[1*], Tan Sin Yun[2], Wan Zuki Azman Wan Ahmad[3]

[1] *Department of Statistics, Faculty of Mathematics and Natural Sciences, Bogor Agricultural University, Indonesia.*

[2] *Department of Mathematics, Faculty of Science, Universiti Putra Malaysia.*

[3] *Institute of Engineering Mathematics, Universiti Malaysia Perlis.*

Abstract

The classical Orthogonal Regression analysis relies heavily on the normality assumption. However, sometimes we might be uncertain of the underlying distribution of our dataset or the sample size might be small, which would cause an inaccurate inference on the parameter if the data is not normally distributed. This leads to the main objective of this paper which is to examine alternative methods to the parametric OR analysis which do not rely on the normality assumption. In this paper, the nonparametric jackknife and bootstrap resampling techniques were applied to assess the bias, standard errors and confidence intervals for the parameters of the model. We studied on the method of delete-one jackknife and bootstrapping the observations and made comparisons between the two methods as well. Under bootstrapping, three methods were considered to construct the confidence intervals which include percentile interval, bias-corrected (BC) interval and bias-corrected and accelerated (BCa) interval. Based on the results, it was found that the bootstrap estimators were closer to the values of classical OR analysis compared to jackknifed estimators. Besides, the jackknife estimates of bias and standard errors were slightly larger than that of bootstrap. Furthermore, we also found that the confidence intervals for the parameters constructed from jackknife have longer lengths and closer to that of OR. This showed that jackknife performed better in constructing confidence interval than the bootstrap.

INTRODUCTION

Orthogonal regression (OR) is a type of structural relationship or errors-in-variables analysis. It is used to examine the linear relationship between two continuous variables X and Y when both variables are measured with error. It is different from ordinary least square (OLS) regression where only the response variable Y contains the measurement error. The best fitting line of OR minimizes the sum of squares of the distances that are perpendicular or in direction other than vertical from the plotted points to the line. A classical use of OR is to determine whether two methods are equivalent in measuring the same quantity.

In the context of regression, our interest is to determine the regression parameters in the model, to estimate the precision of the estimators and to make inference on the parameters. In recent years, many researchers tend to use resampling techniques in regression data analysis instead of the classical parametric procedures. Two of the common resampling techniques are jackknifing and bootstrapping. In the past literatures, many researchers studied on the applications of jackknife and bootstrap in linear regression. However, we found that there are not much efforts and concerns on the OR. This raises our interest to investigate on the resampling in OR.

LITERATURE REVIEW

Orthogonal Regression

In the context of simple linear regression, in OLS, it is assumed that the independent variables to be fixed and the dependent variable to contribute to the only error in the model. However, in practice, we might be facing the case where the independent variables cannot be observed directly and thus, the measurement errors arise from both dependent and independent variables. This type of problem is known as errors-in-variables model. It would be problematic if in that particular situation, OLS is used. Hence, it is more appropriate to apply the OR on the particular situation.

Adcock (1878) was the first who proposed a method to solve errors-in-variables problems. His idea was to fit a line by minimizing the sum of squares of the perpendicular distances from the plotted points to the line. Kummel (1879) then further developed Adcock's idea. Instead of just considering the perpendicular distance, he suggested to minimize the sum of squares of distances in other direction. Deming (1943) also discussed the idea of OR and thus, OR is sometimes referred to Deming regression (DR). Apart from those mentioned above, the errors-in-variables context was also discussed by many other authors such as Kendall and Stuart (1973) and Fuller (1987), to name a few.

According to Carroll and Ruppert (1996), the basic ideas of OR are as shown below. Suppose there is a pair of variables (ξ,η) that are linearly related with the following expression:

$$\eta = \beta_0 + \beta_1\xi, \tag{1}$$

where β_0 is the intercept and β_1 is the slope. In the classical orthogonal regression development, instead of observing (ξ,η), both of the variables are observed with measurement errors such as

$$X = \xi + \delta \qquad \text{and} \qquad Y = \eta + \varepsilon \tag{2}$$

with δ and ε are independent mean zero random variables with variances σ_δ^2 and σ_ε^2 respectively. A regression-like model as shown below is obtained by combining Equation (1) and (2):

$$Y = \beta_0 + \beta_1 \xi + \varepsilon \tag{3}$$

In OR, it is needed to have prior knowledge on the error variance ratio,

$$\lambda = \sigma_\varepsilon^2 / \sigma_\delta^2 , \tag{4}$$

in order to estimate the parameter. Fuller (1987) explained about the situation of over-parameterization if λ is unknown, which will cause β_1 to be not identifiable.

Carroll and Ruppert (1996) mentioned that the OR estimators can be obtained based on a n sized random sample of $X's$ and $Y's$. By minimizing

$$\sum_{i=1}^{n} \left\{ \left(Y_i - \beta_0 - \beta_1 \xi_i\right)^2 / \lambda + \left(X_i - \xi_i\right)^2 \right\}, \tag{5}$$

the slope estimate $\hat{\beta}_0$ and intercept estimate $\hat{\beta}_1$ are computed as

$$\hat{\beta}_1 = \frac{\left(S_Y^2 - \lambda S_X^2\right) + \sqrt{\left(S_Y^2 - \lambda S_X^2\right)^2 + 4\lambda S_{XY}^2}}{2S_{XY}} \tag{6}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \tag{7}$$

where $\bar{X} = \frac{\sum X_i}{n}$, $\bar{Y} = \frac{\sum Y_i}{n}$, $S_X^2 = \sum \left(X_i - \bar{X}\right)^2$, $S_Y^2 = \sum \left(Y_i - \bar{Y}\right)^2$ and $S_{XY} = \sum \left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)$. The estimator in Equation (6) has been justified in several approaches. One of them is the maximum likelihood estimation (MLE). According to Fuller (1987), in MLE, the $X's$ and $Y's$ are assumed to be independent and normally distributed.

Meanwhile, Patefield (1977) came out with the derivation of the asymptotic variance-covariance matrix of the maximum likelihood estimators of β_0 and β_1. He suggested that when σ^2 is unknown, a consistent estimator of the variance-covariance matrix would be

$$\frac{\left(1+\hat{\beta}_1^2\right)\tilde{\sigma}^2 \hat{\beta}_1}{nS_{XY}} \begin{bmatrix} \bar{X}^2 (1+\hat{\tau}) + \left(S_{XY}/\hat{\beta}_1\right) & -\bar{X}(1+\hat{\tau}) \\ -\bar{X}(1+\hat{\tau}) & (1+\hat{\tau}) \end{bmatrix} \tag{8}$$

where $\hat{\tau} = \frac{\tilde{\sigma}^2 \hat{\beta}_1}{\left(1+\hat{\beta}_1^2\right) S_{XY}}$ and $\tilde{\sigma}^2 = \frac{2n\hat{\sigma}^2}{(n-2)}$. $\tilde{\sigma}^2$ is the consistent estimator of σ^2 whereas $\hat{\sigma}^2$ is the maximum likelihood estimator of σ^2.

The $100(1-\alpha)\%$ confidence interval for β_1 of OR model was constructed by Kendall and Stuart (1973) based on the normality assumptions. The interval is the transformed of

$$\tan^{-1}\beta_1 \in \tan^{-1}\hat{\beta}_1 \pm \frac{1}{2}\sin^{-1}\left\{ 2t_{n-2;\alpha/2} \left[\frac{S_{XX} S_{YY} - S_{XY}^2}{(n-2)\left[\left(S_{YY} - S_{XX}\right)^2 + 4S_{XY}^2\right]} \right]^{1/2} \right\} \tag{9}$$

where $t_{n-2;\alpha/2}$ is the $\left(1-\frac{\alpha}{2}\right)$ percentile point of the *t*-distribution with $(n-2)$ degrees of freedom.

Resampling Techniques

Resampling technique has been widely applied in data analysis. It is used to estimate the precision of sample statistics and make inference by taking repeated samples within the same sample. Two common types of resampling techniques are jackknife and bootstrap.

Jackknife Resampling

Jackknife preceded the bootstrap resampling technique. Quenouille (1956) first introduced the idea of jackknife procedure to reduce the bias of an estimator. Tukey (1958) then expanded the use of jackknife to estimate variance and construct confidence limits based on pseudo-values. He also tailored the name of jackknife. According to Friedl and Stampher (2002), there are two types of jackknife algorithms which are delete-one jackknife and delete-d jackknife. The delete-one jackknife is carried out by deleting single observation from the original sample sequentially while delete-d jackknife is a more generalized technique which is based on multiple observations deletion.

Bootstrap Resampling

The bootstrap resampling method was first introduced by Efron (1979) to be applied on a variety of estimation problems. He demonstrated the method on a series of examples and compared with the jackknife results. The examples included variance of the sample median, error rates in a linear discriminant analysis, ratio estimation and so forth. Based on his results, jackknife was concluded to be a linear approximation method for the bootstrap. Few years later, Efron (1982) proposed three methods for bootstrap confidence intervals namely percentile, bias-corrected (BC) percentile and bootstrap-*t*. The bias-corrected and accelerated (BCa) interval was later introduced by Efron (1987) to achieve a better result on confidence interval.

Resampling in Linear Regression

In linear regression context, Efron (1979) was the first to propose the use of bootstrap and jackknife to estimate regression parameters. The idea was further developed by Wu (1986) and studied by many other researchers as stated below.

Booth and Hall (1993) suggested bootstrap methods for constructing confidence bands for an unknown linear functional relationship in errors-in-variables model. They applied percentile and percentile-*t* to compute the confidence bands in which no assumptions was made about the ratio of error variances. They found out several interesting features of the bootstrap bands. First, the bands did not shrink to a line as the sample size approached infinity. Second, the percentile-*t* confidence bands showed only a first-order coverage accuracy instead of a second-order accuracy which is normally found in simpler statistical problems. They believed that the accuracy problem was due to the multivariate nature of the particular problem and suggested to use the iterated bootstrap to improve the coverage accuracy.

Besides, Abdullah (1995) also discussed about the use of nonparametric bootstrap method in errors-in-variables model. He compared between four types of bootstrap confidence intervals for the parameter, which included percentile, BC percentile, BCa and also calibrated or iterated BCa. His results showed that the iterated BCa interval was more reliable compared to the other types of intervals.

Sahinler and Topuz (2007) demonstrated the application of bootstrap and jackknife in regression analysis. Instead of only constructing the confidence intervals, they also estimated the bias and standard errors to be compared with the concerning estimates of parametric OLS. Based on their results, they found that bootstrap was preferable in linear regression compared to jackknife.

Furthermore, the application of bootstrap and jackknife in linear regression were also studied by Algamal and Rasheed (2010). They focused on the accuracy of the two methods in estimating the distribution of the linear regression parameters through different sample sizes (n) and different bootstrap replications (B). Based on their results, the jackknife estimators were reliable when the sample size was large enough $(n \geq 50)$. Besides, they also found better results with less bias in bootstrap as B increases.

Francq (2014) investigated on the DR and Bivariate Least-Squares (BLS) regression in constructing the confidence intervals of the regression parameters in errors-in-variables model. He found that the confidence intervals constructed by the two regressions were based on approximation (except the one for slope constructed by DR). This caused the coverage probabilities to be lower than the nominal value. Hence, he applied jackknife and bootstrap procedures to improve the coverage probabilities. Two bootstrap procedures named bootstrapping the residuals and bootstrapping the pairs were considered which were then split into three approaches (percentile, bootstrap-*t* on DR and bootstrap-t on BLS). Francq (2014) found that jackknife had lower coverage probabilities than the nominal level for small sample sizes. Therefore, its confidence interval might be too narrow in practice. In bootstrapping the residuals, the coverage probabilities were collapsed when the ratio of the variances of measurement errors (λ_{XY}) decreased and when the sample size increased. This is due to the fact that the randomness of the errors in X was not taken into account by bootstrapping the vertical residuals. Thus, bootstrapping the residuals was not recommended as the coverage probabilities collapsed and the confidence intervals were shifted in practice. In contrast, bootstrapping the pairs was recommended to improve the coverage probabilities especially when $\lambda_{XY} < 1$. It could provide better coverage probabilities than the approximate confidence intervals computed directly by DR and BLS. Moreover, this bootstrap approach took into account the measurement errors in both variables.

METHODOLOGY

Data

The dataset is adapted from Minitab, Inc. (2014). It was originated from a medical equipment company which wanted to compare their new blood pressure monitor with a similar model on the market. They obtained systolic blood pressure readings on a random sample of 60 people using the two instruments.

Jackknife Resampling Technique

Let θ be the parameter of the population of interest. Suppose a sample of size n from the population is denoted as $\{X_1, X_2, ..., X_n\}$, where each of the X_i can be either univariate or multivariate. In this study, the focus is on the case where X_i's are univariate. The sample estimator of θ is denoted as $\hat{\theta}$ which is a function of the observations in the sample. The general form of $\hat{\theta}$ is given by

$$\hat{\theta} = f(X_1, X_2, ..., X_n). \tag{10}$$

According to Abdi and Williams (2010), the procedure of delete-one jackknife to obtain the jackknife estimate of θ is as follows:

Step 1: Compute $\hat{\theta}_{-i}$, the estimator of θ, by omitting the observation i from the sample, where $i = 1, ..., n$, which is also known as the partial prediction i.

$$\hat{\theta}_{-i} = f(X_1, ..., X_{i-1}, X_{i+1}, ..., X_n) \tag{11}$$

Step 2: Based on the partial prediction i, the pseudo-value i is computed as

$$\hat{\theta}_i^* = n\hat{\theta} - (n-1)\hat{\theta}_{-i}. \tag{12}$$

Step 3: Finally, the bias-corrected delete-one jackknifed estimator, $\hat{\theta}_j$, is computed as the mean of the n pseudo-values. It is noted that the estimator can also be computed without using the pseudo-values. The formula is as follows:

$$\hat{\theta}_j = \sum_{i=1}^{n} \frac{\hat{\theta}_i^*}{n} = \frac{1}{n}\sum_{i=1}^{n}\left(n\hat{\theta} - (n-1)\hat{\theta}_{-i}\right) = n\hat{\theta} - (n-1)\bar{\theta} \quad (13)$$

where

$$\bar{\theta} = \frac{\sum_{i=1}^{n}\hat{\theta}_{-i}}{n}. \quad (14)$$

The jackknife is known as a nonparametric method to estimate and reduce the bias. In general, the bias, B, of an estimator is defined as

$$B = E\left(\hat{\theta}\right) - \theta, \quad (15)$$

where $\hat{\theta}$ is the estimator of θ and $E\left(\hat{\theta}\right)$ is the expected value of $\hat{\theta}$. Based on this, the jackknife estimate of bias can be obtained as follows:

$$\hat{B}_j = \hat{\theta} - \hat{\theta}_j = (n-1)\left(\bar{\theta} - \hat{\theta}\right) \quad (16)$$

The expected value of the estimator is being replaced by the biased estimator whereas the parameter is being replaced by the unbiased jackknife estimator.

Meanwhile, the jackknife estimate of standard error of $\hat{\theta}$ can be obtained from the sample variance of the pseudo-values. Similar to the jackknifed estimator, the standard error can also be computed without using the pseudo-values. The formula is as follows:

$$\hat{\sigma}_j = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\frac{\left(\hat{\theta}_i^* - \hat{\theta}_j\right)^2}{(n-1)}} = \sqrt{\frac{n-1}{n}\sum_{i=1}^{n}\left(\hat{\theta}_{-i} - \bar{\theta}\right)^2} \quad (17)$$

The standard error can be used to construct the confidence interval for the jackknife estimate of θ. The particular estimation is distributed as a Student's t distribution with $(n-1)$ degrees of freedom under the independence assumption and thus, the confidence interval is given by

$$\hat{\theta}_j \pm t_{n-1;\alpha/2}\hat{\sigma}_j \quad (18)$$

Bootstrap Resampling Technique

According to Francq (2014), there are two well-known bootstrap resampling procedures in linear regression, which based on the residuals or the observations. Bootstrapping the residuals is used when the regressors are fixed. Instead, when the regressors are as random as the response, bootstrapping the observations could be a better approach to be applied. In this study, the focus is on OR model where both variables X and Y are measured with errors (random). Therefore, the discussion is proceeded with only bootstrapping the observations throughout the study. According to Sahinler and Topuz (2007), bootstrapping the observations can be conducted as follows:

Step 1 : For $r = 1, 2, ..., B,$ draw a n sized bootstrap sample with replacement from the original sample, where each of the observations has the same probability $1/n$ to be selected and B is the number of replications.

Step 2 : Compute $\hat{\theta}_{b(r)}$, the bootstrap estimator of θ, from each of the bootstrap samples,

Step 3 : Obtain the probability distribution $\left(F\left(\hat{\theta}_b\right)\right)$ of bootstrap estimates $\hat{\theta}_{b(1)}, \hat{\theta}_{b(2)}, \ldots \hat{\theta}_{b(B)}$ which can be used to estimate the bootstrap estimator, bias, standard error and confidence interval.

The bootstrap estimator, $\hat{\theta}_b$, is calculated as the mean of the distribution $F\left(\hat{\theta}_b\right)$,

$$\hat{\theta}_b = \frac{\sum_{r=1}^{B}\hat{\theta}_{b(r)}}{B} \quad (19)$$

and the bootstrap estimate of bias and standard error are given by

$$B_b = \hat{\theta}_b - \hat{\theta}, \quad \text{and} \quad (20)$$

$$\hat{\sigma}_b = \sqrt{\sum_{r=1}^{B}\frac{\left(\hat{\theta}_{b(r)} - \hat{\theta}_b\right)^2}{B-1}}, \quad (21)$$

respectively.

Percentile Interval

According to Efron (1982), the percentile interval is constructed based on the cumulative distribution function which is written as

$$\hat{C}(t) = \Pr{}^*\left\{\hat{\theta}_b \le t\right\} = \frac{\#\left\{\hat{\theta}_b \le t\right\}}{B} \quad (22)$$

of the bootstrap distribution of $\hat{\theta}_b$. The 100 $(1-2\alpha)$ % confidence interval is obtained by finding the α^{th} percentile and $(1-\alpha)^{th}$ percentile. The general form of the confidence interval is written as

$$\left[\hat{C}^{-1}(\alpha), \hat{C}^{-1}(1-\alpha)\right] \quad (23)$$

where $\hat{C}^{-1}$ is the inverse cumulative function based on Equation (22).

Bias-corrected Percentile Interval

Suppose the bias correction is defined as

$$z_0 = \Phi^{-1}\left(\hat{C}\left(\hat{\theta}\right)\right), \tag{24}$$

where Φ is the cumulative distribution function for a standard normal variate and $\hat{C}\left(\hat{\theta}\right)$ is the cumulative distribution function as shown in Equation (22). The 100 $(1-2\alpha)$% confidence interval is given by

$$\left[\hat{C}^{-1}\left(\Phi\left(2z_0 + z_\alpha\right)\right), \hat{C}^{-1}\left(\Phi\left(2z_0 + z_{1-\alpha}\right)\right)\right] \tag{25}$$

where z_α and $z_{1-\alpha}$ are both standard normal distribution functions. Notice if $z_0 = 0$, the BC interval reduces to the percentile interval (Efron, 1982).

Bias-corrected and Accelerated Interval

According to Efron (1987), the bias correction and acceleration factor are needed to calculate the percentiles that are used to determine the limits of the interval in BCa method. The bias correction, z_0, is the same as the one in BC method as shown in (23). The acceleration factor a is proposed to be computed by using delete-one jackknife estimates of the statistics. The jackknife estimates that would be used are the partial predictions of n jackknife samples and the mean of the partial predictions as shown in Equation (11) and (14), respectively. The acceleration is then calculated as

$$a = \frac{\sum_{i=1}^{n}\left(\bar{\theta} - \hat{\theta}_{-i}\right)^3}{6\left\{\sum_{i=1}^{n}\left(\bar{\theta} - \hat{\theta}_{-i}\right)^2\right\}^{3/2}} \tag{3.17}$$

and the 100 $(1-2\alpha)$% BCa confidence interval is given by

$$\left[\hat{C}^{-1}\left(\Phi\left(z[\alpha]\right)\right), \hat{C}^{-1}\left(\Phi\left(z[1-\alpha]\right)\right)\right] \tag{3.18}$$

where $\hat{C}^{-1}$ is the inverse cumulative distribution function of the bootstrap distribution of $\hat{\theta}_b$ based on Equation (22). $z[\alpha]$ is given by

$$z[\alpha] = z_0 + \frac{z_0 + \Phi^{-1}(\alpha)}{1 - a\left(z_0 + \Phi^{-1}(\alpha)\right)} \tag{3.19}$$

and likewise for $z[1-\alpha]$.

RESULTS AND DISCUSSION

Orthogonal Regression Analysis

Figure 1 showed the plot of the 60 pairs of observations of blood pressure of the dataset. There was a strong linear correlation between the readings of the new method and the current method.

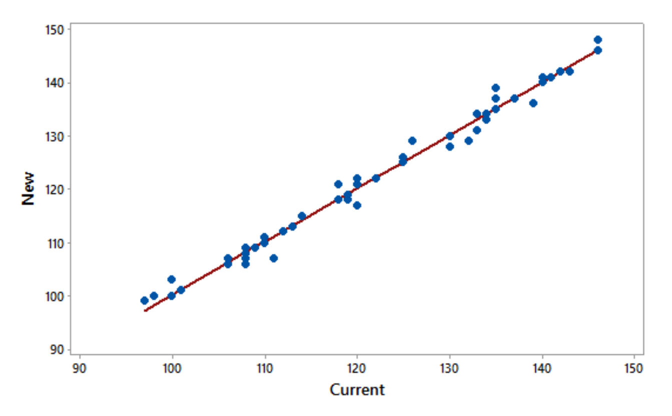

Figure 4.1: Scatter Plot with Fitted Line of Blood Pressure Levels

The calculate error variance ratio for the blood pressure data was 0.90. The estimates of the intercept and slope parameter of OR were $\hat{\beta}_0 = 0.64441$ and $\hat{\beta}_1 = 0.99542$. Their standard errors were 1.74470 and 0.01415, respectively with the 95% confidence interval of β_0 and β_1 are $(-2.77513, 4.06395)$ and $(0.96769, 1.02315)$.

Results of Jackknifing and Bootstrapping

In this study, 60 jackknife samples were generated by omitting one pair of observations at a time (size = $n-1$ = 59). Moreover, 1000 bootstrap samples with the sample each of size 60 from the original dataset were also generated. Both resampling processes were carried out by using SAS 9.3. Summary results of the OR analysis, jackknifing and bootstrapping regression procedures were shown in Table 1.

Table 1: OR, Jackknife and Bootstrap Results of Intercept β_0 and Slope Parameter β_1

Method	β_0			β_1		
	Estimator	Bias	Standard error	Estimator	Bias	Standard error
OR	0.64441	-	1.74470	0.99542	-	0.01415
Jackknife	0.68079	-0.03637	1.72224	0.99516	0.00026	0.01395
Bootstrap	0.65175	0.00734	1.66937	0.99545	0.00002	0.01348

It was observed that both of the bootstrap estimators, $\hat{\beta}_{0(b)}$ and $\hat{\beta}_{1(b)}$, were closer to the estimates of OR compared to the jackknifed estimators $\hat{\beta}_{0(j)}$ and $\hat{\beta}_{1(j)}$. It was noted that the jackknife estimators were computed as the bias-corrected jackknifed estimators based on Equation (13). In comparison, the two jackknife estimates of bias were slightly larger than that of bootstrap. Apart from these, it was also observed that the bootstrap standard errors are smaller than the values of jackknife. However, the jackknifed standard errors were closer to the standard errors obtained from the OR analysis.

Comparisons between the 95% confidence intervals computed from the OR analysis, jackknife and bootstrap were conducted. In this context, the interpretation were based on the common concern such as length of confidence intervals. Three bootstrap confidence intervals were considered which included percentile, BC and BCa. The results were tabulated in Table 2.

Table 2: 95% Confidence Intervals of Intercept β_0 and Slope Parameter β_1 computed from OR, Jackknife and Bootstrap

Method		β_0			β_1		
		Lower limit	Upper limit	Length	Lower limit	Upper limit	Length
OR		-2.77513	4.06395	6.83908	0.96769	1.02315	0.05546
Jackknife		-2.76542	4.12699	6.89241	0.96724	1.02308	0.05584
Bootstrap	Percentile	-2.74407	3.79551	6.53958	0.96988	1.02329	0.05341
	BC	-2.94411	3.61893	6.56304	0.97082	1.02411	0.05329
	BCa	-2.94411	3.61893	6.56304	0.97098	1.02411	0.05313

It could be seen the jackknife confidence intervals for both parameters were superior in length compared to bootstrap. Besides, they were also closer to the confidence intervals computed from OR analysis. All the bootstrap confidence intervals had lengths shorter than that of OR analysis. Meanwhile, there was not much difference in length among the percentile, BC and BCa bootstrap confidence intervals. For β_0, the length of the percentile interval was shorter than the other two intervals but it was vice versa for β_1. Thus, there was no way for us to evaluate the performance between the three methods due to the inconsistent results.

In general, the results of jackknife and bootstrap could be different even though both are similar approaches that rely on data resampling. Fan and Wang (1995) stated that the disparity between jackknife and bootstrap is mainly caused by the sample size. This is because the number of jackknife samples is limited by the original sample size. Thus, it may lead to an inappropriate application of the jackknife resampling technique when the sample size is small.

CONCLUSION AND FUTURE WORK

This paper investigated the use of jackknife and bootstrap resampling technique as alternative approaches to the parametric OR analysis in estimating orthogonal regression coefficients. It was shown that the values of jackknife and bootstrap estimators did not differ much from the value of OR analysis. Besides, both approaches were effective in reducing bias and estimating the standard errors. This helps when the classical method is suspected to be inappropriate. Based on the results, the jackknife performed better in constructing confidence intervals for the parameters. However, apart from the percentile, BC and BCa confidence intervals, there might be some other bootstrap methods that are yet to be studied in this paper. In general, the bootstrap resampling technique is more widely applied compared to jackknife even though it is more computational intensive.

In future works, we could apply the Monte Carlo simulation in order to make better comparison between jackknife and bootstrap resampling techniques because it samples from a population universe. In addition, it is suggested to check for the coverage rate to aid in the comparison between different confidence intervals.

REFERENCES

[1] Abdi, H. & Williams, L. J 2010 ‘Jackknife’, Salkind NJ, editor, Encyclopedia of Research Design, Thousand Oaks, CA: Sage.

[2] Abdullah, M. B. 1995 ‘On bootstrap methods in orthogonal regression model’, Pertanika J. Sci. & Technol. vol. 3, no 2, pp. 349-359.

[3] Adcock, R. J 1878 ‘A problem in least squares’, The Analyst, vol 5, no. 2, pp. 53-54.

[4] Algamal, Z. Y. & Rasheed, K. B 2010 ‘Re-sampling in linear regression model using jackknife and bootstrap’, Iraqi Journal of Statistical Science, vol 18, no. 1, pp. 59-73.

[5] Booth, J. G. & Hall, P 1993 ‘Bootstrap confidence regions for functional relationships in errors-in-variables models’, The Annals of Statistics, vol. 21, no. 4, pp. 1780-1791.

[6] Carroll, R., & Ruppert, D 1996 ‘The use and misuse of orthogonal regression in linear errors-in-variables models’, The American Statistician, vol. 50, no. 1, pp. 1-6.

[7] Deming, W. E 1931 ‘The application of least squares’, Philosophical Magazine and Journal of Science, vol. 11, no. 68, pp. 146-158.

[8] Efron, B 1979 ‘Bootstrap methods: another look at the jackknife’, The Annals of Statistics, vol .7, no. 1, pp. 1-26.

[9] Efron, B 1982 ‘The Jackknife, the Bootstrap and Other Resampling Plans’ SIAM: Philadelphia.

[10] Efron, B 1987 ‘Better bootstrap confidence intervals’, Journal of the American Statistical Association, vol. 82, no. 397, pp. 171-185.

[11] Fan, X. & Wang, L 1995 ‘How comparable are the jackknife and bootstrap results: An investigation for a case of canonical correlation analysis’, Annual meeting of the American Educational Research Association, San Francisco, CA, April 18-22.

International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 13, Number 12 (2020), pp. 4118-4124

[12] Francq, B. G 2014 'Bootstrap in errors-in-variables regressions applied to methods comparison studies', Informatica Medica Slovenica, vol. 19, no. 1-2, pp. 1-11.

[13] Friedl, H. & Stampfer, E 2002 'Jackknife resampling', Encyclopedia of Environmetrics, 2nd Eds, Wiley, Chichester

[14] Fuller, W. A 1987 'Measurement Error Models', John Wiley & Sons Ltd., New York.

[15] Kendall, M. G. & Stuart, A 1973, 'The Advanced Theory of Statistics", Griffin, London.

[16] Kummel, C. H 1879 'Reduction of observed equations which contain more than one observed quantity', The Analyst, vol. 6, no. 4, pp. 97-105.

[17] Minitab, Inc 2014 'Minitab Statistical Software, Release 17 for Windows', State College, Pennsylvania.

[18] Quenouille, M. H 1956 'Notes on bias in estimation', Biometrika, vol. 43, no. 3-4, pp. 353–360.

[19] Sahinler, S. & Topuz, D 2007 'Bootstrap and jackknife resampling algorithms for estimation of regression parameters', Journal of Applied Quantitative Methods, vol. 2, no. 2, pp. 188-199.

[20] Tukey, J. W 1958, 'Bias and confidence in not quite large samples', The Annals of Mathematical Statistics, vol. 29, pp. 614–623.

[21] Wu, C. F. J 1986 'Jackknife, bootstrap and other resampling methods in regression analysis, Annals of Statistics, vol. 14, no. 4, pp. 1261-1295.