Estimating Bias and RMSE of Indirect Effects using Rescaled Residual Bootstrap in Mediation Analysis

ANWAR FITRIANTO
Laboratory of Applied and Computational Statistics, Institute for Mathematical Research
University Putra Malaysia
43400 UPM Serdang, Selangor
MALAYSIA
anwarfitrianto@gmail.com

HABSHAH MIDI
Laboratory of Applied and Computational Statistics, Institute for Mathematical Research
University Putra Malaysia
43400 UPM Serdang, Selangor
MALAYSIA
habshahmidi@gmail.com

Abstract: It is a common practice to estimate the parameters of mediation model by using the Ordinary Least Squares (OLS) method. The construction of T statistics and confidence interval estimates for making inferences on the parameters of a mediation model, particularly the indirect effect, is usually are based on the assumption that the estimates are normally distributed. Nonetheless, in practice many estimates are not normal and have a heavy tailed distribution which may be the results of having outliers in the data. An alternative approach is to use bootstrap method which does not rely on the normality assumption. In this paper, we proposed a new bootstrap procedure of indirect effect in mediation model which is resistant to outliers. The proposed approach was based on residual bootstrap which incorporated rescaled studentized residuals, namely the Rescaled Studentized Residual Bootstrap using Least Squares (ReSRB). The Monte Carlo simulations showed that the ReSRB is more efficient than some existing methods in the presence of outliers.

Key-Words: Mediation analysis, outliers, bootstrap, studentized residuals, indirect effect

1 Introduction

In order to improve the understanding of the nature of relationship between an independent and dependent variables, a third variable is often suggested to be included in the model. When the third variable is considered as a mediator, it is hypothesized to be linked in a causal chain between the independent and dependent variables [1]. The search for intermediate causal variables is called mediation analysis. Mediation analysis is common in social science research, as they elaborate upon other relationships. In these studies, participants are randomly assigned to receive different experimental conditions, and differences in means in the conditions are either consistent or inconsistent with a mediation theory. For example, cognitive dissonance is a social psychological theory which explains that persons make decisions at least in part to reduce internal discomfort or dissonance. Another example is related to programs to reduce drug use, which often target mediators as resistance skills, hypothesizing that a program that increases resistance skills will decrease drug use [2].

The bootstrap is now a widely used method ([3], [4], [5], [6], [7]). In the simplest form of bootstrapping, called the percentile bootstrap, upper and lower confidence limits are obtained by finding the values of the statistics in the 1000 samples that correspond to the 2.5% and 97.5% percentiles. There are several variations of the bootstrap method that are useful in some situations. Another bootstrap method, called the bias-corrected bootstrap, is important for mediation analysis because of its accuracy for computing confidence intervals for the mediated effect when the mediated effect is nonzero [8]. The method consists of adjusting each bootstrap sample for potential bias in the estimate of the statistic. The bias-corrected bootstrap method removes bias that arises because the true parameter value is not the median of the distribution of the bootstrap estimates. The bias correction is used to obtain a new upper and lower percentile used to adjust the confidence limits in the bootstrap distribution.

According to [9], a mediation analysis involves studying the following three regression equations: 1) the regression between the independent variable and the mediator, 2) the regression between the predictor and the outcome, and 3) the regression between the mediator and the outcome when controlling for the predictor. If these three relations are significant, the Sobel test, [10], is used to determine whether the predictor's effect on the outcome is significantly less when the effect of the mediator is taken into account. Detailed regression equation needed to establish a mediation analysis is as follows:

$$Y = i_1 + cX + \varepsilon_1 \quad (1)$$

$$M = i_2 + aX + \varepsilon_2 \quad (2)$$

$$Y = i_3 + c'X + bM + \varepsilon_3 \quad (3)$$

where Y is the dependent variable, X is the independent variable, M is the mediating variable or mediator. Ordinary least squares regression methods (OLS) is the common method to estimate c', a, b, and c in the mediation analysis, since it contains a series of regression analyses. In the OLS, there are several assumptions that have to be fulfilled for the regression model to be valid. When the regression model does not meet the fundamental assumptions, the prediction and estimation of the model may become biased.

Baron and Kenny's causal steps method of examining the significance of the mediation effect has been criticized [11]. First, some researchers have questioned the necessity of testing the overall association in step 1 [12]. Second, MacKinnon in [13] indicates that Baron and Kenny's causal steps method have Type I error rates that are too low and have very low power, unless the sample size is large.

To overcome the limitations associated with Baron and Kenny's, [11] argue that with nonsymmetric confidence intervals for $(a \times b)$, bootstrap methods should be used to assess mediation. The development of bootstrapping methods in mediation analysis has received little attention especially when outliers are present in the data. It is worth mentioning that the classical bootstrap method is easily affected by outliers [14]. This motivates us to develop a new robust bootstrap method in the mediation analysis.

2 General Bootstrap Procedures

Resampling methods have become practical with the general availability of rapid computing and new software. As computers have become faster and easier to use, they have made possible more powerful and accurate graphical and numerical methods in statistics. Bootstrap and resampling methods are part of this revolution. They are feasible only with the use of software, but they are in many ways easy to use and they generalize to situations where traditional methods based on the theory of the Normal distribution cannot be applied.

Bootstrapping method becomes more popular as powerful tools used for statistical inference nowadays. Since its introduction by Efron [15], the bootstrap has been the object of research in statistics. Bootstrapping uses a sample data to estimate relevant characteristics of the population. The sampling distribution of a statistic is then constructed empirically by resampling from the sample. The resampling procedure is designed to parallel the process by which sample observations were drawn from the population. For example, if the data represent an independent random sample of size *n* (or a simple random sample of size *n* from a much larger population), then each bootstrap sample selects *n* observations with replacement from the original sample.

Obtaining the desired statistics, it is needed to understand the general bootstrap procedure, which is as follows:

1. Obtain a single sample of size n from the population under study and calculate $\hat{\theta}$ (the estimate of θ based on the sample).
2. Generate a bootstrap sample of the same size n by sampling *with replacement* from your original sample.
3. Calculate $\hat{\theta}*$, the value of the estimate based on the bootstrap resample.
4. Repeat steps 2 and 3 to obtain $\hat{\theta}_1^*$, $\hat{\theta}_2^*$, …, $\hat{\theta}_B^*$.

2.1 Bootstrap bias and standard error

Two typical measures of accuracy for $\hat{\theta}$ is the standard error (SE) and the bias of the estimator. It can be estimated by the bootstrap samples by

$$SE_{boot} = \sqrt{\frac{\sum_{b=1}^{B}\left[\hat{\theta}_b^* - \bar{\hat{\theta}}\right]^2}{B-1}},$$

where $\bar{\hat{\theta}} = \frac{\sum_{b=1}^{B} \hat{\theta}_b^*}{B}$.

Meanwhile the bias is written by

$$Bias(\hat{\theta}) = E(\hat{\theta}) - \theta$$
$$= \bar{\hat{\theta}} - \hat{\theta},$$

where $\hat{\theta}_i^*$ are bootstrap copies of $\hat{\theta}$, as defined earlier.

3 Bootstrapping Regression Models

In regression analysis, there are two basic approaches to perform bootstrapping, and the choice of which to use depends upon the regressors [16]. One is to bootstrap the vector of the response variables and the associated predictor variable. Such bootstrap procedure is commonly called as pair bootstrap. It was proposed by Freedman [17]. The other one is to first fit the model and bootstrap the residuals which have been proposed for the first time by Wu [18]. Bootstrapping the residuals requires that the residuals be independent and identically distributed. If the regressors are *fixed*, as in design experiment, the bootstrap resampling must preserve that structure. Each bootstrap sample should have the same regressors. Then, if this is the case, the appropriate bootstrap is the residual bootstrap. But if the data has random regressors, bootstrap samples should posses this additional variation from the regressors, so that pair bootstrap is appropriate.

Assume that we want to fit a regression model with response variable Y and predictors $\{X_1, \ldots, X_p\}$. Then we can write down that we have a sample of *n* cases of the form

$$\mathbf{z}_i' = [Y_i, X_{1i}, \ldots, X_{1p}].$$

In pair bootstrap, we select *n* cases randomly from the original $\mathbf{z}_1', \mathbf{z}_1', \ldots, \mathbf{z}_n'$ observations with replacement. Then the regression estimator computed for each of the resulting bootstrap samples, $\mathbf{z}_{b1}'^*, \mathbf{z}_{b2}'^*, \ldots, \mathbf{z}_{bn}'^*$, producing B sets of bootstrap regression coefficients, $\boldsymbol{\beta}_b^* = [\beta_{0b}^*, \beta_{1b}^*, \ldots, \beta_{pb}^*]'$. Standard errors and confidence intervals for the regression coefficients can be constructed based on B (a given large number) bootstrap samples. But, [19] pointed out that pair bootstrap is less sensitive to model assumptions such as normality and constant variance than bootstrapping residuals.

Directly resampling the observations $\mathbf{z}_i'$ implicitly treats the regressors $X_1, \ldots, X_p$ as *random* rather than *fixed*. We may want to treat the *X*s as fixed (if, e.g., the data derive from an experimental design). This is accomplished by constructing the distribution F_n that places probability $1/n$ at each e_i. We then generate bootstrap residuals e_i^* for $i = 1,2,\ldots,n$, where the e_i^* are obtained by sampling independently from F_n. Practically, recent book by Fox [20] mentioned the steps of doing residual bootstrap samples in regression as follows:

1. Estimate the regression coefficients $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p$ for the original sample, and calculate the fitted value and residual for each observation:

   $$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \ldots + \hat{\beta}_p X_{ip}$$
   $$e_i = Y_i - \hat{Y}_i.$$

   The e_i is then scaled using studentized residual,

   $$r_i = \frac{e_i}{\sigma\sqrt{1 - h_i}},$$

   where h_i is leverage value at point i. Many researcher also used raw residual, e_i, instead of studentized residual, r_i.

2. Select bootstrap samples of the *residuals*, $\mathbf{e}_b^* = [r_{1b}^*, r_{2b}^*, \ldots, r_{nb}^*]'$, and from these, calculate bootstrapped Y values, $\mathbf{y}_b^* = [Y_{1b}^*, Y_{2b}^*, \ldots, Y_{nb}^*]'$, where $Y_{ib}^* = \hat{Y}_i + e_{ib}^*$.
3. Regress the bootstrapped *Y* values on the *fixed X* values to obtain bootstrap regression coefficients. If, for example, estimates are calculated by least-squares regression, then $\mathbf{b}_b^* = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_b^*$ for $b = 1,2,\ldots,B$.
4. The resampled $\mathbf{b}_b^* = [\beta_{0b}^*, \beta_{1b}^*, \ldots, \beta_{pb}^*]'$ can be used in the usual manner to construct bootstrap standard errors and confidence intervals for the regression coefficients.

Bootstrapping with fixed X draws an analogy between the fitted value $\hat{Y}$ in the sample and the conditional expectation of Y in the population, and between the residual e in the sample and the error ε in the population. Although no assumption is made about the *shape* of the error distribution, the bootstrapping procedure, by constructing the Y_{ib}^{*} according to the linear model, implicitly assumes that the functional form of the model is correct.

3.1 Bootstrap Estimates of the Indirect Effect in Mediation Analysis

The principle of bootstrapping is fundamental to Shrout and Bolger's method [11] for testing the statistical significance of an indirect effect. In statistics of mediation analysis, formulas for the point estimator of the indirect effect (mediated effect) and its standard error can be used to test the significance of a mediated effect and construct confidence intervals for the effect in the population based on data from the single sample. One assumption of this approach is that the population distribution underlying both the point and standard error estimates of the mediated effect is normal. The mediated effect is a product of a and b. In fact, the distribution of the product of two random variables does not have a normal distribution in all situations, and information regarding the distribution of the product of two random variables has been shown to provide more accurate confidence limits and statistical tests [21].

Traditional statistical inference is based on assumptions about the distribution of the population from which the sample is drawn. In standard multiple regression analysis, the estimates of the standard errors of parameter estimates are derived based on the assumption of normality. Bootstrapping uses computer intensive resampling to make inferences rather than making assumptions about the population. In other words, bootstrapping treats a given sample as the population.

Bootstrapping can be used for any statistics, but in this study, we focus only on the mediated effect (also see [22], for a treatment of bootstrapping a mediated effect). The following steps of a pair bootstrap procedure outlined by [11] will be presented below. Parameter estimates are based on Eq. (2) and Eq. (3).

1. Using the original data set of n cases as a population reservoir, create a bootstrap sample of n cases by random sampling with replacement,
2. Calculate a, b, and $(a \times b)$ based on this bootstrap sample,
3. Repeat Steps 1 and 2 a total of B times,
4. Examine the distribution of the estimates.

The probability distribution constructed in step 4 becomes the basis for making inferences. Thus, the standard deviation of this distribution is the bootstrapped standard error of the mediated effect.

Bootstrap estimates of the mediated effect are straightforward to obtain. First, the mediated effect estimate is obtained from the sample of data. In a bootstrap analysis of a sample of $B = 100$, for example, a new sample of 100 is taken with replacement from the original sample of 100 and the mediated effect is estimated. Then a second sample of 100 is taken from the original sample, and the mediated effect is estimated. The process is repeated a large number of times, usually at least 1000. The mediated effect estimated in each bootstrap sample is used to form a distribution of the bootstrap mediated effect estimates.

4 The Proposed Bootstrap Approach

Due to its strength-proven in statistics, bootstrap residual approach should be taken into account in doing bootstrap method in mediation analysis. The approach is needed since mediation analysis involves several linear regression equations.

MacKinnon in [23] developed a computer program of bootstrapping mediation analysis. He used a pair bootstrap regression in obtaining statistics a, b and $(a \times b)$. In its original form, the pairs bootstrap is implemented by randomly resampling the data directly with replacement. And as has been mentioned earlier, [24] also have done the same approaches in bootstrapping mediated effects. Monte Carlo evidence suggests that inference based on this procedure is not always accurate [9]. This work has motivated us to perform residuals bootstrapping algorithm based on the mediation model. Here we confine our study on the bootstrap estimate of the indirect effects. In this paper we propose a new bootstrap algorithm for the indirect effects of the simple mediation model. The proposed algorithm is based on the studentized residuals.

Residual bootstrap is very common technique in many literatures such as [10], [11], and [12]. Different residuals formulations have been proposed in the literatures, including the ordinary (raw) residuals [15], standardized residuals [13] and studentized residuals [14]. [14] has shown that the bootstrap from studentized residuals is asymptotically valid when the number of cases increases and the number of variables is fixed. [16]

proposed a bootstrap approach in least square regression to construct critical points for the studentized residuals by employing a jackknife-after-bootstrap technique.

Residual vector, $\hat{\boldsymbol{\varepsilon}}$ can be expressed in the form of hat matrix $\mathbf{H}$ or $\mathbf{W}$ as follows:

$$\begin{aligned}\hat{\boldsymbol{\varepsilon}} &= \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} \\ &= \mathbf{Y} - \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \\ &= (\mathbf{I} - \mathbf{W})\,\mathbf{Y} \\ &= (\mathbf{I} - \mathbf{H})\,\mathbf{Y}.\end{aligned}$$

Then the variance of the ith case residual $\hat{e}_i$ can be written as:

$$\begin{aligned}Var(\hat{e}_i) &= Var\{(1 - h_{ii})y_i\} \\ &= (1 - h_{ii})\sigma^2,\end{aligned}$$

where h_{ii} is the ith leverage, $i = 1,\ldots,n$. It means that the variance of residuals varies inversely with leverage. The leverage depends on x_i, and so if any leverages are particularly high, resampling the residuals will not do a good job of estimating the distribution of $\mathbf{e}$.

In the residual bootstrap resample of linear regression models, [14] proposed using studentized residual , r_i, as a function of ith leverage as follows:

$$r_i = \frac{\hat{e}_i}{\hat{\sigma}\sqrt{1 - h_{ii}}}. \tag{4}$$

Equation (4) means that residuals can be standardized to correct for unequal variance caused by unequal leverages by dividing the residual by an estimate of its standard deviation. [17] pointed out that the studentized residual can also be used to identify outliers.

In order to improve the performance of the studentized residuals, we propose to rescale the well known studentized residuals in Eq. (4). The rescaled studentized residuals is as follows:

$$\begin{aligned}r_{rest} &= r_i(h_{ii}) \\ &= \frac{\hat{e}_i}{\hat{\sigma}\sqrt{1 - h_{ii}}}(h_{ii}).\end{aligned} \tag{5}$$

We call this method as Rescaled Studentized Residual Bootstrap based on Least Squares (ReSRB).

As already known, leverage value h_{ii} can be used for identifying outliers with respect to their x values, which indicates how much a certain point is far apart from its data cloud. It measures the weight of the ith observation in terms of its importance in the model's fit. The value of h_{ii} will always be in between 0 and 1, and, technically, represents the diagonal portion of a $(n \times n)$ hat matrix, $\mathbf{H}$, where,

$$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})\mathbf{X}^{-1}. \tag{6}$$

Note that $0 \le h_{ii} \le 1$ and $\sum_{i=1}^{n} h_{ii} = k$, where k = the number of parameters in a linear regression model. If the leverage value h_{ii} is large, the ith observation is outlying with respect to its x values. Furthermore, for correcting the unequal variances caused by unequal leverages, [14] suggested using the standardized residuals, r_i, not the usual residuals. This works have motivated us to proposed the new bootstrap algorithm in the mediation model by incorporating a new improved version of studentized residuals in the bootstrapping the residuals algorithm.

4.1 The proposed bootstrap algorithm

We will incorporate the proposed ReSRB in our proposed bootstrap algorithm for estimating the indirect effects of a mediation model. As already mentioned, to the best of our knowledge, this is the first bootstrap algorithm introduced in the mediation model. The proposed bootstrap algorithm is developed by resampling the residuals from the original data instead of employing the random resampling of the data as was done by [24], [21], and [23]. According to [26], inferences based on bootstrapping by random resampling are not always accurate.

As discussed in previous section, a simple mediation analysis involves three regression equations as in Eq. (1), Eq. (2), and Eq. (3). However, in order to obtain a bootstrap mediation effects, we need to know the value of α and β which can be obtained using Eq. (2) and Eq. (3), respectively. The proposed approach, which is designed for simple mediation analysis with the presence of single or multiple high leverage points, is as follows:

1. Estimate the regression coefficients a and b, respectively from Eq. (2) and Eq. (3) for the screened original sample using least squares regression, and calculate the fitted values and residuals for each observation:

   For Equation (2), it will be:

   $$\begin{aligned}\hat{M}_i &= \hat{i}_2 + \hat{a}X_i \\ \hat{e}_{2i} &= M_i - \hat{M}_i.\end{aligned}$$

Then for rescaling the residual $\hat{e}_{2i}$, instead of using studentized residual $\hat{r}_{2i} = \frac{\hat{e}_{2i}}{\hat{\sigma}\sqrt{1-h_{ii}}}$,we give weight to the studentized residual using inverse of its leverage value. In other words, the studentized residual is magnified by its leverage value, h_{ii}. As a result, the modified scaled residual become

$$\hat{r}_{2i} = \frac{\hat{e}_{2i}}{\hat{\sigma}_{MX}\sqrt{1-h_{ii}}} h_{ii},$$

where $\hat{\sigma}_{MX}$ is the estimated standard deviation of residual of the $M_i = i_2 + aX_i + e_2$ equation.

And for Equation (3), it will be:

$$\hat{Y}_i = \hat{i}_3 + \hat{b}M_i + \hat{c}'X_i$$

$$\hat{e}_{3i} = Y_i - \hat{M}_i.$$

Similarly, its scaled residual for this equation becomes

$$r_{3i} = \frac{\hat{e}_{3i}}{\hat{\sigma}_{YMX}\sqrt{1-h_{ii}}} h_{ii},$$

where $\hat{\sigma}_{YMX}$ is the estimated standard deviation of residual of the $Y_i = i_3 + \hat{b}M_i + \hat{c}'X_i + e_3$ equation.

2. Select bootstrap samples of the residuals from screened original sample,

$$\mathbf{e}^*_{3B} = [r_{3B1}^*, r_{3B2}^*, \ldots, r_{3Bn}^*]',$$

$$\mathbf{e}^*_{2B} = [r_{2B1}^*, r_{2B2}^*, \ldots, r_{2Bn}^*]'$$

and from these, calculate bootstrapped *M* and *Y* values,

$$\mathbf{m}^*_B = [M_{B1}^*, M_{B2}^*, \ldots, M_{Bn}^*]',$$

$$\mathbf{y}^*_B = [Y_{B1}^*, Y_{B2}^*, \ldots, Y_{Bn}^*]',$$

where $M_{Bi}^* = \hat{M}_i + e_{3Bi}^*$ and $Y_{Bi}^* = \hat{Y}_i + e_{2Bi}^*$.

3. Regress the bootstrapped *M* and *Y* values on the *fixed X* and *M*, *X* values, respectively to obtain bootstrap regression coefficients needed in mediation analysis.
4. Repeat step 2 and 3 for B times to obtain B bootstrap replications.

4.2 Simulation design

The performance of the new method of Rescaled Studentized Residual Bootstrap (ReSRB) is examined on simulated data to further validation. A series of simulation was conducted.

We simulated a clean data (0% outliers) sized *n* with design structure as follows. We generated $X \sim (0,20)$, $M = f(X) + e = 1.75X + e$, and $Y = f(X, M) + e = -0.1X + 0.85M + e$. The errors in the model were generated from $e \sim N(0,1)$. Then the clean data were contaminated with of 5%, 10%, 15%, 20%, 25% and 30% percentage outliers in the *X*, *M* and *Y* directions. To generate the outlying observations for each of the three variables, we delete the respective observations and replace them with their original values plus a fixed shift of 10σ distance.

About 2000 bootstrap samples were drawn from a sample size of 25, 50, 100, and 200 to represent small, medium and large sample respectively. In this simulation, we consider sample size of $n = 25$, $n = 50$, $n = 100$, and $n = 200$. The results of the simulation studies are shown in Table 1 till Table 3. We expect by doing this kind of simulated data, the performance of the proposed ReSRB method can be legitimated.

5 Simulation Results

Table 1 – Table 3 display the bias and the root mean square error (RMSE) of the indirect effect, $a \times b$, of the proposed bootstrap method, ReSRB and the other three methods at various percentage of outliers (PO) and sample size (*n*). We compared the proposed method, namely ReSRB with different methods of bootstrap. Those methods are RRB (Raw Residual Bootstrap), SRB (Studentized Residual Bootstrap), and JRB (Jackknifed Residual Bootstrap).

Our main focus is to evaluate the performances of the proposed method, namely the ReSRB, which is based on the bias and RMSE. We expect that the ReSRB can reduce bias and RMSE more than the other existing methods. A statistic used to estimate a population parameter is *unbiased* if the sampling distribution of the statistic is centered at the value of the population parameter; that is, the mean or average value of the statistic equals the true value of the parameter. An unbiased statistic gives a correct estimate, on average. We knew that the mean square error (MSE) is considered as efficiency measurement. A good estimator is the one that have desirable properties which are unbiasedness and efficiency.

From the Table 1, which is for contamination in *X* direction, we see that after $B = 2000$ bootstrap simulation, the bias and RMSE of each combination of percentage of outliers (PO) and sample size (n) produced by the ReSRB method are smaller than the RRB, SRB, and JRB. The behavior occurs on both clean data and contaminated data. But, the results are a little bit different on the contamination in M direction.

In addition, for the contaminated data, the better performance of our proposed method with regard of bias and RMSE does not only happen for small or medium percentage of outliers, but also at large percentage of outliers. Among the methods we have studied, we can say that a bad performance is shown by RRB which has the highest bias and RMSE.

Table 1 Average Bias and RMSE for the Indirect Effect, of the Simulated Data with Outliers in X

PO (%)	n	Bootstrap Methods							
		RRB		SRB		JRB		**ReSRB**	
		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
0%	25	0.0500	0.4785	0.0729	0.4008	0.0807	0.4271	**0.0145**	**0.0521**
	50	0.0070	0.1751	0.0184	0.2171	0.0101	0.2214	**0.0017**	**0.0111**
	100	-0.1005	0.2261	-0.1102	0.2352	-0.0987	0.2274	**-0.0037**	**0.0083**
	200	0.0011	0.1234	0.0034	0.1184	0.0037	0.1190	**-0.0002**	**0.0021**
5%	25	-0.0934	0.1551	-0.0302	0.0509	-0.2481	0.4001	**-0.0285**	**0.0460**
	50	-0.1035	0.1542	-0.0148	0.0238	-0.0163	0.0258	**-0.0071**	**0.0107**
	100	-0.1546	0.2327	-0.0190	0.0284	-0.0191	0.0287	**-0.0031**	**0.0045**
	200	-0.1484	0.2176	-0.0174	0.0256	-0.0179	0.0263	**-0.0014**	**0.0020**
10%	25	-0.0576	0.1162	-0.0056	0.0136	-0.0060	0.0142	**-0.0038**	**0.0062**
	50	-0.0656	0.1163	-0.0079	0.0139	-0.0085	0.0145	**-0.0012**	**0.0021**
	100	-0.0907	0.1362	-0.0091	0.0138	-0.0092	0.0138	**-0.0007**	**0.0010**
	200	-0.0888	0.1292	-0.0101	0.0146	-0.0101	0.0146	**-0.0004**	**0.0005**
15%	25	-0.0793	0.1408	-0.0080	0.0160	-0.0097	0.0186	**-0.0013**	**0.0030**
	50	-0.0796	0.1303	-0.0091	0.0146	-0.0090	0.0147	**-0.0008**	**0.0013**
	100	-0.0688	0.1037	-0.0078	0.0118	-0.0082	0.0123	**-0.0003**	**0.0005**
	200	-0.0766	0.1111	-0.0079	0.0115	-0.0079	0.0115	**-0.0002**	**0.0002**
20%	25	-0.0407	0.0903	-0.0042	0.0092	-0.0047	0.0102	**-0.0002**	**0.0011**
	50	-0.0674	0.1082	-0.0069	0.0111	-0.0070	0.0113	**-0.0004**	**0.0006**
	100	-0.0623	0.0932	-0.0071	0.0106	-0.0071	0.0107	**-0.0002**	**0.0003**
	200	-0.0661	0.0963	-0.0064	0.0094	-0.0064	0.0094	**-0.0001**	**0.0001**
25%	25	-0.0282	0.0653	-0.0021	0.0068	-0.0019	0.0068	**-0.0003**	**0.0008**
	50	-0.0259	0.0578	-0.0014	0.0057	-0.0017	0.0059	**0.0001**	**0.0004**
	100	-0.0396	0.0630	-0.0041	0.0069	-0.0042	0.0070	**0.0000**	**0.0001**
	200	-0.0617	0.0901	-0.0066	0.0096	-0.0066	0.0096	**0.0000**	**0.0001**
30%	25	-0.0479	0.0941	-0.0042	0.0088	-0.0044	0.0091	**0.0000**	**0.0008**
	50	-0.0447	0.0727	-0.0057	0.0092	-0.0057	0.0092	**-0.0001**	**0.0003**
	100	-0.0454	0.0711	-0.0041	0.0065	-0.0042	0.0067	**0.0000**	**0.0001**
	200	-0.0612	0.0893	-0.0054	0.0079	-0.0055	0.0080	**0.0000**	**0.0000**

Table 2 Average Bias and RMSE for the Indirect Effect, of the Simulated Data with Outliers in M

PO (%)	n	Bootstrap Methods							
		RRB		SRB		JRB		**ReSRB**	
		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
0%	25	0.0500	0.4785	0.0729	0.4008	0.0807	0.4271	**0.0145**	**0.0521**
	50	0.0070	0.1751	0.0184	0.2171	0.0101	0.2214	**0.0017**	**0.0111**
	100	-0.1005	0.2261	-0.1102	0.2352	-0.0987	0.2274	**-0.0037**	**0.0083**

	200	0.0011	0.1234	0.0034	0.1184	0.0037	0.1190	**-0.0002**	**0.0021**
5%	25	**-0.0337**	**0.0542**	-0.0649	0.0961	-0.1157	0.1747	-0.0516	0.0821
	50	-0.0151	0.0319	-0.0116	0.0262	-0.0105	0.0262	**-0.0048**	**0.0100**
	100	-0.0014	0.0076	-0.0012	0.0072	-0.0010	0.0069	**-0.0003**	**0.0013**
	200	-0.0029	0.0111	-0.0026	0.0081	-0.0028	0.0082	**-0.0001**	**0.0007**
10%	25	-0.0094	0.0307	-0.0089	0.0225	-0.0091	0.0237	**-0.0035**	**0.0105**
	50	0.0043	0.0140	0.0032	0.0103	0.0033	0.0105	**0.0005**	**0.0019**
	100	-0.0031	0.0110	-0.0026	0.0082	-0.0023	0.0080	**-0.0001**	**0.0008**
	200	-0.0015	0.0059	-0.0013	0.0043	-0.0012	0.0043	**-0.0001**	**0.0002**
15%	25	-0.0008	0.0177	0.0014	0.0144	**0.0002**	0.0148	0.0015	**0.0047**
	50	0.0020	0.0041	-0.0002	0.0045	-0.0001	0.0046	**-0.0001**	**0.0007**
	100	0.0010	0.0073	0.0009	0.0054	0.0008	0.0055	**0.0000**	**0.0003**
	200	0.0007	0.0055	0.0002	0.0040	0.0003	0.0040	**0.0000**	**0.0001**
20%	25	0.0066	0.0180	0.0016	0.0090	0.0023	0.0094	**0.0003**	**0.0019**
	50	-0.0015	0.0031	0.0002	0.0035	0.0003	0.0037	**-0.0001**	**0.0004**
	100	0.0019	0.0100	0.0004	0.0067	0.0007	0.0067	**0.0000**	**0.0003**
	200	-0.0008	0.0031	-0.0007	0.0027	-0.0006	0.0027	**0.0000**	**0.0001**
25%	25	0.0055	0.0246	-0.0021	0.0117	-0.0016	0.0125	**-0.0006**	**0.0020**
	50	0.0034	0.0145	0.0025	0.0115	0.0038	0.0124	**0.0003**	**0.0011**
	100	-0.0025	0.0171	-0.0026	0.0103	-0.0027	0.0103	**-0.0001**	**0.0005**
	200	0.0014	0.0023	-0.0006	0.0021	-0.0006	0.0020	**0.0000**	**0.0000**
30%	25	0.0014	0.0068	-0.0002	0.0035	0.0000	0.0035	**0.0000**	**0.0005**
	50	-0.0011	0.0099	-0.0009	0.0077	-0.0010	0.0075	**-0.0001**	**0.0005**
	100	-0.0014	0.0039	-0.0004	0.0029	-0.0005	0.0029	**0.0000**	**0.0001**
	200	-0.0033	0.0049	0.0000	0.0005	0.0000	0.0005	**0.0000**	**0.0000**

Table 3 Average Bias and RMSE for the Indirect Effect, of the Simulated Data with Outliers in Y

PO (%)	n	**Bootstrap Methods**							
		RRBLS		SRBLS		JRBLS		**ReSRBLS**	
		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
0%	25	0.0500	0.4785	0.0729	0.4008	0.0807	0.4271	**0.0145**	**0.0521**
	50	0.0070	0.1751	0.0184	0.2171	0.0101	0.2214	**0.0017**	**0.0111**
	100	-0.1005	0.2261	-0.1102	0.2352	-0.0987	0.2274	**-0.0037**	**0.0083**
	200	0.0011	0.1234	0.0034	0.1184	0.0037	0.1190	**-0.0002**	**0.0021**
5%	25	-4.5868	9.0741	-0.3223	0.6413	-4.3843	8.0193	**-0.0305**	**0.0668**
	50	-1.4651	2.6002	-0.1019	0.1866	-0.1470	0.2529	**-0.0071**	**0.0134**
	100	4.2181	6.2273	0.2544	0.3813	0.2859	0.4264	**0.0070**	**0.0106**
	200	-1.3757	2.2588	-0.0776	0.1271	-0.0835	0.1373	**-0.0018**	**0.0029**
10%	25	-5.7141	9.6460	-0.2413	0.4077	-0.3106	0.5287	**-0.0289**	**0.0470**
	50	8.9667	13.2568	0.3564	0.5424	0.3967	0.5986	**0.0199**	**0.0305**
	100	-2.5160	3.6832	-0.0956	0.1410	-0.0996	0.1471	**-0.0025**	**0.0039**
	200	0.8853	1.4663	0.0360	0.0636	0.0371	0.0653	**0.0008**	**0.0015**
15%	25	9.5893	14.6621	0.4639	0.7759	0.5383	0.8838	**0.0720**	**0.1106**
	50	1.1876	3.0290	0.0411	0.1085	0.0431	0.1157	**0.0027**	**0.0100**
	100	1.5392	2.4013	0.0571	0.0917	0.0577	0.0929	**0.0023**	**0.0040**
	200	5.7833	8.2671	0.1900	0.2747	0.1922	0.2783	**0.0030**	**0.0043**
20%	25	12.3292	17.9899	0.3066	0.5137	0.3261	0.5439	**0.0516**	**0.0769**
	50	-8.0347	11.6951	-0.2750	0.4054	-0.2802	0.4126	**-0.0193**	**0.0289**
	100	5.8611	8.7454	0.2072	0.3140	0.2095	0.3161	**0.0052**	**0.0082**
	200	1.0894	1.6822	0.0383	0.0589	0.0372	0.0574	**0.0004**	**0.0010**

25%	25	2.3127	7.4833	0.1143	0.2484	0.1236	0.2644	**0.0156**	**0.0311**
	50	8.5629	12.7035	0.2755	0.4200	0.2836	0.4341	**0.0206**	**0.0306**
	100	2.1040	3.6496	0.0791	0.1397	0.0823	0.1444	**0.0027**	**0.0054**
	200	-2.2683	3.4293	-0.0711	0.1072	-0.0714	0.1079	**-0.0012**	**0.0020**
30%	25	-4.7326	9.5680	-0.1147	0.2384	-0.1227	0.2525	**-0.0171**	**0.0397**
	50	-3.5949	8.1575	-0.0975	0.2333	-0.0991	0.2323	**-0.0080**	**0.0196**
	100	-1.2551	2.9218	-0.0329	0.0765	-0.0338	0.0778	**-0.0013**	**0.0039**
	200	5.1383	7.4558	0.1284	0.1908	0.1285	0.1908	**0.0017**	**0.0027**

The advantages of the ReSRB over other methods are even more apparent in data sets with medium or large percentage of outliers. In summary, the results from the experiments indicate that the ReSRB procedure performs well in most of the given situations.

6 Conclusion

Resampling approaches such as bootstrap for testing mediation effects hold considerable promise for mediation analysis. There is evidence that the resampling methods generally have more accurate Type I error rates and more statistical power than single sample methods that assume a normal distribution for the mediated effect [18].

Our proposed ReSRB method has provides an alternative to the bootstrap for estimating the mediated effects in mediation models. It is mainly based on residual bootstrap. Empirical result and the simulation study have shown that the ReSRB is better than the other methods. The rescaling term of h_i to the studentized residuals introduced in the ReSRB procedure help to improve the performance of the bootstrapped estimates. Our conclusion is based on its performances with regard to bias and root of mean square error (RMSE).

The advantages of the ReSRB over other methods are even more apparent in data sets with medium or large percentage of outliers. In summary, the results from the simulation experiments indicate that the ReSRB method performs well in most of the given situations.

References:

[1] James, L. R., & Brett, J. M. . Mediators, moderators, and tests for mediation. *Journal of Applied Psychology, 69*, 1984, pp.307-321.

[2] MacKinnon, D. P., & Dwyer, J. H. Estimating mediated effects in prevention studies. Evaluation Review, 17(2), 1993, 144-158.

[3] Efron, B. The bootstrap and modern statistics. *Journal of the American Statistical Association, 95*, 2000, 1293–1296.

[4] Biebler, K.E., Michael W., and Kerstin W.. Bootstrap evaluation of mixed data type discriminant analysis. *WSEAS Transactions on Systems.* 3(8), 2004, pp.2743-5.

[5] Zaharim, A. Ibrahim M.A.A., and Mohd. S.Y. An Evaluation of Test Statistics for Detecting Level Change in BL(1,1,1,1) Models. *WSEAS TRANSACTIONS on Mathematics.* 2(7), 2008, pp. 67-70.

[6] Norazan, M.R, Habshah, M., and Imon, A.H.M.R. Estimating Regression Coefficients using Weighted Bootstrap with Probability. *WSEAS Transaction on Mathematics*, 7, 2009, pp. 362-371.

[7] Hyriena, O., Margot M. P., Mark N., and Andrei Y. The Statistical Analysis of Longitudinal Clonal Data on Oligodendrocyte Generation. Proceeding of the Proceedings of the 2006 WSEAS International Conference on Mathematical Biology and Ecology, Miami, Florida, USA, January 18-20, 2006, pp. 227-232.

[8] Efron, B. Better bootstrap confidence intervals. *Journal of the American Statistical Association, 82*, 1987, 171–185.

[9] Baron, R. M., and Kenny, D. A. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51 (6), 1986. 1173-1182.

[10] Sobel, M. E. Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology, 13*, 1982, 290–312.

[11] Shrout, P. E., and Bolger, N. Mediation in experimental and nonexperimental studies: Newprocedures and recommendations. *Psychological Methods, 7*, 2002, pp.422-445.

[12] MacKinnon, D. P., Krull, J. L., and Lockwood, C. M. Equivalence of the mediation,

confounding, and suppression effect. *Prevention Science, 1*, 2000, 173–181.
[13] MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., and Sheets, V. A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods*, 7, 2002, 83-104.
[14] Imon, A. H. M. R. and Ali, M. M. Bootstrapping Regression Residuals. *Journal of the Korean Data & Information Science Society*, Vol. 16, No. 3, 2005, pp. 665-682.
[15] Efron, B. Bootstrap methods: another look at the jacknife". *The Annals of Statistics 7*, 1979, pp.1-26.
[16] Chernick, M. R. *Bootstrap Methods: A Guide for Practitioners and Researchers. Second Ed.* John Wiley & Sons, Inc. Newtown, PA. 2007.
[17] Freedman, D. A. Bootstrapping regression models. *Annals of Statistics*, 9, 1981, 1218–1228.
[18] Wu, C. F. J. Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis. *The Annals of Statistics*, Vol. 14, No. 4, 1986, pp. 1261-1295.
[19] Efron, B. & Tibshirani, R. *An Introduction to the Bootstrap*. London: Chapman & Hall. 1993.
[20] Fox, John *Applied Regression Analysis and Generalized Linear Models. Second Ed.* Sage Publications, Ontario, Canada. 2008.
[21] MacKinnon, D. P., Lockwood C. M., & Williams, J. Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research, 39*, 2004, 99–128.
[22] Bollen, K. A., & Stine, R. Direct and indirect effects: Classical and bootstrap estimates of variability. *Sociological Methodology, 20*, 1990, 115–140.
[23] MacKinnon, D. P. *Introduction to Statistical Mediation Analysis*. New York: Taylor & Francis. 2008.
[24] Mooney, C. Z. and Duval, R. D. *Bootstrapping: A non-parametric approach to statistical inference*. Newbury Park, CA: Sage. 1993.
[25] Horowitz, J. L. The bootstrap". In *Handbook of Econometrics*, Vol. 5. *J. J. Heckman and E. E. Leamer (eds)*, Elsevier Science. 2000.
[26] Horowitz, J. L. The bootstrap. In *Handbook of Econometrics*, Vol. 5. *J. J. Heckman and E. E. Leamer (eds)*, Elsevier Science. 2000.