

The Consistency of Blindfolding in the Path Analysis Model with Various Number of Resampling

Solimun^{*}, Adji Achmad Rinaldo Fernandes

Department of Statistics, Faculty of Mathematics and Natural Sciences, Indonesia.

Received October 29, 2019; Revised March 23, 2020; Accepted March 28, 2020

Copyright©2020 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract The use of regression analysis has not been able to deal with the problems of complex relationships with several response variables and the presence of intervening endogenous variables in a relationship. Analysis that is able to handle these problems is path analysis. In path analysis there are several assumptions, one of which is the assumption of residual normality. If the normality residual assumptions are not met, then estimating the parameters can produce a biased estimator, a large and not consistent range of estimators. Unmet residual normality problems can be overcome by using resampling. Therefore in this study, a simulation study was conducted to apply resampling with the blindfold method to the condition that the normality assumption is not met with various levels of resampling in the path analysis. Based on the simulation results, different levels of closeness occur consistently at different resampling quantities. At a low level of closeness, it is consistent with the resampling magnitude of 1000. At a moderate level, a consistent level of resampling of 500 occurs. At a high level of closeness, it is consistent with the amount of resampling 1400.

Keywords Amount of Resampling, Blindfold, Consistency

1. Introduction

According to [1], path analysis is a form of multiple regression statistical analysis used to analyze causal relationships in which independent variables produce both direct and indirect effects on a dependent variable. In path analysis, when the normality assumption of residuals cannot be fulfilled, several things can be done, for example transforming data, trimming data outlier, or adding observations. In addition to the methods mentioned, one other method that can be used to overcome violations of the

normality assumption of residuals is resampling.

Resampling is a sampling activity from an existing data sample to draw a new sample with a larger size. According to the central limit theorem, the sampling distribution of the sample means approaches of a normal distribution for samples from a non-normal distribution with a sufficiently large sample size [2]. Several types of resampling include bootstrapping, jackknifing, blindfolding, k-nearest neighbor, Randomization Exact Test, and Cross-validation.

Basically, resampling is the method of repeatedly drawing samples from the original data samples. One of the most basic resampling techniques is Blindfolding. The advantage of the Blindfolding method is that it has a more flexible performance. It is because the sampling is conducted with part of the data that are set and part of the data that are omitted.

Generally, one issue to consider in resampling is the number of resampling. The number of resampling shows the number of replications of sampling taken. The more the number of resampling is, the more consistent the estimators obtained will be [3]. A simulation study can be done to investigate the number of resampling approaching the consistency of estimators in path analysis.

Based on the description above, this simulation study is conducted to examine the number of resampling on the consistency of path coefficient estimators if the normality assumption of residuals is not fulfilled under various conditions of the closeness of relationships between variables. This study aims to identify the consistency of blindfolding. Blindfolding does resampling with part of the data set. The consistency is needed for path coefficient estimation.

2. Materials and Methods

Path analysis is an extension of multiple linear regression. Path analysis, developed by Sewall Wright as in [7], is a method for studying the direct and indirect effects

of variables in which some of the variables are often taken as causes, whereas others are taken as effects. Meanwhile, according to [8], path analysis is defined as an extension of the regression model used to test the fit of the correlation matrix against two or more causal models which are being compared by the researcher. The model is usually depicted in a circle-and-arrow figure in which single-headed arrows indicate causation. A regression is done for each variable in the model as a dependent on others which the model indicates are causes. The regression weights predicted by the model are compared with the observed correlation matrix for the variables, and a goodness-of-fit statistic is calculated [8].

In path analysis, the terms exogenous and endogenous variables are used. An exogenous variable is one whose variation is assumed to be determined by causes outside the model. It means that exogenous variables are not determined in the model. An endogenous variable, conversely, is one whose variation is explained by exogenous or other endogenous variables in the model. Endogenous variables are divided into intervening endogenous variables and pure endogenous variables. An intervening variable is one whose variation is explained by other variables as well as being the explanatory to other variables. Meanwhile, a pure endogenous variable is one that only explains the variation of other variables. In path analysis, there is at least one exogenous variable, one intervening endogenous variable, and one pure endogenous variable [9].

2.1. Steps of Path Analysis

According to Solimun (2010), the steps in path analysis are:

1. Designing a model based on concepts and theories. The designed model will be presented in the form of a path diagram or in the form of an equation. For the equation model, because path analysis consists of several equations, a system of equations will be formed.
2. Checking assumptions of path analysis.
3. Calculating path coefficients. According to Solimun (2010), there are three methods used to calculate path coefficients:
 - a. A correlation matrix, if the model formed is not tiered ($p = R_{x-1}R_y$)
 - b. Regression coefficients, followed by mathematical calculations.
 - c. Standardized regression coefficients. In this third method, the goodness of fit in the form of the coefficient of determination can be calculated simply, and the theory trimming can be implemented more easily.
4. Checking model validity.
5. Interpreting analysis results.

2.2 Steps of Path Analysis

Path analysis distinguishes three types of effects [10] with the following explanations.

2.2.1. The Direct Effect

The direct effect occurs when the relationship between the exogenous variable and the endogenous variable is not mediated by the intermediary variable. The direct effect can be described as follows.

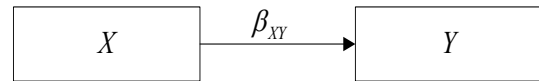


Figure 1. The Direct Effect

Based on Figure 1, it can be seen that the magnitude of the direct effect of the exogenous variable X on the endogenous variable Y is β_{XY} .

2.2.2. The Indirect Effect

The indirect effect occurs when the relationship between the exogenous variable and the endogenous variable is mediated by the intermediary variable. The indirect effect can be described as follows.

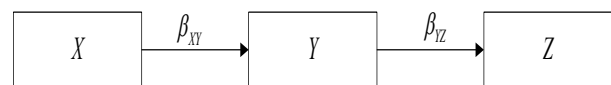


Figure 2. The Indirect Effect

Based on Figure 2, it can be seen that the exogenous variable X has an effect on the endogenous variable Z through the endogenous variable Y . The magnitude of the indirect effect is calculated by multiplying the direct effect of X on Y with the direct effect of Y on Z . Mathematically, the indirect effect in this example is $\beta_{XY} \times \beta_{YZ}$.

2.2.3. The Total Effect

The total effect is the sum of the direct and indirect effects. From the examples in Figures 1 and 2., the total effect is $\beta_{XY} + \beta_{YZ} + (\beta_{XY} \times \beta_{YZ})$.

2.3. A Path Diagram

One important component in path analysis is a path diagram. A path diagram is used to illustrate the causal relationships between variables [7]. In the path diagram, the direct and indirect effects of each variable can be seen.

An example of the relationship between the exogenous variable and the endogenous variable can be illustrated in the following path diagram.

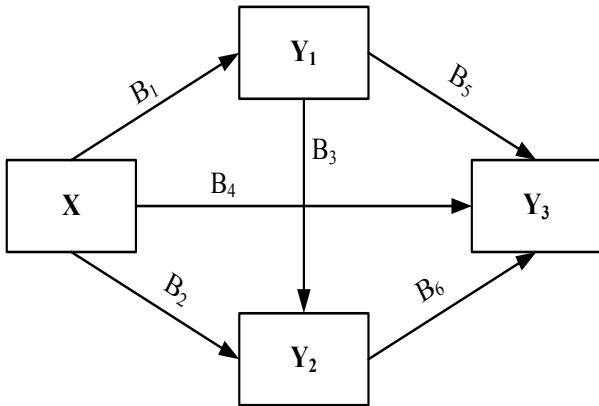


Figure 3. A Path Diagram with One Exogenous Variable, Two Intervening Endogenous Variables, and One Pure Endogenous Variable

In Figure 3., the variable X is an exogenous variable, while the variables Y_1 , Y_2 and Y_3 are endogenous variables. The causal relationship in the path diagram is drawn in the direction of the arrows. In Figure 3, it can be seen that the variable Y_1 is influenced by the variable X , the variable Y_1 influences the variable Y_2 and the variable Y_3 , the variable Y_2 is influenced by the variable X and the variable Y_1 , and the variable Y_2 influences the variable Y_3 . Thus, it can be said that the variables Y_1 and Y_2 are intervening endogenous variables. The variable Y_3 is influenced by variables X , Y_1 and Y_2 , but does not influence other variables. Thus, it can be said that the variable Y_3 is a pure endogenous variable.

Before making a model, the variables in path analysis need to be standardized in order to have the same means and variations, so that the path coefficients obtained can have the same units and can be compared.

The following is a form of standardization conducted on the exogenous variable X [4].

$$Z_{X_i} = \frac{X_i - \bar{X}}{S} \text{ with } S = \sqrt{\frac{(X_i - \bar{X})^2}{n-1}} \quad (1)$$

Where:

Z_{X_i} : the value of the standardized variable X for the i^{th} observation

X_i : the value of the variable X for the i^{th} observation

\bar{X} : the mean value of the variable X

S : the standard deviation of the variable X

n : sample size

By doing standardization, each variable will follow the standard normal distribution with mean = 0 and variance = 1.

The path analysis model consists of a system of equations. This model can be formed based on a path

diagram. The system of equations needs to be solved simultaneously, starting from parameter estimation, hypothesis testing, to interpretation.

The system of equations obtained from the path diagram in Figure 3. is shown in equation (2).

$$\begin{aligned} Z_{Y1i} &= \beta_{Z_X Z_{Y1}} Z_{X_i} + \varepsilon_{Z_{Y1i}} \\ Z_{Y2i} &= \beta_{Z_X Z_{Y2}} Z_{X_i} + \beta_{Z_{Y1} Z_{Y2}} Z_{Y1i} + \varepsilon_{Z_{Y2i}} \\ Z_{Y3i} &= \beta_{Z_X Z_{Y3}} Z_{X_i} + \beta_{Z_{Y1} Z_{Y3}} Z_{Y1i} + \beta_{Z_{Y2} Z_{Y3}} Z_{Y2i} + \varepsilon_{Z_{Y3i}} \end{aligned} \quad (2)$$

The matrix form above can be written as follows.

$$\begin{bmatrix} Z_{Y11} \\ \vdots \\ Z_{Y1n} \\ Z_{Y21} \\ \vdots \\ Z_{Y2n} \\ Z_{Y31} \\ \vdots \\ Z_{Y3n} \end{bmatrix} = \begin{bmatrix} Z_{X1} & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ Z_{Xn} & 0 & 0 & 0 & 0 & 0 \\ 0 & Z_{X1} & Z_{Y11} & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & Z_{Xn} & Z_{Y1n} & 0 & 0 & 0 \\ 0 & 0 & 0 & Z_{X11} & Z_{Y11} & Z_{Y21} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & Z_{X1n} & Z_{Y1n} & Z_{Y2n} \end{bmatrix} \begin{bmatrix} \beta_{Z_X Z_{Y1}} \\ \beta_{Z_X Z_{Y2}} \\ \beta_{Z_{Y1} Z_{Y2}} \\ \beta_{Z_X Z_{Y3}} \\ \beta_{Z_{Y1} Z_{Y3}} \\ \beta_{Z_{Y2} Z_{Y3}} \end{bmatrix} + \begin{bmatrix} \varepsilon_{Z_{Y11}} \\ \vdots \\ \varepsilon_{Z_{Y1n}} \\ \varepsilon_{Z_{Y21}} \\ \vdots \\ \varepsilon_{Z_{Y2n}} \\ \varepsilon_{Z_{Y31}} \\ \vdots \\ \varepsilon_{Z_{Y3n}} \end{bmatrix}$$

$$\underline{Z} = \underline{Z}\beta + \underline{\varepsilon} \quad (3)$$

2.4. Blindfolding Resampling

Characteristics that describe a population are called parameters. Information about the population is needed to estimate parameters. However, in the application, not all observations in the population can be known. It can be caused by several factors, for example, considerable cost and time that are required to obtain complete data. Thus, sampling is more often done.

If the number of observations taken as a sample is insufficient, parameter estimation with the sample becomes less accurate. To overcome this problem, a tool departing from theoretical distributions can be used, i.e. resampling. As the name implies, the resampling method is done by repeated sampling within the same sample. Resampling is used in hypothesis testing. However, resampling can be used to bring up all possible combinations. It is certainly highly time-consuming, so the computation is required [11].

One resampling method is the blindfolding method. Blindfolding is a sample re-use technique, which systematically deletes data points and provides a prognosis of their original values. For this purpose, the procedure requires an omission distance D . A value for the omission distance D between 5 and 12 is recommended in literature [11]. An omission distance of seven ($D=7$) implies that every fifth data point of a latent variable's indicators will be eliminated in a single blindfolding round. Since the blindfolding procedure has to omit and predict every data

point of the indicators used in the measurement model of the selected latent variable, an omission distance of $D=7$ results in seven blindfolding rounds. Hence, the number of blindfolding rounds always equals the omission distance. Blindfolding employs a resampling algorithm that creates a number of resamples by a method whereby each resample has a certain number of rows replaced with the means of the respective columns (variable). Replications are performed on the first row, continued on the second row until the last row. The number of rows modified in this way in each resample equals the sample size divided by the number of resamples. For example, if the sample size is 108 and the number of resamples selected is 108, then each resample will have 1 row modified. As with the bootstrapping method, the Blindfolding method will obtain convergent sample estimators on at least 100 replications [5].

For example, the variable X has seven sizes of $X = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$ and 5 sizes are taken in the resampling process. The following is the resampling process in the blindfolding method:

- a. The first resampling $x_1^* = \{x_1, x_2, x_4, x_3, x_2\}$
- b. The second resampling $x_2^* = \{x_1, x_2, x_3, x_5, x_5\}$
- c. The third resampling $x_3^* = \{x_1, x_2, x_3, x_5, x_7\}$

Based on this process, it can be seen that in each blindfold sample, x_1 and x_2 will always be taken at every repeated sample taken. The Blindfolding Process is illustrated in Figure 4.

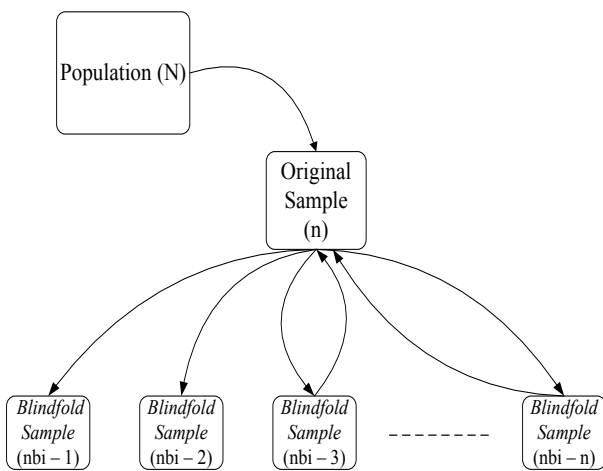


Figure 4. The Blindfolding Process

2.5. Path Coefficient Estimation

A path coefficient indicates the magnitude of the direct effect of an exogenous variable on an endogenous variable in a system of equations. One method that can be used to estimate path coefficients is Ordinary Least Squares (OLS).

2.5.1. The Ordinary Least Squares (OLS) Method

The Ordinary Least Squares (OLS) method is used when the model is linear in parameters. The principle of the OLS method is by minimizing the sum of squared residuals. Based on equation (2.3), $Z = Z\beta + \varepsilon$ can be written as $\varepsilon = Z - Z\beta$. Thus, the sum of squared residuals can be written as $Q = \varepsilon^T \varepsilon$. The OLS method minimizes the following functions.

$$\begin{aligned} \min \{Q\} &= \min \{ \varepsilon^T \varepsilon \} \\ &= \min \left\{ (Y - Z\beta)^T (Y - Z\beta) \right\} \end{aligned} \quad (4)$$

Completion of equation optimization (4) is by minimizing Q to β .

$$\hat{\beta} = (Z^T Z)^{-1} Z^T Y \quad (5)$$

2.5.2. Hypothesis Testing and Consistency with Blindfolding

The blindfolding steps to estimate the standard error are as follows:

- a. Determining the number of F times in the blindfold sample $(x_1^*, x_2^*, \dots, x_F^*)$ obtained from random sampling with returns of n elements from the original sample (x_1, x_2, \dots, x_n) , with x_1 to x_4 are set.
- b. Calculating the related blindfolding replications for each blindfold sample.
- c. Estimating the standard error using the standard deviation for blindfolding replicated F times with the following formula:

$$\begin{aligned} \hat{\beta}^*(f) &= s(x_f^*); f = 1, 2, \dots, F \\ Se_F &= \sqrt{\frac{\sum_{f=1}^F [\hat{\beta}^*(f) - \hat{\beta}^*(.)]^2}{(F-1)}} \end{aligned} \quad (6)$$

with $\hat{\beta}^*(.) = \frac{\sum_{f=1}^F \hat{\beta}^*(f)}{F}$

with the B value approaches infinity, \hat{se}_B is an ideal estimator for $Se_{\hat{\beta}}(\hat{\beta}^*)$.

$$\lim_{B \rightarrow \infty} \hat{se}_B = se_{\hat{\beta}}(\hat{\beta}^*) \quad (7)$$

The use of resampling method causes the data to be free from distribution so the assumption of normally distributed data and a large sample are not required. Thus, hypothesis testing can be done using the t-test on each coefficient obtained using the following formula.

$$t_i = \frac{\hat{\beta}_i^*}{\hat{se}_B(\hat{\beta}_i^*)}, i = 1, 2, \dots, k \quad (8)$$

In equation (8), k shows the number of estimated coefficients. The statistical hypothesis used is as follows.

$$\begin{aligned} H_0 : \hat{\beta}_i &= 0 \\ H_1 : \hat{\beta}_i &\neq 0 \end{aligned}, i = 1, 2, \dots, k$$

If the t -test statistic obtained is greater than t_{table} , then the null hypothesis is rejected. It means that there is a significant effect between variables. In this case, the effect between variables is the effect of the exogenous variable on the endogenous variable and the effect of the endogenous variable on another endogenous variable.

According to [6], the consistency of an estimator in blindfolding can be shown by the bias value, which is the difference (distance) between the estimator and the parameter. The following is the formula for calculating the estimator bias in resampling for the relationship between the exogenous variable X and the endogenous variable Y_1 .

$$bias_F = bias_F(\hat{\beta}_{XY_1}, \beta_{XY_1}) = |E_F[s(x)] - t(F)| \quad (9)$$

Good resampling results will follow a Monte Carlo simulation approach based on means.

$$\hat{\beta}^*(.) = \sum_{b=1}^B \frac{\hat{\beta}^*(b)}{B} = \sum_{b=1}^B \frac{s(x_b^*)}{B} \quad (10)$$

In equation (10), $\hat{\beta}^*(.)$ is the mean of parameter estimators obtained from the resampling process. Thus, the bias based on the replica B is to replace $E_F[s(x)]$ with $\hat{\beta}^*(.)$.

$$bias_F = \|\hat{\beta}^*(.) - t(F)\| \quad (11)$$

The bias in equation (11) can be used to determine the consistency of estimators obtained from blindfold samples.

2.6. Methods

This research is a simulation study using data generation with the following criteria.

- 1) Cross-section data with four variables, i.e. one exogenous variable, two intervening endogenous variables, and one pure endogenous variable. The four variables were measured directly (observable variables), so they did not require a measurement model. Thus, data had interval or ratio scales.
- 2) The exogenous variable was set by the standardization of $\bar{x} \pm 2s$, where $\bar{x} = 0$ and $s = 1$. The distance between observations on the exogenous variable was made the same.
- 3) Intervening and pure endogenous variables were calculated through a linear regression function with

three path coefficient variations. The path coefficient with the range of $|0,05 - 0,25|$ represented the low closeness of the relationship, the range of $|0,30 - 0,50|$ represented the moderate closeness of the relationship, and the range of $|0,60 - 0,90|$ represented the high closeness of the relationship.

- 4) The software used in this research was R software.

2.7. Forming Simulation Data

The following are the steps conducted in this study to form simulation data.

1. Setting data for the exogenous variable

The exogenous variable (X) was set by the standardization of $\bar{x} \pm 2s$, where $\bar{x} = 0$ and $s = 1$. Three sample sizes (n) were used, i.e. n small = 25, n medium = 50, and n large = 100. The distance between observations on the exogenous variable was made the same using arithmetic progression of $a_n = a + (n - 1)b$.

2. Making a path diagram

A path diagram illustrates the causal relationships between variables. In this study, the path diagram as shown in Figure (3) was used. Furthermore, a system of equations was formed according to the path diagram using standardized variables as in equation (16).

3. Making the path coefficient values from data generation

The path coefficient (β) values were $-1 < 0 < 1$ due to the standardization of the exogenous variable. In this study, the path coefficient values were divided into three conditions, i.e. those that describe the low ($|0,05 - 0,25|$), moderate ($|0,30 - 0,50|$), and high ($|0,60 - 0,90|$) closeness of the relationship. Thus, path coefficients can be generated by a uniform distribution with lower and upper limits adjusted to the conditions of the closeness of the relationship.

4. Generating the residual values

In this study, the residual \mathcal{E} used followed an exponential distribution. It aimed to find out the resampling results on data which was not fulfilled the normality assumption of residuals.

5. Forming data for the intervening and pure endogenous variables

The intervening endogenous variables (Y_1) and (Y_2), as well as the pure endogenous variable (Y_3) were calculated based on a system of equations formed in step two.

6. Forming simulation data

Simulation data was obtained by combining the exogenous variable, the intervening endogenous variables, and the pure endogenous variable obtained.

Next, the following steps were taken to do resampling with the Blindfolding method on simulation data.

1) Path coefficient estimation on simulation data

Path coefficient estimation on simulation data was performed using the OLS method. From this process, six path coefficients were obtained.

2) Sampling using the blindfolding method

Resampling was conducted using the blindfolding method. Ten variations of the number of resampling (B) were used, i.e. 100, 250, 500, 750, 1000, 1200, 1400, 1600, 1800, and 2000.

3) Path coefficient estimation in each blindfold sample

Path coefficient estimation was conducted on the B sample set. Thus, this process generated path coefficients $\hat{\beta}^*(b) = (\hat{\beta}_{XY_1}^*(b), \hat{\beta}_{XY_2}^*(b), \hat{\beta}_{XY_3}^*(b), \hat{\beta}_{Y_1Y_2}^*(b), \hat{\beta}_{Y_1Y_3}^*(b), \hat{\beta}_{Y_2Y_3}^*(b))$ with $b = 1, 2, \dots, B$. After that, average path coefficient was calculated

notated as $\hat{\beta}^* = (\hat{\beta}_{XY_1}^*(.), \hat{\beta}_{XY_2}^*(.), \hat{\beta}_{XY_3}^*(.), \hat{\beta}_{Y_1Y_2}^*(.), \hat{\beta}_{Y_1Y_3}^*(.), \hat{\beta}_{Y_2Y_3}^*(.))$

4) Hypothesis testing of blindfolding

Hypothesis testing was performed for each path coefficient value.

5) Calculation of the bias

The bias is the difference (distance) between the estimator and the parameter. There were six estimators and six parameters in this study. The formula for calculating the bias is shown in equation (2.21). In equation (2.21), the

bias can be calculated using the vector norm. Estimators were represented by a point

$$A(\hat{\beta}_{XY_1}^*(.), \hat{\beta}_{XY_2}^*(.), \hat{\beta}_{XY_3}^*(.), \hat{\beta}_{Y_1Y_2}^*(.), \hat{\beta}_{Y_1Y_3}^*(.), \hat{\beta}_{Y_2Y_3}^*(.))$$

Meanwhile, parameters were represented by a point $B(\beta_{XY_1}, \beta_{XY_2}, \beta_{XY_3}, \beta_{Y_1Y_2}, \beta_{Y_1Y_3}, \beta_{Y_2Y_3})$. Thus, the bias was calculated as the distance between the two points written as

$$d = \sqrt{(\hat{\beta}_{XY_1}^*(.) - \beta_{XY_1})^2 + (\hat{\beta}_{XY_2}^*(.) - \beta_{XY_2})^2 + (\hat{\beta}_{XY_3}^*(.) - \beta_{XY_3})^2 + (\hat{\beta}_{Y_1Y_2}^*(.) - \beta_{Y_1Y_2})^2 + (\hat{\beta}_{Y_1Y_3}^*(.) - \beta_{Y_1Y_3})^2 + (\hat{\beta}_{Y_2Y_3}^*(.) - \beta_{Y_2Y_3})^2}$$

3. Results and Discussion

The resampling process with the Blindfolding method was performed on simulation data whose residuals did not fulfill the normality assumption. In this case, violations of the normality assumption of residuals were represented by exponentially distributed residuals. The following were the calculation results of the bias at various levels of the closeness of the relationship between variables, various conditions of the number of resampling, and different sample sizes.

3.1. Low Level of the Closeness of the Relationship between Variables

The level of the closeness of the relationship between variables included in the low category was indicated by the path coefficient value in a range of 0.05 - 0.25.

3.1.1. Sample Size $n = 25$

Calculation of the bias was performed using the vector norm. In Figure 5 below, a comparison of the bias values in each condition of the number of resampling is presented.



Figure 5. The Bias of Path Coefficient Estimators under Unfulfilled Normality Assumption Condition, the Low Closeness of the Relationship, And $n = 25$ With Various Number of Resampling

Based on Figure 5, it can be seen that the bias values get smaller as the number of resampling increases. The difference in the bias at $B = 750$ and the bias at $B = 1000$ was very small of only 0.0448. After that, the bias value experienced a quite low decrease. Thus, it can be said that the consistency of path coefficient estimators under unfulfilled normality assumption condition and the low closeness of the relationship between variables was reached at the number of resampling of 750.

3.1.2. Sample Size $n = 50$

Calculation of the bias was performed using the vector norm. In Figure 6, a comparison of the bias values in each condition of the number of resampling is presented.

Based on Figure 6, it can be seen that the bias values get smaller as the number of resampling increases. The difference in the bias at $B = 1000$ and the bias at $B = 1200$ was very small of only 0.0372. After that, the bias value experienced a quite low decrease. Thus, it can be

said that the consistency of path coefficient estimators under unfulfilled normality assumption condition and the low closeness of the relationship between variables was reached at the number of resampling of 1000.

3.1.3. Sample Size $n = 100$

Calculation of the bias was performed using the vector norm. In Figure 7, a comparison of the bias values in each condition of the number of resampling is presented.

Based on Figure 7, it can be seen that the bias values get smaller as the number of resampling increases. The difference in the bias at $B = 1000$ and the bias at $B = 1200$ was very small of only 0.0276. After that, the bias value experienced a quite low decrease. Thus, it can be said that the consistency of path coefficient estimators under unfulfilled normality assumption condition and the low closeness of the relationship between variables was reached at the number of resampling of 1000.

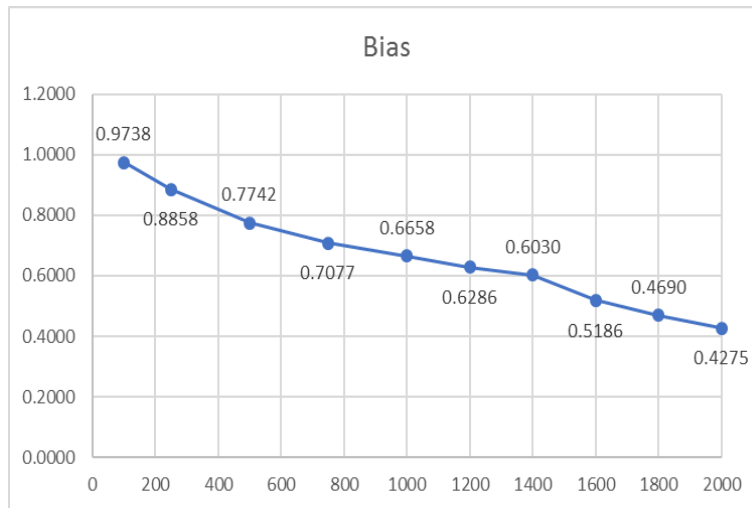


Figure 6. The Bias of Path Coefficient Estimators under Unfulfilled Normality Assumption Condition, the Low Closeness of the Relationship, And $n = 50$ With Various Number of Resampling

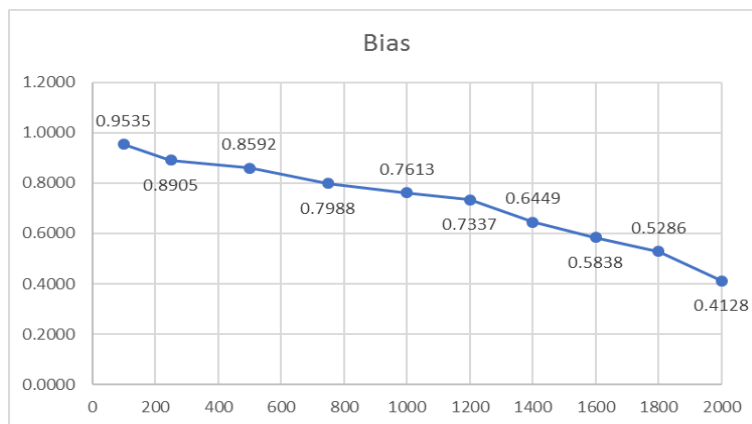


Figure 7. The Bias of Path Coefficient Estimators under Unfulfilled Normality Assumption Condition, the Low Closeness of the Relationship, And $n = 100$ With Various Number of Resampling

3.2. Moderate Level of the Closeness of the Relationship between Variables

The level of the closeness of the relationship between variables included in the moderate category was indicated by the path coefficient value in a range of 0.30 - 0.50.

3.2.1. Sample Size $n = 25$

Calculation of the bias was performed using the vector norm. In Figure 8, a comparison of the bias values in each condition of the number of resampling is presented.

Based on Figure 8, it can be seen that the bias values get smaller as the number of resampling increases. The difference in the bias at $B = 500$ and the bias at $B = 750$ was very small of only 0.0345. After that, the bias value experienced a quite low decrease. Thus, it can be said that the consistency of path coefficient estimators under

unfulfilled normality assumption condition and the moderate closeness of the relationship between variables was reached at the number of resampling of 500.

3.2.2. Sample Size $n = 50$

Calculation of the bias was performed using the vector norm. In Figure 9, a comparison of the bias values in each condition of the number of resampling is presented.

Based on Figure 9, it can be seen that the bias values get smaller as the number of resampling increases. The difference in the bias at $B = 1000$ and the bias at $B = 1200$ was very small of only 0.0367. After that, the bias value experienced a quite low decrease. Thus, it can be said that the consistency of path coefficient estimators under unfulfilled normality assumption condition and the moderate closeness of the relationship between variables was reached at the number of resampling of 1000.



Figure 8. The Bias of Path Coefficient Estimators under Unfulfilled Normality Assumption Condition, the Moderate Closeness of the Relationship, And $n = 25$ With Various Number of Resampling



Figure 9. The Bias of Path Coefficient Estimators under Unfulfilled Normality Assumption Condition, the Moderate Closeness of the Relationship, And $n = 50$ With Various Number of Resampling

3.2.3. Sample Size $n = 100$

Calculation of the bias was performed using the vector norm. In Figure 10, a comparison of the bias values in each condition of the number of resampling is presented.

Based on Figure 10, it can be seen that the bias values get smaller as the number of resampling increases. The difference in the bias at $B = 500$ and the bias at $B = 750$ was very small of only 0.0483. After that, the bias value experienced a quite low decrease. Thus, it can be said that the consistency of path coefficient estimators under unfulfilled normality assumption condition and the moderate closeness of the relationship between variables was reached at the number of resampling of 500.

3.3. High Level of the Closeness of the Relationship between Variables

The level of the closeness of the relationship between variables included in the high category was indicated by the path coefficient value in a range of 0.6 - 0.9.

3.3.1. Sample Size $n = 25$

Calculation of the bias was performed using the vector norm. In Figure 11, a comparison of the bias values in each condition of the number of resampling is presented.

Based on Figure 11, it can be seen that the bias values get smaller as the number of resampling increases. The difference in the bias at $B = 1400$ and the bias at $B = 1600$ was very small of only 0.0541. After that, the bias value experienced a quite low decrease. Thus, it can be said that the consistency of path coefficient estimators under unfulfilled normality assumption condition and the high closeness of the relationship between variables was reached at the number of resampling of 1400.

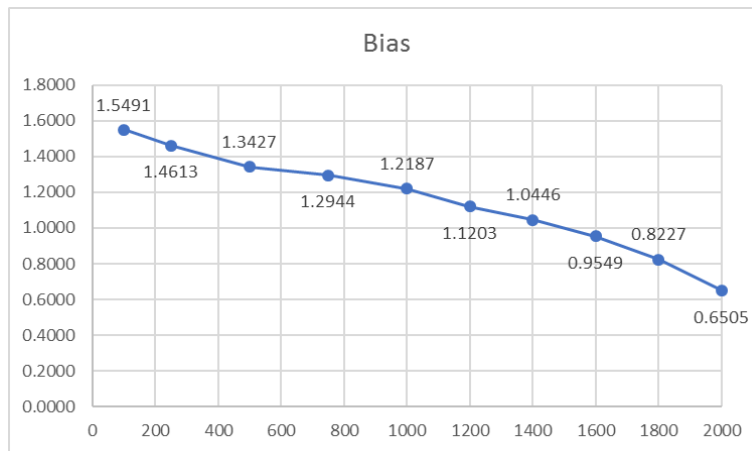


Figure 10. The Bias of Path Coefficient Estimators under Unfulfilled Normality Assumption Condition, the Moderate Closeness of the Relationship, And $n = 100$ With Various Number of Resampling

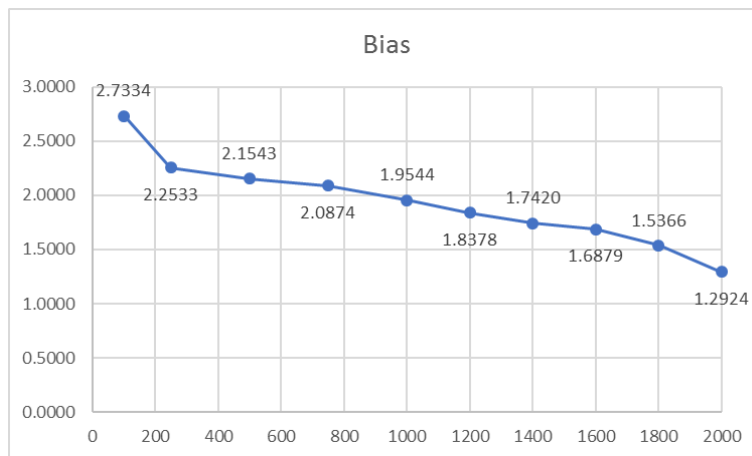


Figure 11. The Bias of Path Coefficient Estimators under Unfulfilled Normality Assumption Condition, the High Closeness of the Relationship, And $n = 25$ With Various Number of Resampling

3.3.2. Sample Size $n = 50$

Calculation of the bias was performed using the vector norm. In Figure 12, a comparison of the bias values in each condition of the number of resampling is presented.

Based on Figure 12, it can be seen that the bias values get smaller as the number of resampling increases. The difference in the bias at $B = 1400$ and the bias at $B = 1600$ was very small of only 0.0257. After that, the bias value experienced a quite low decrease. Thus, it can be said that the consistency of path coefficient estimators under unfulfilled normality assumption condition and the high closeness of the relationship between variables was reached at the number of resampling of 1400.

3.3.3. Sample Size $n = 100$

Calculation of the bias was performed using the vector norm. In Figure 13, a comparison of the bias values in each condition of the number of resampling is presented.

Based on Figure 13, it can be seen that the bias values get smaller as the number of resampling increases. The difference in the bias at $B = 1000$ and the bias at $B = 1200$ was very small of only 0.0584. After that, the bias value experienced a quite low decrease. Thus, it can be said that the consistency of path coefficient estimators under unfulfilled normality assumption condition and the high closeness of the relationship between variables was reached at the number of resampling of 1400.

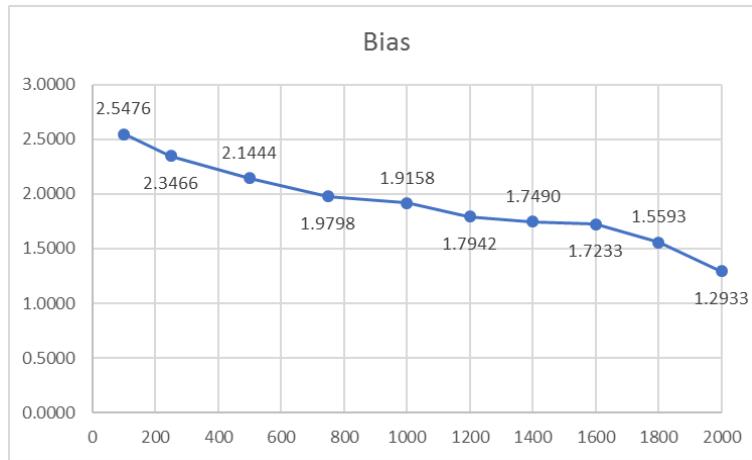


Figure 12. The Bias of Path Coefficient Estimators under Unfulfilled Normality Assumption Condition, the High Closeness of the Relationship, And $n = 50$ With Various Number of Resampling

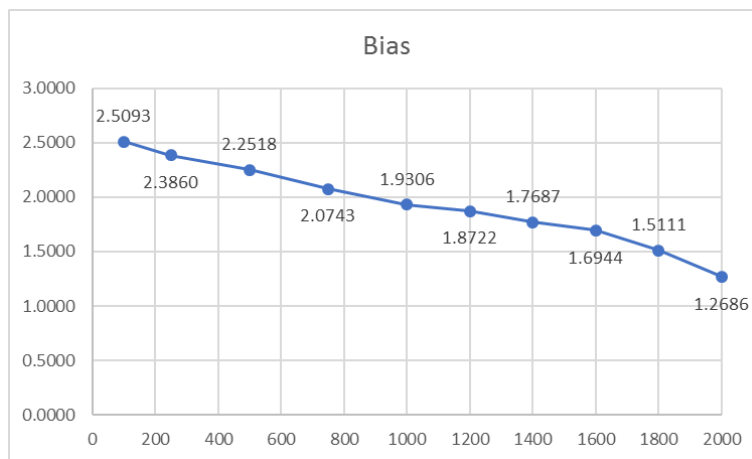


Figure 13. The Bias of Path Coefficient Estimators under Unfulfilled Normality Assumption Condition, the High Closeness of the Relationship, And $n = 100$ With Various Number of Resampling

3.4. Consistency of Blindfolding under Unfulfilled Normality Assumption Condition

This section presents a comparison of the bias values obtained under the unfulfilled normality assumption of residuals condition. In Table 1 below, the number of resampling able to reach the consistency of path coefficient estimators for each condition is presented.

Table 1. The Number of Resampling Resulting in Consistent Path Coefficient Estimators under Various Conditions

Closeness of Relationship Between	Large Sample	Scale Resampling
Low (0,05 – 0,25)	25	750
	50	1000
	100	1000
Medium (0,30 – 0,50)	25	500
	50	1000
	100	500
High (0,60 – 0,90)	25	1400
	50	1400
	100	1000

In Table 1, it can be seen that the consistency was achieved in the low closeness of the relationship between variables with a sample size of 25 at the number of resampling of 750, a sample size of 50 at the number of resampling of 1000, and a sample size of 100 at the number of resampling of 1000. The consistency was achieved in the moderate closeness of the relationship between variables with a sample size of 25 at the number of resampling of 500, a sample size of 50 at the number of resampling of 1000, and a sample size of 100 at the number of resampling of 500. The consistency was achieved in the high closeness of the relationship between variables with a sample size of 25 at the number of resampling of 1400, a sample size of 50 at the number of resampling of 1400, and a sample size of 100 at the number of resampling of 100.

4. Conclusions

Based on the simulation study results, it can be concluded that:

- 1) The number of resampling with the blindfolding method needed to achieve consistent path coefficient estimators in the data with exponentially distributed residuals for the condition of the low closeness of relationships between variables was 1000.
- 2) The number of resampling with the blindfolding method needed to achieve consistent path coefficient estimators in the data with exponentially distributed residuals for the condition of the moderate closeness of relationships between variables was 500.
- 3) The number of resampling with the blindfolding method needed to achieve consistent path coefficient

estimators in the data with exponentially distributed residuals for the condition of the high closeness of relationships between variables was 1400.

REFERENCES

- [1] Rutherford, R.D. 1993. Statistical Model For Causal Analysis. John Wiley & Sons Inc, New York.
- [2] Gujarati, D. N., dan Porter, D. C. 2004. Basic econometrics (ed.) McGraw-Hill. Irwin, a business.
- [3] Efron, B. dan Tibhirani, R.J. 1993. *An Intoduction to the Bootstrap*. US: Springer.
- [4] Solimun, A. A. R. Fernandes, dan Nurjannah. 2017. Metode Statistika Multivariat: Pemodelan Persamaan structural (SEM) perndekatan WarpPLS. Malang: UB Press.
- [5] Kock, N. 2015. Common method bias in PLS-SEM: A full collinearity assessment approach. *International Journal of e-Collaboration (IJeC)*, 11(4), 1-10.
- [6] Mammen, E. 1993. Bootstrap and Wild Bootstrap for High Dimensional Linear Models. *The Annals of Statistics*. Vol. 21 No. 1. Hal. 255-285.
- [7] Dillon, W. R., dan Goldstein, M. 1984. *Multivariate analysis methods and applications* (No. 519.535 D5).
- [8] Garson, G. D. 2013. Path analysis. Asheboro: Statistical Associates Publishing
- [9] Pedhazur, E. J. 1973. *Multiple Regression in Behavioral Research Explanation and Prediction*, Second Edition. New York: Holt, Rinehart and Winston, Inc.
- [10] Solimun. 2010. Metode Partial Least Square, CV. Citra Malang, Malang.
- [11] Yu, C. H. 2003. *Resampling methods: concepts, applications, and justification, Practical Assessment, Research and Evaluation*, 8(19).
- [12] Hair, J. F., Hult, G. T. M., Ringle, C. M., and Sarstedt, M. (2017). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*, 2nd Ed., Sage: Thousand Oaks