Journal of Engineering and Applied Sciences 14 (Special Issue 2): 5541-5546, 2019
ISSN: 1816-949X

Mediation Effect in Bootstrap Logistic Regression

Shrook A.S. AL-Sabbah and Enas Abdu Alhafudh Mohammed
Department of Statistics, Administration and Economics College,
University of Kerbala, Karbala, Iraq
shorouqa.@uokerbala.edu.iq, enas.albasri@uokerbala.edu.iq

Abstract: In this study applied the method of mediation (indirect effects) which is based on the method of regression analysis and multiple correlations between variables, it was also part of the total effects of independent variables on the dependent variable. The research adopted the method of causal steps to estimate the parameters of the binary logistic regression according to resampling (Bootstrap approach) to extracting the bootstrap estimates values. The data analyzed based on the broad money supply and some variables affecting it, namely the volume of the currency in circulation and the current deposits using the SPSS program. The result showed significant relation between the broad money supply and the independent variables (the volume of the currency) and the (current deposits). Therefore, the estimations can be adopted for using the estimated equation for the forecasting the broad money supply in the future and we concluded that the use of mediation allows the possibility of clarifying the indirect effects of the explanatory variables. The current deposits transferred the indirect effect of the explanatory variables on the broad money supply it is thus, an intermediate variable and the type of mediation is partial.

Key words: Bootstrap approach, mediation, logistic regression, variable, money, intermediate

INTRODUCTION

The aim of the research is to propose the use of mediation in the bootstrap logistic regression model to test significance of the effect of the mediation variable which contributes to the interpretation of the total effects of the explanatory variables on the dependent variable.

Kazem (2003) used the Bootstrap method for a multi-regression model in quantitative data. While in this study, we suggest the effect of mediation in the Bootstrap logistic regression model when quantitative data were not available (we have qualitative variable). The mediation is the transferring of the effect of the explanatory variable on the dependent variable by another variable or several other variables. This is what is called a mediator.

A regression relationship is a relationship between two or more variables, one of which is a dependent variable and the rest of the other variables are explanatory variables. Fitrianto and Midi (2013) hence, the model may contain two explanatory variables, one of which may be an intermediate variable on the dependent variable and thus, we can know mediation as the indirect effect of the first explanatory variable. (Indirect effect) as the mean variable interferes with the effect as in the following form It is noticed from Fig. 1 that there are two paths, direct and

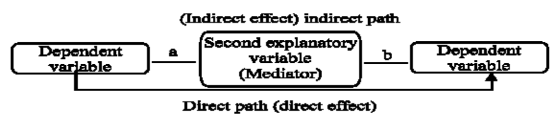

Fig. 1: Direct and indirect effect

indirect, of the effect of the first explanatory variable on the adopted variable, the mediation is two types (Mackinnon *et al.*, 2012):

Complete mediator: The first explanatory variable shows no effects on the dependent variable after the control of the second explanatory variable.

Partial mediation: The direct effect (the first explanatory variable on the dependent variable) is low but it is still not equal to zero when we inter the second explanatory variable (mediator). Here, it noted here that when there is one variable in the logistic regression model, we have mediation but if there is two available then it called multiple mediation. This research will be limited to simple mediation (Muller *et al.*, 2008).

Test of mediation variables: The presence of the mediator in the mediation models does not necessarily mean its significance. Therefore, the mediation variable is tested by the method of causal steps.

Corresponding Author: Shrook A.S. AL-Sabbah, Department of Statistics, Administration and Economics College, University of Kerbala, Karbala, Iraq, shorouqa.@uokerbala.edu.iq

Method of causal steps: This method was proposed in 1986 by researchers Baron and Kenny and is based on four steps as follows (Precher and Hayes, 2004):

The first step: Formation of logistic regression model between the dependent variable and the first explanatory variable to choose the significance of the latter as follows:

$$\text{In}\left(\frac{p_i}{1-p_i}\right) = \text{logit}\quad E(y/x) = y_i = \beta_{01}+\beta_{11}x_i+U_{i1} \qquad (1)$$

The second step: Formation of the logistic regression model is simple between the intermediate variable as a dependent variable and the first explanatory variable to test the significance of the latter as follows:

$$\text{In}\left(\frac{p_i}{1-p_i}\right) = \text{logit}\quad E(M/x) = M_i = \beta_{02}+\beta_{12}x_i+U_{i2} \qquad (2)$$

The third step: The composition of the simple logistic regression model between the dependent variable and the mean variable for the last significant selection as follows:

$$\text{In}\left(\frac{p_i}{1-p_i}\right) = \text{logit}\quad E(y/M) = y = \beta_{03}+\beta_{13}M+U_{i3} \qquad (3)$$

The fourth step: Formation of the model of logistic regression between the dependent variable and the explanatory variables and the medium to choose the significance of explanatory and intermediate variables is as follows:

$$\text{In}\left(\frac{p_i}{1-p_i}\right) = \text{logit}\quad E(y/x, M)\ \ y = \beta_{04}+\beta_{14}x_i+\beta_{24}M+U_{i4} \qquad (4)$$

MATERIALS AND METHODS

Logistic regression model: The logistic model of the logarithm of odds ratio, derived from the logistic regression model, deals with the analysis of a particular phenomenon to find the factors that influence it in order to predict its occurrence or non-occurrence in light of the availability of information about that problem, thus, the logistic regression model can be defined as one of the nonlinear regression models that characterize conventional models limited to estimating linear models only because they are more flexible, since, any (arbitrary) relationship between dependent and independent variables is assumed (Shaver, 2005). This model is based on the logistic curve function which is expressed as follows:

$$\text{Pi} = \frac{e^{(\dot{x}\theta+z\gamma)}}{1+e^{(\dot{x}\theta+z\gamma)}} = \frac{1}{1+e^{(\dot{x}\theta+z\gamma)}}$$

where, Pi represents, event probability (response). If Yi represents the response state of the dependent variable, Y takes the value (1) or (0) as follow:

$$Y_i = \begin{cases} 1 & \text{when response occur} \\ 0 & \text{otherwise} \end{cases}$$

Thus, the response variable (the dependent variable) can follow the Bernoulli distribution with the probability of Pi succeeding when Y = 1 and the probability of 1-Pi failure when Y = 0 and according to Bernoulli's distribution:

$$\left.\begin{aligned} E(Yi) &= 1(Pi)+0(1-Pi) \\ E(Yi) &= Pi \end{aligned}\right\}$$

where, Pi is success probability, $0\leq Pi\leq 1$, thus, $0\leq E(Yi)\leq 1$, then:

$$P_i = P_r(Yi = 1)\frac{e^{(\dot{x}\theta+z\gamma)}}{1+e^{(\dot{x}\theta+z\gamma)}}$$

$$(1-Pi) = P_r(Yi = 0) = \frac{1}{1+e^{(\dot{x}\theta+z\gamma)}}$$

Thus, the logit will then be In(Pi/1-Pi) which can be expressed in the following linear equation (Al-Hamdi, 2003):

$$\text{In}\left(\frac{P_i}{1-P_i}\right) = \text{logit}\{E(Y|x, z)\} = \dot{x}\theta+\dot{z}\gamma$$

Now, let μ = In (Pi/1-Pi) with limits $-\infty<\mu<+\infty$ then:

$$\mu = \dot{x}\theta+\dot{z}\gamma \rightarrow \text{Pi} = \frac{e^{(\mu)}}{1+e^{(\mu)}}$$

μ represents a linear regression equation whose parameters can be estimated.

Let $\beta = (\theta, \gamma)$ parameter vector with dimension p+q because that the value of partial vectors has rank (qx1). Thus, (px1) explains the amount by which the logarithm of the variance changes due to the change of the event due to one unit change of the independent variable when other independent variables are stable. If group S includes all the explanatory variables for exclusion from the model in the response model, the large model would be:

$$\text{logit } P(Y=1) = Bo + B_1X_1 + \sum_{s \in S} B_S Z_S$$

When S is an empty set $S = \varnothing$, the smallest response model containing the main explanatory variables including the Xo variable which conjugate with the constant, then the model be as following:

$$\text{logit } P(Y=1) = Bo + B_1\dot{X}_1$$

Goodness of fit of the model: This method shows the Goodness of fit statistic of the model which containing all the independent variables and the statistic (R^{-2} Nagelkerke and R^2 Cox and Snell). It aims at determining the variance ratio explained in the logistic regression model and expressed as follows (Hrcks and Tingley, 2011):

$$R^2_{cox} = 1 - \left(\frac{Lo}{L\beta}\right)^{\frac{N}{2}}$$

$$R^2_N = \frac{R^2}{MaxR^2}, \quad MaxR^2 = 1 - (Lo)^{\frac{N}{2}}$$

Bootstrap approach: The basic idea of the choice of the model according to this method depend on the theory of statistical sampling to recognize the appropriate model while the former method is based on the theory of information and specifically on the information function which measures the amount of information provided by the unknown parameter. This method is based on sampling with replacement with the return of any retested sampling unit before the next unit that follows. It is used for several purposes, especially in estimating the frequency of selection of the model indicated by it (Π_B) and in measuring the accuracy of estimating the parameters of the unknown model. It is an arithmetic method for estimating standard error and bias and can be considered a type of Monte-Carlo method using repetition by using the sampling with replacement to the real X data to generate the bootstrap distribution (approximate or theoretical) according to the format in Fig. 2.

The technique is according to principle: Drawing a partial random samples ($x^*_1, x^*_2, ..., x^*_{bt}$) are representative of the bootstrap samples from true data $x_1, x_2, ..., x_n$, their number equals (bt) which are at least 20 times, then we estimate the parameter by using the maximum likelihood method for each model ($J = 1, ..., R$) and for each model j can be calculated ($j = 1, 2, ..., R$) for each bootstrap sample and for R models as the following Fig. 3.

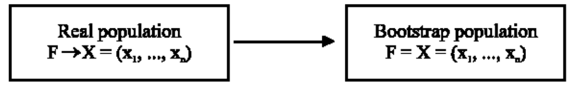

Fig. 2: Bootstrap approach:

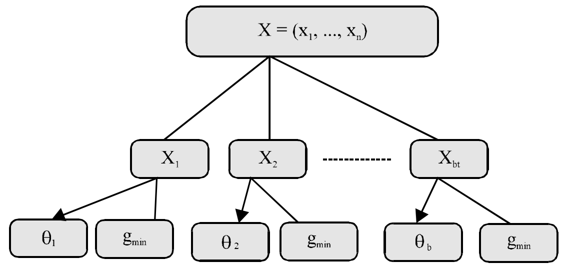

Fig. 3: Samples of bootstrap

Let the real distribution (actual) has unknown mean (μ) and unknown variance (σ^2) and the traditional estimation for (μ) is mean of sample μ_n.

In case of bootstrap, the theoretical distribution of $x_1, x_2, ..., x_n$ and by resampling for data assuming ($x^*_1, x^*_2, ..., x^*_{bt}$) are independent, then the distribution have μ^*_n which is represent the estimated for μ_n. According to the Central Limited Theory (C.L) and assuming resampling of (bt) of the times, the distribution of the bootstrap distribution is calculated as the approximation of the real (actual) distribution:

$$\sqrt{n}(\mu^*_n - \mu_n) \rightarrow \sqrt{n}(\mu_n - \mu)$$

Applied side: The research was based on a simple random sample of the central Bank's data on the broad money supply, the volume of the currency and the current deposits for the year 2010-2015, to estimate the indirect effect (mediation) of the independent variables on the dependent variable. Variables can be defined as follows (Hadi and Sheibah, 2002).

Broad money supply: The narrow money supply of all sectors (excluding the central government) is represented by commercial banks and is symbolized in the statistical analysis of Y_I.

The currency of circulation: The total amounts traded in the economy of the country of the currency of cash as well as money between people and the central bank to collect information on criticism of its importance in determining prices and their impact on inflation as well as the movement of trading and symbolized in the statistical analysis of X.

Current deposits: These deposits are deposited by individuals and institutions at banks, so that, they can be withdrawn at any time without prior notice and symbolized in the statistical analysis of M.

RESULTS AND DISCUSSION

Analysis of data and results: This study presents the result of the analysis by using the statistical program SPSS for analyzing the data according to the method of causation steps and using the bootstrap method of the binary logistic model.

First step: significance of model: Table 1 shows the significance of the relationship between the dependent variable (the broad money supply), the independent variable (the volume of the currency) and the value of the R_2 by the independent variable (the volume of the currency), i.e., approximately 78% of the variance in the dependent variable is 0.783.

The percentage of change in the value of the likelihood is (186.667) which represents the percentage of the likelihood of the independent variable (volume of currency), the bootstrap parameters and standard errors have significance as shown in the following Table 2.

The Table 2 shows the significance of independent variable (the volume of currency) and the estimated function is:

$$\hat{y} = -7.596+5.229X_i$$

Second equation: significance of model:

$$\text{In}\left(\frac{p_i}{1-p_i}\right) = \text{logit } E(M/X) = M_i = \beta_{02}+\beta_{12}X_iU_{i2}$$

Table 3 shows the significance of the relationship between current deposits and the independent variables (the volume of the currency). The R^2 is equal to 0.399, i.e., approximately 40% of the variance of the median variable (current deposits) and is explained by the independent variable (volume of currency). The percentage change in the value of the likelihood is 15.812 for the independent variable (the volume of the currency), the bootstrap parameters and its standard errors have significant and confidence intervals as following Table 4.

From the Table 4 we note the significance of the independent variable (the volume of the currency) and the estimated equation is:

$$\hat{M} = 3.411+2.761X_i$$

Third equation: significance of model:

$$\text{In}\left(\frac{p_i}{1-p_i}\right) = \text{logit } E(Y/M =) Y_i = \beta_{03}+\beta_{13}X_iU_{i3}$$

Table 5 shows the significance of the relationship between (broad money supply) and the independent variable (current deposits). The R^2 is equal to 0.752, i.e., approximately 75% of the variance of the median variable is explained by the independent variable.

Table 1: One-way ANOVA components

Models	Sum of squares	df	Mean square	F-values	Sig.
Regression	6241564978.220	1	6241564978.220	548.781	0.000
Residual	796144858.655	70	11373497.981		
Total	7037709836.875	71			

Table 2: Significant of bootstrap parameters and standard errors

		Bootstrap[a]				
					Confidence interval (95 %)	
Parameters	B	Bias	SE	Sig. (2-tailed)	Lower	Upper
X	5.229	2.968	6.749	0.001	3.866	24.203
Constant	-7.596	-3.805	9.157	0.001	-45.152	-5.527

[a]Significant values

Table 3: One-way ANOVA components

Models	Sum of squares	df	Mean square	F-values	Sig.
Regression	1172648745.479	1	1172648745.479	101.924	0.000
Residual	805357864.396	70	11505112.349		
Total	1978006609.875	71			

Table 4: Bootstrap for variables in the equation

		Bootstrap[a]				
					Confidence interval (95 %)	
Parameters	B	Bias	SE	Sig. (2-tailed)	Lower	Upper
X	2.761	0.434	2.604	0.001	1.645	4.632
Constant	-3.411	-.430	2.666	0.001	-5.782	-1.740

[a]Significant values

Table 5: One-way ANOVA components

Models	Sum of squares	df	Mean square	F-values	Sig.
Regression	6213088553.019	1	6213088553.019	527.413	0.000
Residual	824621283.856	70	11780304.055		
Total	7037709836.875	71			

Table 6: Bootstrap for variables in the equation

		Bootstrap[a]				
					Confidence interval (95 %)	
Parameters	B	Bias	SE	Sig. (2-tailed)	Lower	Upper
M	3.912	2.221	6.054	0.001	2.646	22.908
Constant	-6.438	-4.364	12.048	0.001	-44.111	-4.186

[a]Significant values

Table 7: One-way ANOVA components

El	Sum of squares	df	Mean square	F-values	Sig.
Regression	7036734339.282	2	3518367169.641	248865.129	0.000
Residual	975497.593	69	14137.646		
Total	7037709836.875	71			

Table 8: Bootstrap for variables in the equation

		Bootstrap[a]				
					Confidence interval (95 %)	
Parameters	B	Bias	SE	Sig. (2-tailed)	Lower	Upper
X	21.952	2.569	5.968	0.001	21.043	39.881
m	20.854	1.788	5.724	0.001	18.373	39.014
Constant	-64.759	-6.970	13.414	0.001	-102.284	-63.238

[a]Significant values

The percentage change in the value of the likelihood is 50 for the independent variable (current deposits), the bootstrap parameters and its standard errors have significant and confidence intervals as following Table 6. From the Table 6 we note the significance of the independent variable (current deposits) and the estimated equation is:

$$\hat{Y}_i = 6.438+3.912M_i$$

Fourth equation: significance of model

$$\text{In}\left(\frac{pi}{1-pi}\right) = \text{logit } E(Y/M, X =) Y_i = \beta_{04}+\beta_{14}X_i+\beta_{24}M_i+U_{i3}$$

Table 7 shows the significance of the relationship between (the broad money supply) and the independent variable (current deposits). The R^2 is equal to 0.899, i.e., approximately 90% of the variance of the dependent variable is explained by the independent variable.

The percentage change in the value of the likelihood is 3.419 for the independent variable, the bootstrap parameters and its standard errors have significant and confidence intervals as shown in the following Table 8.

From the Table 8 we note the significance of the independent variable (current deposits) and the estimated equation is:

$$\hat{Y}_i = -64.769+21.952Xi+20.854M_i$$

CONCLUSION

Through the research, the following conclusions were reached: as a result of the above, we conclude that the variable of current deposits represents an intermediate variable that transfers the indirect effect of the volume of the currency variable and the type of mediation is partial mediation because the first explanatory variable (the volume of the currency) affects the dependent variable (broad money supply) after the control of the second explanatory variable (current deposits).

The use of mediation allows the possibility of clarifying the indirect effects of the explanatory variables which that it is part of the method of analysis of the path. The ratio of the current deposits variable and the volume of the currency in the determining factor is approximately 90%, i.e., 90% of the total deviations are explained by the independent variables while 10% of total deviations are explained by random error.

The significance of the relationship between the broad money supply and the explanatory variables is the volume of currency and the current deposits is significant according to F value which means we depend on the estimations for the forecasting b the broad money supply for the coming years.

REFERENCES

Al-Hamdi, F.H.H., 2003. Marketing dimensions of the social responsibility of the organizations and their impact on consumer satisfaction. Ph.D Thesis, University of Mustansiriya, Baghdad, Iraq.

Fitrianto, A. and H. Midi, 2013. Standardized simple mediation model: A numerical example. World Appl. Sci. J., 22: 1135-1139.

Hadi, H.A. and Q.B. Shleibah, 2002. Advanced Economic Measurement Theory and Practice. Duniaal-Amal Library, Baghdad, Iraq,.

Hicks, R. and D. Tingley, 2011. Causal mediation analysis. Stata J., 11: 605-619.

Kazem, M.H., 2003. Bootstrap in the analysis of regression models: Practical application. Ph.D Thesis, Baghdad University, Baghdad, Iraq.

Mackinnon, D.P., J.W. Cheong and A.G. Pirlott, 2012. Statistical Mediation Analysis. American Psychological Association, Washington, DC., USA.,.

Muller, D., V.Y. Yzerbyt and C.M. Judd, 2008. Adjusting for a mediator in models with two crossed treatment variables. Organizational Res. Methods, 11: 224-240.

Preacher, K.J. and A.F. Hayes, 2004. SPSS and SAS procedures for estimating indirect effects in simple mediation models. Behav. Res. Meth. Instrum. Comput., 36: 717-731.

Shaver, J.M., 2005. Testing for mediating variables in management research: Concerns, implications and alternative strategies. J. Manage., 31: 330-353.