



EndoS: An SPSS macro to assess endogeneity

Ahmad Daryanto ^a  

^aDepartment of Marketing, Lancaster University Management School, Lancaster University, UK

Abstract ■ In this tutorial, I demonstrate how to address endogeneity with instrumental variable (IV) regression using my newly developed EndoS macro for SPSS. EndoS implements the generalized instrumental variable estimator in the form of a two-stage least square (TSLS) estimator and is capable of performing three endogeneity diagnostic tests indispensable in empirical IV regression papers: Hausman’s specification tests, overidentifying restriction tests and weak instrument tests. I present two illustrative examples using data from published studies.

Keywords ■ Endogeneity, diagnostic test for endogeneity, weak instrument. **Tools** ■ EndoS, SPSS.

Acting Editor ■ Roland Pfister (Universität Würzburg)

Reviewers ■ Two anonymous reviewers.

 a.daryanto@lancaster.ac.uk

 [10.20982/tqmp.16.1.p056](https://doi.org/10.20982/tqmp.16.1.p056)

Introduction

One of the challenges faced by researchers who use regression analysis is that their hypothesized model should not suffer from an endogeneity problem. This problem arises when an independent variable in a regression model correlates with the regression error term. A variable that possesses this characteristic is commonly termed as an endogenous variable. The most common source of endogeneity is when an influential variable is omitted from the model creating a so-called omitted variable bias.¹ In fact, omitted variable is considered to be the most common source of endogeneity (Abdallah, Goergen, & O’Sullivan, 2015; Papies, Ebbes, & Van Heerde, 2017; Sajons, 2020). The endogeneity problem can be remedied by applying a regression method called an instrumental variable regression (IV regression, for short; Stock & Watson, 2015).² Recent editorials or publications in management/applied psychology that discuss endogeneity and IV estimation as a possible solution are Antonakis, Bendahan, Jacquart, and Lalive (2010); Bascle (2008); Bollen (2012); Gennetian, Magnuson, and Morris (2008); Ketokivi and McIntosh (2017); Papies et al. (2017); Reeb, Sakakibara, and Mahmood (2012); Rutz and Watson (2019); Shaver (2019) and Sajons (2020).

Instrumental variable (IV) regression requires the use

of an external variable that is not part of the core model. This variable is labelled as an instrumental variable, or instrument for short. An instrument separates the part of the endogenous variable that is not correlated with the error term and uses this part instead in the regression model to explain the dependent variable. In doing so, an instrument must be correlated with an endogenous variable (the relevance condition), and not be correlated with the regression error term, i.e., an instrument must be exogeneous. The latter implies that an instrument must be related to the dependent variable only via the endogenous variable (i.e., the exclusion restriction condition; Stock & Watson, 2015).

The relationship between an endogenous variable X , and instrumental variable Z , an omitted variable O and a dependent variable Y can be illustrated by a mediation model as depicted in Figure 1 where the relationship among the variables is assumed to be linear. As can be seen in Figure 1, information about the omitted variable O is kept in the regression error term U if OLS regression is applied to explain Y from X . Note that the model in Figure 1 can be estimated using path or structural equation modelling (SEM) with maximum likelihood estimation where the error term for X and the error term for Y —due to an omitted variable, should be allowed to covary (Maydeu-Olivares, Shi, & Rosseel, 2019).

The IV regression method, to the best of my knowl-

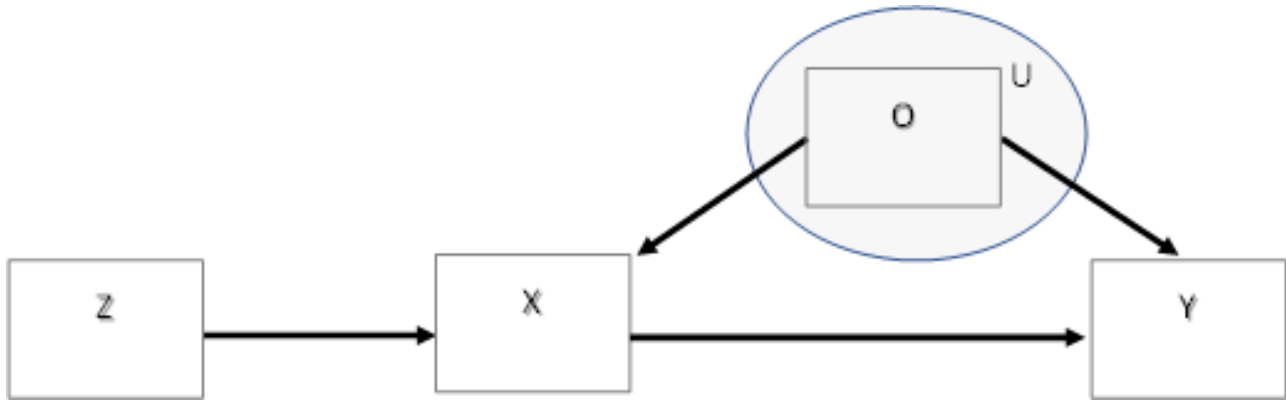
¹Other sources are simultaneity and measurement error (Stock & Watson, 2015).

²Others statistical methods are Gaussian copula (Park & Gupta, 2012), control function approach (Petrin & Train, 2010), latent instrumental variable approach (Ebbes, Wedel, Steerneman, & Bockenholt, 2005), and the higher moments approach (Lewbel, 2012). These methods are beyond the scope of this paper.

³The author has taught quantitative research methods for marketing students for more than 10 years using SPSS and, so far, has survived his teaching



Figure 1 ■ Instrumental variable (Z) in a mediation model. See text for details.



edge³, is rarely taught in the quantitative modules for graduate students in psychology, business and management. There are two reasons for the lack of teaching. Firstly, the textbooks commonly used in the quantitative modules for these students do not sufficiently cover the IV regression method. For instance, the popular multivariate statistics books by Tabachnick and Fidell (2017), Lattin, Carroll, and Green (2003), and Hair, Black, Anderson, and Babin (2018), which are widely taught at the graduate level in psychology, business and management, do not cover IV regression at all. Secondly, SPSS is the typical software used in these modules, and has very limited features in performing IV regression. For instance, SPSS ver. 25 has a ‘two-stage least squares’ option, but its outputs are very basic and do not provide essential information usually required by journals to assess endogeneity. For example, diagnostic tests for endogeneity in instrumental variable estimation covered in this paper (also known as specification tests for endogeneity, Hausman’s specification tests, overidentifying restriction tests and weak instrument tests) are not available in the SPSS two-stage least square (TSLS; another popular notation in the econometric literature is 2SLS) regression option. For these reasons, I have created an SPSS macro called EndoS, which allows newcomers in IV regression and seasoned SPSS researchers to undertake IV regression analysis. Specifically, I hope that it helps students to learn the implementation of IV regression and the interpretations of the IV regression outputs. Furthermore, with the use of EndoS, it will help researchers to present results for diagnostic tests for endogeneity in instrumental variable estimation, especially with the reporting of weak instrument tests that have limited discussion in recent psychology, business and management journals (e.g., Abdallah

et al., 2015; Jean, Deng, Kim, & Yuan, 2016; Rutz & Watson, 2019; Zaefarian, Kadile, Henneberg, & Leischnig, 2017).

In the next section, I will explain the main features of EndoS and show users how to use the macro using two illustrative examples. For the first illustration, I use the Mroz.sav data (Mroz, 1987; Wooldridge, 2000) which is discussed in one of the chapters in Wooldridge’s econometrics textbook about IV regression (Wooldridge, 2016).⁴ For the second illustration, I replicate part of the results obtained by Acemoglu, Johnson, and Robinson (2001) using data used in their paper.

Estimation procedure used in EndoS

The common form of IV regression is the two-stage least squares (TSLS). As the name suggests, TSLS procedures are conducted in two stages. For simplicity, let us consider a regression model that has one endogenous variable.

$$Y = \beta_0 + \beta_1 X + u \tag{1}$$

where X is an endogenous variable that is correlated with the error term u and β_0 and β_1 are the regression slopes. In the first stage of TSLS, X is regressed on an instrument Z .

$$X = \pi_0 + \pi_1 Z + v \tag{2}$$

where Z must be uncorrelated with the error term v and π_0 and π_1 are regression slopes. In this stage, the predicted score for X , i.e., $\hat{X} = \hat{\pi}_0 + \hat{\pi}_1 Z$ is calculated by using regression estimates ($\hat{\pi}_0$, $\hat{\pi}_1$) obtained from the ordinary least squares (OLS) regression. In the second stage, Y is regressed on \hat{X} using OLS, i.e.,

$$Y = \gamma_0 + \gamma_1 \hat{X} + \varepsilon \tag{3}$$

career without covering the IV regression method.

⁴The data in a STATA file format easily, converted into an SPSS file format, can be downloaded at <https://www.stata.com/texts/eacsap>



where γ_0 , γ_1 are called the TSLS estimates. The two-stages procedure described above is not the one usually employed in any statistical or econometric software. In fact, the two-step procedure should never be done manually because that would yield incorrect standard errors (Stock & Watson, 2015). Rather, the general formula in a matrix form is used. Like any other econometric software, EndoS uses the generalized IV estimator, also commonly known as the two-stage least square (TSLS) estimator, which takes the following matrix form (Stock & Watson, 2015):

$$\hat{B}_{IV} = (X^T P_Z X)^{-1} X^T P_Z y \quad (4)$$

where P_Z is an idempotent projection matrix that takes the following form:

$$P_Z = Z(Z^T Z)^{-1} Z^T \quad (5)$$

The matrix Z in the above equation contains instruments. The matrix X includes both the endogenous variables and covariates assumed to be exogenous. EndoS distinguishes between external and internal instruments.⁵ Z are external instruments that are truly exogenous whereas internal instruments refer to covariates that can act as their own instruments so as to achieve model identification (i.e., variables that are used as instruments such as variable ‘age’ discussed in illustration 2 later).⁶ EndoS computes the variance of the \hat{B}_{IV} by applying the following formula:

$$V(\hat{B}_{IV}) = MSE(X^T P_Z X)^{-1} \quad (6)$$

where MSE is the mean square of residuals obtained from calculating the average squared difference between the estimated values and the actual values. The estimated values are calculated using the TSLS estimates, \hat{B}_{IV} .

Diagnostic tests for endogeneity in EndoS

Model specification tests, also called diagnostic tests, are vital in endogeneity studies. There are three fundamental tests covered by EndoS, which assess the adequacy of the IV regression models: (1) Hausman’s specification test, (2) the overidentifying restriction test, and (3) the weak instrument test. I explain these three tests in a separate subsection below.

Hausman’s specification test. The aim of the Hausman’s specification test is to test whether X is truly endogenous. It is also framed in a textbook as a procedure to test whether IV regression is a better estimator compared to

the ordinary least square (OLS) estimator. If OLS is better than IV regression, i.e., produces smaller variances than that of IV, hence better precision, then one does not necessarily use IV regression (Wooldridge, 2016). As the name suggests, the test was originally developed by Hausman (1978) who suggested directly comparing and testing the regression coefficients from OLS and IV regressions. If the regression coefficients obtained from OLS regression are not significantly different from those obtained from IV regression, then the IV regression is not necessary because, as explained above, the OLS variances are smaller than the IV variances.

The direct estimates comparison approach can be achieved via another attractive computational technique—commonly known as the Durbin-Wu-Hausman test (or the omitted variables version of the Hausman test; Kennedy, 2003) which produces equivalent results (Wooldridge, 2016). The alternative method tests whether a variable is, or set of independent variables are, endogenous. The rationale is that if the OLS and IV regression produce significantly different estimates, then the suspected independent variables are endogenous. EndoS produces results for testing the null hypothesis that variables that are suspected to be endogenous are indeed endogenous.

To test for the endogeneity of a set of independent variables, EndoS conducts two-stage OLS regression using residuals as the independent variables. In the first stage, each endogenous variable is regressed on the independent variables and the external instruments. EndoS stores the residuals from the first stage regressions in a matrix to be used in the second stage. In the second stage, the dependent variable is regressed on the independent variables and residuals obtained from stage 1. EndoS has a maximum capacity of storing 10 residuals named as `resid1`, `resid2`, ..., `resid10` in the outputs. The number of residuals corresponds to the number of endogenous variables that appear in the model. This threshold can be easily increased by changing the number in the EndoS syntax file, but it is quite rare in empirical studies that one will have more than 10 endogenous variables, thus 10 is kept as the maximum. EndoS produces an F test in the case of one endogenous variable and joint F test if there are multiple endogenous variables. If the F statistic is significant then the null hypothesis of all instruments being exogeneous will be rejected.

Overidentifying restriction test. The overidentifying restriction test is used to test whether extra instruments are really needed. The test requires that the extra instruments must be exogeneous or valid—the null hypothesis is that

⁵I follow the terminology used in Hill, Griffiths, and Lim (2018). Other terms known in literature for internal and external instruments are included and excluded instruments (Stock & Watson, 2015).

⁶In Stock and Watson (2015) notation, X denotes endogenous variables, Z denotes external instruments (or also termed as excluded instruments in the literature, and W denotes covariates.



all instruments are not correlated with a regression error term. However, the overidentifying restriction test does not test for the exogeneity of the instruments overall but only a part of it, i.e., exclusion restriction.⁷ This is because assessing exogeneity requires an expert judgment about the domain of the research context (Stock & Watson, 2015).

For the overidentifying restriction test, EndoS calculates the J -statistic, $J = mF$, where F is the F value obtained from the OLS regression with the TOLS residuals as the dependent variable and the instruments and exogenous variables as the independent variables, and m is the number of the instruments in the regression. In a large sample, the J -statistic follows a chi-square distribution with $m - k$ degree of freedom, where k is the number of endogenous variables in the model. If there are as many instruments as endogenous variables (i.e., $m = k$), the model is just-identified. For this just-identified model, EndoS will display in the outputs that the statistic cannot be computed because the model is just-identified (in fact J -statistic equals zero). For the just-identified model, it is impossible to test whether the instruments are exogenous (Stock & Watson, 2015). If the number of instruments is less than the number of endogenous variables, i.e., when the model is un-identified, EndoS will stop its estimation procedure.

Weak instrument test. As explained at the beginning of the paper, an instrument must correlate with the endogenous variable. However, when an instrument is only weakly correlated with an endogenous variable, little of the exogenous variation in the instrument can be used to predict the potentially endogenous X , thus producing biased and inconsistent IV estimates (Andrews, Stock, & Sun, 2019; Bound, Jaeger, & Baker, 1995).⁸ To test whether instruments are weak, EndoS calculates the so-called Cragg-Donald- F statistic. For a special case when there is only one endogenous variable, the Cragg-Donald- F statistic is just the F value obtained from the first stage of the two-stage least squares regression. If there is only one endogenous variable, an instrument is considered weak if the F -statistic is less than 10 (Staiger & Stock, 1997). If there are multiple endogenous variables, EndoS reports the critical values for TOLS size and the TOLS relative bias at the 5% significance level, which is known as the Stock-Yogo test (Stock & Yogo, 2005). The null hypothesis in the Stock-Yogo test is that instruments are weak, against the alternative hypothesis that the instruments are strong. Instruments are considered strong if the bias of the TOLS estimator is at most 10% of the bias of the OLS estimator (Stock & Watson, 2015, p. 517).

⁷I wish to thank an anonymous reviewer for pointing this out.

⁸I wish to thank an anonymous reviewer for his/her suggestion to include this explanation.

Note on the binary decisions in the diagnostic tests

Note that in the above three procedures, researchers are forced to make a binary decision (reject vs. not reject) upon seeing the statistical results related to the null hypothesis significance testing (NHST). For example, in the Hausman tests, the null hypothesis is H_0 : Instrumented variables are exogenous (OLS is efficient); in the overidentifying restriction test, the null hypothesis is H_0 : All instruments are exogenous (overidentifying restrictions are valid); and in the weak instruments test, the null hypothesis is H_0 : Instruments are weak (e.g., first-stage F statistic < 10). In fact, the possibility of making these types of binary decisions is one of the reasons why the specification or diagnostic tests are very popular in applied research such as economics as the tests allow researchers to evaluate different models (e.g., IV vs. OLS models; Woutersen & Hausman, 2019). However, these are also possible limitations of the above tests as false decisions due to NHST are not unlikely as the tests might lack power (Gill, 1999). Recently, Woutersen and Hausman (2019) introduced a novel method to increase the power of the specification test, however this method is beyond of the scope of my paper. Thus, one must be aware of these limitations.

Using EndoS with illustrative examples

In this section, I demonstrate how to use EndoS (see Appendix for installation and the cautionary note) using two illustrative examples. In the first example, the focus is on the interpretation of the outputs with less attention given to the quality of the instruments being used. In the second example, more attention is given to development of the instrumental variable regression model and the search for the instruments.

Illustration 1. Using data from (Mroz, 1987)

I use Mroz.sav data that contains 428 observations of working women (Mroz, 1987; Wooldridge, 2000), which is also discussed in Wooldridge (2016). As stated in Mroz (1987), this data is part of the University of Michigan Panel Study of Income Dynamics for the year 1975. Mroz (1987) used this data to study married women's labour force participation. The sample was white married women aged between 30 and 60 years in 1975. Data collected by interviewing respondents (i.e., self-reported data), include, among others, the following variables: yes/no variable indicating whether or not a respondent worked for a wage outside home during the year, age, years of education, hourly wage (in dollars), husband's age, husband's hourly wage (in dollars), husband's hours of work, number of children under



six years of age, number of kids between six and 18 years of age, and father's and mother's level of education. For this illustration, I focus on this famous return-to-education regression model:

$$\log(\text{Wage}) = \beta_0 + \beta_1 \text{Educ} + u \quad (7)$$

where `Educ` is the respondent's years of schooling and `Wage` is the respondent's reported earnings at the time data was collected. Because the dependent variable is in the log form, the coefficient β_1 if multiplied by 100% tells us about the percentage return for another year in education. Estimating Equation (7) with OLS will give the coefficient estimate for `Educ` as 0.109 ($t = 7.545$), which means that an additional year in education will give a 10.9% increase in wage.

As an individual's level of education might not be the only variable that affects his/her wages, $\log(\text{Wage})$ in equation (7) will be affected by an unobserved individual ability—highly able individuals will earn more than the less talented individuals. Ability will also correlate with `Educ` because more intelligent individuals will tend to stay longer in education. Because ability exists in the error term u , this implies that `Educ` correlates with u , thus by definition, `Educ` is endogenous. The 10.9% return on education is biased—no longer equal to its true value—due to the variable ability being omitted in equation (7).

To overcome this endogeneity problem, for the sake of this demonstration, I use two instruments: mother's and father's education captured by variables `motheduc` and `fatheduc` in the Mroz data set, respectively. As in Wooldridge (2016), I assume that `motheduc` and `fatheduc` are valid instruments for `Educ`. This assumption may not be true as `motheduc` and `fatheduc` may be correlated with children's ability (e.g., highly educated parents tend to pay more attention to their children's educational progress and attainment than lowly educated parents). However, for the present illustration of using the macro, I assume that the use of the two instruments is justified. In this illustration, I also use women's age (in years) as a covariate assuming that older individuals earn more than the younger individuals. After EndoS is installed on SPSS, EndoS can be found by clicking Analyze: Regression. The screen capture of the EndoS for this particular example can be seen in Figure 2. Clicking the 'Option' button, users can tick all boxes under the 'Diagnostic' tests (see the right image of Figure 2) to instruct EndoS to perform the three diagnostic tests explained in the previous section.

Interpretations of the EndoS outputs

EndoS outputs can be seen in Listing 1 (see Appendix) and will be annotated as follows. First, EndoS produces information on the sample size and the input variables. In the

example, the instrumental variables are `age`, `motheduc`, and `fatheduc` with the last two variables being the external instruments (Z). Notice that `age` appears both in the 'independent variable' box and the 'instrument' box. `Age` is the covariate that can be used as an internal instrument. If `age` is omitted from the instrument box, `age` will be considered as another endogenous variable in which case the model will have two endogenous variables ($m = 2$) and two instruments ($k = 2$). In this new situation, the model is said to be just identified as the number of the endogenous variable being equal to the number of instruments. When the model is identified ($m = k$), EndoS will still produce the TSLS outputs, but no diagnostic tests for overidentifying restrictions are performed.

EndoS produces the OLS outputs presented in lines 1-34. As can be seen from the output, the regression coefficient for `Educ` equals 0.11, which means that an additional year in education will give a 11% increase in wage.

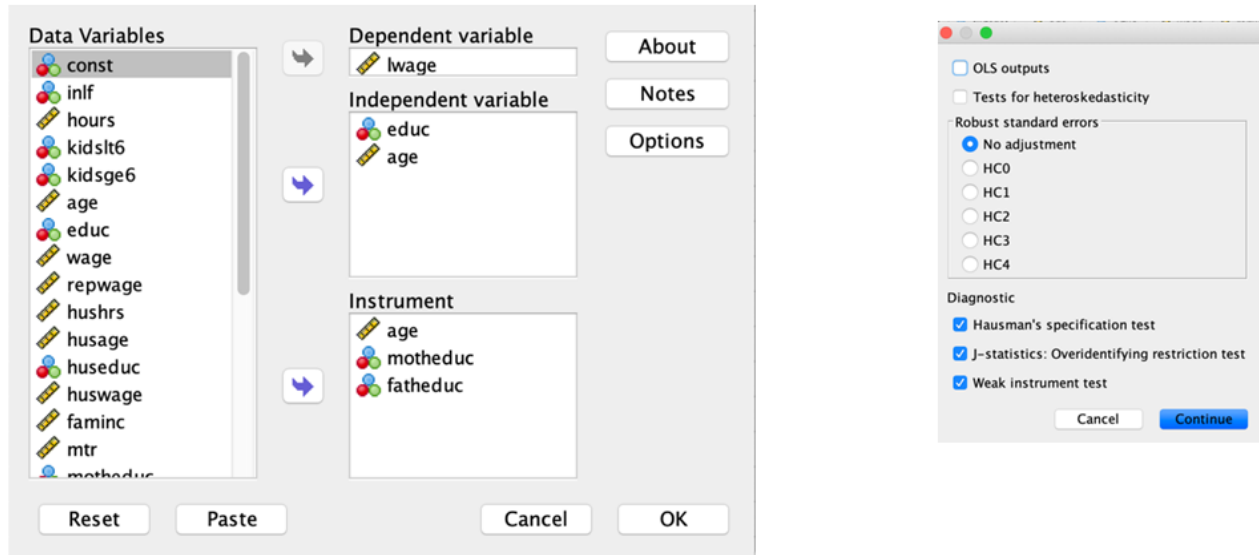
The TSLS outputs are presented in lines 37-151. As can be seen from the output, the regression coefficient for `Educ` now equals 0.058, which means that an additional year in education will give a 5.8% increase in wage, which is far below the coefficient obtained from OLS regression—about 5% lower representing a drop of almost 50%. This suggests that the original value of the return on education from OLS is upward biased. At this stage, users might be tempted to interpret the R-square of IV regression reported in the output, where they should ignore it as R-square is meaningless in IV regression. In IV regression, R-square can be negative because the model sum of squares can be less than zero (Stock & Watson, 2015).

Lines 77-117 produce outputs for Hausman's specification test. The information about the two-step OLS procedure was also given in the first few lines of the outputs. The most important information is presented in lines 109-111, which shows the F statistic for the significance of the 'resid' variable. As there is only one endogenous variable in the model, only one 'resid' variable (`resid1`), appears in the output. The F value is 3.322 ($p = 0.069$), which can be considered as marginally significant. One can conclude that `Educ` might be endogenous. The null hypothesis printed in the output is: H_0 : Instrumented variables are exogenous (OLS is efficient; i.e., estimates of all residuals are equal to zero).

Lines 119-127 are outputs from the overidentifying restriction test. The outputs of the test will tell us whether an extra instrument (`motheduc` or `fatheduc`) is really necessary—an extra instrument will not reduce TSLS standard errors or increases precision. However, as explained above, for a just-identified model, the test cannot be conducted. The overidentifying restriction test assumes a large sample size, the presence of strong instruments and



Figure 2 ■ Screen capture of the EndoS macro.



error to be homoskedastic. The J -statistic for this example is 1.72 with p-value equal to 0.678. The null hypothesis printed in the output is that: H_0 : All instruments are exogenous. The results show that this null hypothesis is not rejected at the 5% significance level, which means that *motheduc* and *fatheduc* are indeed valid instruments.

Lines 129-147 are outputs for the weak instrument test. The Cragg-Donald F -statistic is 55.3015. Because there are two instruments, one cannot use the rule-of-thumb ($F < 10$) by Staiger and Stock (1997). In the next line, the output shows the results of the weak instrument test based on TSLS size. For this example, for 10% of the maximal IV size, the critical value is 19.93; for 15% of the maximal IV size, the critical value is 11.59, and so on. Thus, the Cragg-Donald F -statistic from the two instruments in this example ($F = 55.3015$) exceeds the critical value of 10% maximal bias ($p=0.05$), and is even larger than all Stock-Yogo's critical values, thus one can conclude that *motheduc* and *fatheduc* are not weak instruments. The critical values for the relative bias are not available for models with less than three instruments as noted in the EndoS outputs.

Illustration 2. Replicating Acemoglu et al. (2001) study

In the second illustration, I present the replication of the results of the famous endogeneity paper by Acemoglu et al. (2001) that examined the effect of a country's current political institutions on the country's GDP. I use *Acemoglu.sav* data that contains 163 observations. The data were converted to SPSS for-

mat from the Stata format available to download at <https://economics.mit.edu/faculty/acemoglu/data/ajr2001>. The number of countries or observations actually used is 64, which are the base countries in the data set—those with variable *baseco* equals 1. Finding the answer to this question was aimed at addressing the more fundamental question: why are there large differences in income per capita across the colonized countries in Africa, Asia and Latin America? They estimated the following model:

$$\log(GDP) = \beta_0 + \beta_1 Risk + u \quad (8)$$

where GDP is the country's GDP per capita in 1995, and $Risk$ is a proxy variable that captures the quality of the country's current political institutions. It is an index that "measures differences in institutions originating from different types of states and state policies" (Acemoglu et al., 2001, p. 1370). Estimating model (8) with OLS gives $\beta_1 = 0.522$. The interpretation of the coefficient is that the difference between the GDP of two countries with one point different in the index of political institution is about 5.22%.

The regression model (8) suffers from an endogeneity problem as $Risk$ is correlated with the error term, thus is said to be endogenous. The error term includes factors that affect political institutions, which in turn will affect GDP.

In the search for an instrumental variable, Acemoglu et al. (2001) argued that the differences in the GDP that still exist today are related to the current political institutions that were shaped during the European colonial time during which the colonizers implemented either an extractive



or migrant strategy. An extractive strategy means that the colonizers were more interested in “transferring as much of the resources of the colony to the colonizer” (Acemoglu et al., 2001, p. 1370) rather than migrating or settling down in the colonized countries, thus setting up an extractive institution in the colonized countries. The migrant strategy is the opposite of the extractive strategy where colonizers aimed to establish settlements in the colonized countries.

What factors affect the choice of strategy? Acemoglu et al. (2001) argues that the mortality rates of the European settlers in the colony influences the choice. European settlers that faced high mortality rates (e.g., in a colony with a higher chance to contract a disease) would choose to implement the extractive strategy. Those who faced low mortality rates in a colony would choose the migrant strategy and would set up political institutions that resembled those in Europe. However, mortality rates at the time of the colonization have no influence on the current GDP. Thus, mortality rates satisfy both the relevance (i.e., correlated with *Risk*) and exogeneity condition (i.e., uncorrelated with GDP) to be an instrument.

Listing 2 shows EndoS outputs from OLS on the top lines followed those of TSLS estimations. The outputs show that the coefficient for *Risk* now equals $\beta_1 = 0.944$. This means that the difference between the GDP of two countries with one point different in the index of political institution is about 9.44%, which is nearly 10%. This value clearly shows that the OLS coefficient is downward biased. The Hausman’s test reveals that *Risk* is indeed endogenous ($F = 17.621$). As there is only one instrument, the overidentifying restriction test is not relevant. For a just-identified model, EndoS gives a message in the outputs that the statistic cannot be computed (see Lines 119–120 in Listing 2). The weak instrument test gives the Cragg-Donald F -statistic as equal to 22.9468. This value is larger than any critical values in the table, showing evidence to reject the null hypothesis of weak instrument. Thus, *mortality rate* is a strong instrument.

Limitations

It is worth mentioning here that the computation in EndoS assumes that the error term is homoskedastic—variance of the error term is constant. Therefore, all the three diagnostic tests discussed in the paper are not valid for heteroskedastic errors—e.g., distribution of errors depends on an independent variable. For instance, if the errors are heteroskedastic, then the J -statistic no longer has a chi-squared distribution, in which case the Generalized Method of Moments (GMM) can be used and is more efficient than the TSLS estimator (Stock & Watson, 2015). The GMM estimator is not yet implemented in EndoS. Nevertheless, like the OLS estimator, the calculation of the

standard errors that are robust to heteroskedasticity of unknown form is possible for the IV estimator by using the sandwich form that modifies equation (4) (see Davidson & MacKinnon, 2004, p. 335) Davidson & MacKinnon, 2004, p. 335). EndoS offers an option to conduct homoskedasticity tests in the form of the Breusch-Pagan and Koenker tests. Furthermore, EndoS computes the heteroskedasticity-consistent standard errors for the IV estimator giving four types of error corrections as possible options, i.e., HC1, HC2, HC3, and HC4.

Conclusion

The issue of endogeneity is increasingly recognized and widely discussed in psychology, business and management research. However, the use of IV regression in applied research is still very sporadic. Although IV regression is not the only tool to address endogeneity, and TSLS is not the only estimation procedure, the technique can help researchers in uncovering a causal mechanism that is hidden beyond the model. With the help of two illustrations, this paper has provided working examples of how to assess endogeneity.

This paper also introduces EndoS as a pedagogic and research tool. It covers three diagnostic tests for endogeneity in instrumental variable estimation (i.e., Hausman’s specification test, overidentifying regression test, and weak instrument test) that are commonly addressed in endogeneity empirical studies. Having EndoS written in the SPSS syntax file means that users can easily access its source code (i.e., by clicking ‘paste’ instead of OK in the EndoS windows), which may help users in understanding the mechanics of the method.

I hope this paper serves as helpful guidance on how to use the EndoS macro in general and explains how to undertake diagnostic tests for endogeneity in particular. For pedagogical reasons, because of its ease of use, EndoS can be recommended for use in SPSS-based quantitative modules for social science students in general (eg., psychology, education, health, business and management). Furthermore, using EndoS can be students’ first step to learn endogeneity and SPSS matrix programming because of easy access to the macro’s source codes. Lastly, I hope that seasoned empirical researchers might find this macro useful for their work and the macro can be considered for addition to SPSS statistical training courses aimed at researchers and professionals.

References

- Abdallah, W., Goergen, M., & O’Sullivan, N. (2015). Endogeneity: How failure to correct for it can cause wrong inferences and some remedies. *British Journal of Management*, 26, 791–804.



- Acemoglu, D., Johnson, S., & Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation. *American economic review*, 91(5), 1369–1401.
- Andrews, I., Stock, J. H., & Sun, L. (2019). Weak instruments in instrumental variables regression: Theory and practice. *Annual Review of Economics*, 11, 727–753.
- Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. *The leadership quarterly*, 21(6), 1086–1120.
- Bascle, G. (2008). Controlling for endogeneity with instrumental variables in strategic management research. *Strategic organization*, 6(3), 285–327.
- Bollen, K. A. (2012). Instrumental variables in sociology and the social sciences. *Annual Review of Sociology*, 38, 37–72.
- Bound, J., Jaeger, D. A., & Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association*, 90(430), 443–450. doi:10.1080/01621459.1995.10476536
- Davidson, R., & MacKinnon, J. G. (2004). *Econometric theory and methods*. New York: Oxford University Press.
- Ebbes, P., Wedel, M., Steerneman, T., & Bockenholt, U. (2005). New evidence for the effect of education on income: Solving endogeneity with latent instrumental variables. *Quant. Marketing Econom*, 3(4), 365–392.
- Gennetian, L. A., Magnuson, K., & Morris, P. A. (2008). From statistical associations to causation: What developmentalists can learn from instrumental variables techniques coupled with experimental data. *Developmental psychology*, 44(2), 381–381.
- Gill, J. (1999). The insignificance of null hypothesis significance testing. *Political research quarterly*, 52(3), 647–674.
- Hair, J., Black, W., Anderson, R., & Babin, B. (2018). *Multivariate data analysis (8, ilustra ed.)* Washington, DC: Cengage Learning EMEA.
- Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica: Journal of the econometric society*, 11, 1251–1271.
- Hill, R. C., Griffiths, W. E., & Lim, G. C. (2018). *Principles of econometrics*: John Wiley & Sons.
- Jean, R.-J. B., Deng, Z., Kim, D., & Yuan, X. (2016). Assessing endogeneity issues in international marketing research. *International Marketing Review*, 33(3), 483–512.
- Kennedy, P. (2003). *A guide to econometrics*: MIT press.
- Ketokivi, M., & McIntosh, C. N. (2017). Addressing the endogeneity dilemma in operations management research: Theoretical, empirical, and pragmatic considerations. *Journal of Operations Management*, 52, 1–14.
- Lattin, J. M., Carroll, J. D., & Green, P. E. (2003). *Analyzing multivariate data*. Washington, DC.: Springer.
- Lewbel, A. (2012). Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *Journal of Business & Economic Statistics*, 30(1), 67–80.
- Maydeu-Olivares, A., Shi, D., & Rosseel, Y. (2019). Instrumental variables two-stage least squares (2sls) vs. maximum likelihood structural equation modeling of causal effects in linear regression models. *Structural Equation Modeling: A Multidisciplinary Journal*, 26(6), 876–892.
- Mroz, T. A. (1987). The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. *Econometrica: Journal of the econometric society*, 11, 765–799.
- Papies, D., Ebbes, P., & Van Heerde, H. J. (2017). Addressing endogeneity in marketing models. In M. Mouse (Ed.), *Advanced methods for modeling markets* (pp. 581–627). New York: Springer.
- Park, S., & Gupta, S. (2012). Handling endogenous regressors by joint estimation using copulas. *Marketing Science*, 31(4), 567–586.
- Petrin, A., & Train, K. (2010). A control function approach to endogeneity in consumer choice models. *Journal of marketing research*, 47(1), 3–13.
- Reeb, D., Sakakibara, M., & Mahmood, I. P. (2012). *From the editors: Endogeneity in international business research*. In: Springer.
- Rutz, O. J., & Watson, G. F. (2019). Endogeneity and marketing strategy research: An overview. *Journal of the Academy of Marketing Science*, 47(3), 479–498.
- Sajons, G. (2020). Estimating the causal effect of measured endogenous variables: A tutorial on experimentally randomized instrumental variables. *Leadership Quarterly*, 55, 11. Forthcoming.
- Shaver, J. M. (2019). Causal identification through a cumulative body of research in the study of strategy and organizations. *Journal of Management*, 22, 1–11.
- Staiger, D., & Stock, J. H. (1997). Instrumental variables regression with weak instruments, 557. *Econometrica: Journal of the econometric society*(, 65, 557–586.
- Stock, J. H., & Watson, M. W. (2015). *Introduction to econometrics*. London: Pearson.
- Stock, J. H., & Yogo, M. (2005). Testing for weak instruments in linear iv regression. In D. W. K. Andrews & J. H. Stock (Eds.), *Identification and inference for econometric models: Essays in honor of thomas j. rothenberg* (pp. 1–11). Cambridge, UK: Cambridge University Press.



- Tabachnick, B., & Fidell, L. (2017). *Using multivariate statistics*. New York: In.
- Wooldridge, J. M. (2000). *Mroz: Instructional stata datasets for econometrics*. New York: Springer.
- Wooldridge, J. M. (2016). *Introductory econometrics: A modern approach*. Washington, DC.: Nelson Education.
- Woutersen, T., & Hausman, J. A. (2019). Increasing the power of specification tests. *Journal of Econometrics*, 211(1), 166–175.
- Zaefarian, G., Kadile, V., Henneberg, S. C., & Leischnig, A. (2017). Endogeneity bias in marketing research: Problem, causes and remedies. *Industrial Marketing Management*, 65, 39–46. doi:<https://doi.org/10.1016/j.indmarman.2017.05.006>

Appendix A: EndoS installation

EndoS is written in the spd format. To install the macro, one must download the EndoS.spd, which is available on my google website or the journal website. Next, one must start a SPSS session. For SPSS ver 24 and 25, the macro can be installed by clicking Extensions: Utilities: Install Custom Dialog. For SPSS ver 23 and earlier, it can be installed within SPSS by clicking: Utilities: Custom Dialogs: Install Custom Dialog. Next, locate the EndoS.spd file and click the file. The dialog box will appear on the screen; click OK. The macro will be installed under SPSS menu: Analyze: Regression: EndoS.

The above way of installing the macro will work just fine for personal PCs on which users have admin rights to change the default folder where the macro should be installed. For PCs that required admin rights and are part of a network (e.g., a desktop PC in the workplace), installing via the above procedure will prompt users with an installation error. To circumvent this problem, I have provided three files: defaultExtensionDialog.png, EndoS.properties, and EndoS.xml, which can be downloaded from my google website and placed into this location: C:\Program Files\IBM\SPSS\Statistics\XX\ext\lib\spss.command, where XX is the version of the IBM SPSS statistics.⁹ After this step has been done, open SPSS, and the macro should be found under the SPSS menu: Analyze: Regression: EndoS.

Cautionary note

First, EndoS puts great faith in its users that they must understand that the input matrix in a regression must not be singular, e.g., no independent variable is a linear combination of any other independent variables. If the input matrix is singular, EndoS will produce a list of error messages. Second, EndoS assumes no missing data, if otherwise, SPSS error messages will appear, and no outputs will be printed out.

Listing 1. Outputs from EndoS for illustration 1.

```
1 Run MATRIX procedure:
2
3 Two-Stage Least Squares
4
5 Written by Ahmad Daryanto
6
7 https://sites.google.com/site/ahmaddaryanto/
8
9 -----
10 OLS regression
11
12 Sample size
13 428
14
15 Dependent variable (DV)
16 lwage
17
18 Independent variable (IVs)
19 educ      age
20
```

⁹I wish to thank the action editor who pointed out this solution.



21 Model summary

| 22 | R ² | Adj | SER |
|----|----------------|-------|-------|
| 23 | .1232 | .1170 | .6788 |

24
25 Note: Adj = Adjusted R², SER = Std. error of regression.

26
27 OLS outputs

| 28 | | b | se | t | sig | 95%LB | 95%UB |
|----|----------|-------|------|--------|------|--------|-------|
| 29 | constant | -.488 | .264 | -1.848 | .065 | -1.005 | .030 |
| 30 | educ | .110 | .014 | 7.633 | .000 | .082 | .138 |
| 31 | age | .007 | .004 | 1.606 | .109 | -.002 | .015 |

32 * Note: standard errors are assumed to be homoskedastic--no adjustments.

33
34
35 =====

36
37 Two-Stage Least Squares

38
39 Sample size
40 428

41
42 Dependent Variable
43 lwage

44
45 Endogeneous (instrumented) variable
46 educ

47
48 Instrument(s)
49 age motheduc fatheduc

50
51 External instrument(s)
52 motheduc fatheduc

53
54 Note: External instrument(s) is truly exogeneous, not part of the original model.

55
56 Model summary

| 57 | R ² | Adj | SER |
|----|----------------|-------|-------|
| 58 | .0969 | .0905 | .6889 |

| 59 | df(m) | df(res) | F | Sig. |
|----|--------|----------|--------|-------|
| 60 | 2.0000 | 425.0000 | 2.3683 | .0949 |

61
62 Note: Adj = Adjusted R², SER = Std. error of regression.

63
64
65
66 TSLS outputs

| 67 | | b | se | t | sig | 95%LB | 95%UB |
|----|----------|------|------|-------|------|--------|-------|
| 68 | constant | .196 | .465 | .421 | .674 | -1.005 | .030 |
| 69 | educ | .058 | .032 | 1.821 | .069 | .082 | .138 |
| 70 | age | .006 | .004 | 1.391 | .165 | -.002 | .015 |

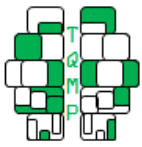
71
72



```

73 -----
74 * Note: standard errors are assumed to be homoskedastic--no adjustments.
75 -----
76
77 Hausman's specification test: Two-stage OLS regression using residuals as
    predictors.
78
79 Note: Results obtained via two stages.
80
81 Stage 1: Endogeneous variables are regressed on the IVs and external instruments.
82
83 Stage 2: DV is regressed on IVs and residuals obtained from stage 1.
84
85 The test assumes that the errors in stage 2 are homoskedastic.
86
87 Dependent variable
88 lwage
89
90 Model summary
91   R^2      Adj      SER
92 .1301     .1218     .6769
93
94
95 df(m)   df(res)      F      Sig.
96 3.0000  424.0000    21.1291  .0000
97
98 Note: Adj = Adjusted R^2, SER = Std. error of regression.
99
100 =====
101           b           se           t           sig           95%LB           95%UB
102 constant   .196         .457         .429         .668         -.700         1.092
103 educ       .058         .032         1.853         .065         -.003         .120
104 age        .006         .004         1.416         .158         -.002         .014
105 resid1     .065         .035         1.828         .068         -.005         .134
106
107 ----
108
109 Joint F test for the significance of the 'resid' variables
110   df1      df2      F      Sig.
111 1.000    424.000    3.322    .069
112
113 H0: Instrumented variables are exogeneous (OLS is efficient) (i.e., estimates of
    all resids are equal to zero).
114
115 H1: Instrumented variables are endogeneous.
116
117 -----
118
119 Overidentifying restrictions test (The $J$-statistic)
120 Value      .172
121 Sig.       .678
122

```



```
123 H0: All instruments are exogenous.
124
125 The test assumes a large sample size, strong instruments, and homoskedastic errors.
126
127 -----
128
129 Weak instrument tests
130
131 Cragg-Donald F-statistic
132 55.3015
133
134 Stock-Yogo critical values for the weak instrument test based on TSLS size at 5%
    significance level:
135   Size      Values
136   .10      19.93
137   .15      11.59
138   .20       8.75
139   .25       7.25
140
141 Stock-Yogo critical values for relative bias are not available for models with less
    than 3 instruments.
142
143 Note: For one endogeneous variable, Cragg-Donald F-statistic is the F-value of
    first stage regression.
144
145 For one endogeneous regressor, instruments are weak if the F-statistics is less
    than 10 (Staiger and Stock, 1997).
146
147 Note: Tables for the critical values (Stock & Yogo, 2002) are reproduced by
    permission--email communication (Stock, 16-11-2019).
148
149 =====
150
151 ----- END MATRIX -----
```

Listing 2. Outputs from EndoS for illustration 2.

```
1 Run MATRIX procedure:
2
3 Two-Stage Least Squares
4
5 Written by Ahmad Daryanto
6
7 https://sites.google.com/site/ahmaddaryanto/
8
9 -----
10
11 OLS regression
12
13 Sample size
14 64
15
```



16 Dependent variable (DV)

17 logppp95

18

19 Independent variable (IVs)

20 avexpr

21

22 Model summary

23

| R ² | Adj | SER |
|----------------|-----|-----|
|----------------|-----|-----|

| | | |
|-------|-------|-------|
| .5401 | .5250 | .7132 |
|-------|-------|-------|

26

27 Note: Adj = Adjusted R², SER = Std. error of regression.

28

29 OLS outputs

30

| | b | se | t | sig | 95%LB | 95%UB |
|--|---|----|---|-----|-------|-------|
|--|---|----|---|-----|-------|-------|

| | | | | | | |
|----------|-------|------|--------|------|-------|-------|
| constant | 4.660 | .409 | 11.408 | .000 | 3.860 | 5.461 |
|----------|-------|------|--------|------|-------|-------|

| | | | | | | |
|--------|------|------|-------|------|------|------|
| avexpr | .522 | .061 | 8.533 | .000 | .402 | .642 |
|--------|------|------|-------|------|------|------|

34 Note: standard errors are assumed to be homoskedastic--no adjustments.

35

36 Two-Stage Least Squares

37

38 Sample size

39 64

40

41 Dependent Variable

42 logppp95

43

44 Endogeneous (instrumented) variable

45 avexpr

46

47 Instrument(s)

48 logem4

49

50 External instrument(s)

51 logem4

52

53 Note: External instrument(s) is truly exogeneous, not part of the original model.

54

55 Model summary

56

| R ² | Adj | SER |
|----------------|-----|-----|
|----------------|-----|-----|

| | | |
|-------|-------|-------|
| .1870 | .1603 | .9483 |
|-------|-------|-------|

59

| df(m) | df(res) | F | Sig. |
|-------|---------|---|------|
|-------|---------|---|------|

| | | | |
|--------|---------|---------|-------|
| 1.0000 | 62.0000 | 36.3941 | .0000 |
|--------|---------|---------|-------|

62

63 Note: Adj = Adjusted R², SER = Std. error of regression.

64

65 TSLS outputs

66

| | b | se | t | sig | 95%LB | 95%UB |
|--|---|----|---|-----|-------|-------|
|--|---|----|---|-----|-------|-------|



| | | | | | | | |
|----|----------|-------|-------|-------|------|-------|-------|
| 68 | constant | 1.910 | 1.027 | 1.860 | .068 | 3.860 | 5.461 |
| 69 | avexpr | .944 | .157 | 6.033 | .000 | .402 | .642 |

70

71 -----

72 Note: standard errors are assumed to be homoskedastic--no adjustments.

73 -----

74

75 Hausman's specification test: Two-stage OLS regression using residuals as predictors.

76

77 Note: Results obtained via two stages.

78

79 Stage 1: Endogeneous variables are regressed on the IVs and external instruments.

80

81 Stage 2: DV is regressed on IVs and residuals obtained from stage 1.

82

83 The test assumes that the errors in stage 2 are homoskedastic.

84

85 Dependent variable

86 logppp95

87

88 Model summary

89

| | | | |
|----|----------------|-------|-------|
| 90 | R ² | Adj | SER |
| 91 | .6708 | .6544 | .6084 |

92

93

| | | | | |
|----|--------|---------|---------|-------|
| 94 | df(m) | df(res) | F | Sig. |
| 95 | 2.0000 | 61.0000 | 62.1532 | .0000 |

96

97 Note: Adj = Adjusted R², SER = Std. error of regression.

98

99 =====

| | | | | | | | |
|-----|----------|-------|------|--------|------|-------|-------|
| 100 | | b | se | t | sig | 95%LB | 95%UB |
| 101 | constant | 1.910 | .659 | 2.899 | .005 | .619 | 3.201 |
| 102 | avexpr | .944 | .100 | 9.404 | .000 | .747 | 1.141 |
| 103 | resid1 | -.578 | .118 | -4.921 | .000 | -.809 | -.348 |

104

105 ----

106

107 Joint F test for the significance of the 'resid' variables

108

| | | | | |
|-----|-------|--------|--------|------|
| 109 | df1 | df2 | F | Sig. |
| 110 | 1.000 | 61.000 | 17.621 | .000 |

111

112 H0: Instrumented variables are exogeneous (OLS is efficient) (i.e., estimates of all resids are equal to zero).

113

114

115 H1: Instrumented variables are endogeneous.

116

117



```

118 -----
119 Overidentifying restrictions test (The $J$-statistic).
120 The statistic cannot be computed because the model is just-identified.
121 -----
122
123 Weak instrument tests
124
125 Cragg-Donald F-statistic
126 22.9468
127
128 Stock-Yogo critical values for the weak instrument test based on TSLS size at 5%
    significance level:
129 Size      Values
130 .10      16.38
131 .15       8.96
132 .20       6.66
133 .25       5.53
134
135 Stock-Yogo critical values for relative bias are not available for models with less
    than 3 instruments.
136
137 Note: For one endogeneous variable, Cragg-Donald F-statistic is the F-value of
    first stage regression.
138
139 For one endogeneous regressor, instruments are weak if the F-statistics is less
    than 10 (Staiger and Stock, 1997).
140
141 Note: Tables for the critical values are reproduced by permission--email
    communication with JS dated 16-11-19 (Stock & Yogo, 2005)
142
143 =====
144
145 ----- END MATRIX -----

```

Open practices

📄 The *Open Material* badge was earned because supplementary material(s) are available on [the journal's web site](#).

Citation

Daryanto, A. (2020). EndoS: An SPSS macro to assess endogeneity. *The Quantitative Methods for Psychology*, 16(1), 56–70. doi:[10.20982/tqmp.16.1.p056](https://doi.org/10.20982/tqmp.16.1.p056)

Copyright © 2020, Daryanto. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Received: 07/12/2019 ~ Accepted: 14/02/2020