CSCanada
**Cross-Cultural Communication**
Vol. 18, No. 1, 2022, pp. 18-31
**DOI:**10.3968/12498

ISSN 1712-8358[Print]
ISSN 1923-6700[Online]
www.cscanada.net
www.cscanada.org

# Cross-Sectional Analysis of Methods of Computing Partial Correlation Coefficients: A Self-Explained Note With R Syntax

Timothy A. Ogunleye[a],*; Kehinde K. Adesanya[b]; Oluwatosin J. Akinsola[c]; Godwill I. Wilcox[d]

[a] Department of Statistics, University of Ilorin, Nigeria & Head, Tim-R Programming Consult, Nigeria.
[b] Department of Biostatistics, Ogun State College of Heath Technology, Ilese-Ijebu, Nigeria.
[c] Department of Community Health and Primary Care, College of Medicine, University of Lagos, Nigeria.
[d] Department of Agricultural Economics, Ignatius Ajuru University of Education, Rumuolumeni, Port Harcourt, Nigeria.
*Corresponding author.

Received 29 December 2021; accepted 24 January 2022
Published online 26 March 2022

## Abstract
This paper examines four different methods of computing partial correlation coefficients. These include conventional method, variance-covariance matrix approach, regression residual's approach, and OLS method. Each of these is fully illustrated with practical examples as well as R syntax. Applicability of each of the methods is discussed in our illustrations. Strength and weakness of each method are extensively detailed. It's, however, discovered that none of the basic assumptions of partial correlation: linearity, normality, and non-existence of outliers is violated after performing statistical checks on the datasets used. The study, therefore, recommends the best method(s) of computing partial correlation coefficients when at least one variable is held constant, thereby adding more invaluable knowledge to the existing literatures. Finally, the study further recommends the best method in each scenario with illustrative examples as evidences.

**Key words:** Conventional method; OLS method; Partial correlation; Regression residual's approach; Variance-covariance matrix method

Ogunleye, T. A., Adesanya, K. K., Akinsola, Q. J., & Wilcox, G. I. (2022). Cross-Sectional Analysis of Methods of Computing Partial Correlation Coefficients: A Self-Explained Note With R Syntax. *Cross-Cultural Communication, 18*(1), 18-31. Available from: http//www.cscanada.net/index.php/ccc/article/view/12498
DOI: http://dx.doi.org/10.3968/12498

**JEL Codes: JEL C63 (Computational Techniques) and JEL C65 (Miscellaneous Mathematical Tools)**

## 1. INTRODUCTION

In both parametric and non-parametric statistical situations, where the term correlation may be mentioned, simply the word *correlation* means an association that exists between or among (as the case may be) two or more variables, where perhaps one depends on the other. On the other hand, correlation coefficient could be defined as a measurement of the strength of such association. This measurement is usually in two decimal places, making it easier to interpret in form of percentage. In another dimension, coefficient of correlation could mean the degree at which one variable is associated with the other.

According to Akoglu (2018) and Merriam-webster dictionary (2020), correlation is defined as *a relation that exists between phenomena or things (or between mathematical or statistical variables) which tend to vary, be associated, or occur together in a way not expected by chance alone*. Also, Mukaka (2012) defines correlation as a statistical method used to assess a possible linear association between two continuous variables. It is assumed to be simple both in calculation and interpretation. However, misuse of correlation is so common among researchers that some statisticians have wished that the method had never been devised at all (Mukaka, 2012; Bishara and James, 2017).

The term 'partial correlation' is defined as the type of correlation that exists between two random variables, where one depends on the other, for which at least one other variable is held constant. Though it exists among at least three variables, the first two variables are the main target variables with all other variables referred to as controlling variables. However, the number of controlling variable(s) determines the order of the partial correlation.

If there is no controlling variable, it is called zero-order partial correlation, and equivalently, it is known as Pearson's correlation coefficient.

The present study examines four different methods of computing partial correlation coefficients. Each of the methods is fully illustrated with practical examples as well as R syntax. Applicability of each of the methods is discussed in our illustrations. Strength and weakness of each method are extensively detailed. This study recommends the best method in each scenario with illustrative evidences. Therefore, the study aims at providing the best method of computing partial correlation coefficients when one, two and more variables are held constant, thereby adding more invaluable knowledge to the existing literatures. The first chapter is introduction followed by literature review, which is subdivided into theoretical and empirical literature. The third chapter, data and methodology, comprising the data, theoretical framework, model specification, and estimation procedure, are extensively discussed. Descriptive statistics, pre-estimation results, and estimation results come next as the fourth, while the last chapter discusses conclusion and recommendation.

## 2. LITERATURE REVIEW

### 2.1 Theoretical Literature

Yule (1907), Fisher (1924) and Fieller, *et. al.* (1957) report that any correlation coefficient shouldn't be more than positive one (+1) and also shouldn't be less than negative one (-1). Occasionally sometimes, the coefficient could be zero, which indicates that no iota of association exist between the variables involved. Therefore, all correlation coefficients should be between -1 and +1 respectively. According to Hinkle, *et. al.* (2018) and Egozcue, *et. al.* (2018), the following rules of thumb are succinctly explained in Table 1 for the interpretation of correlation coefficients.

**Table 1**
**Rules of Thumb for the Interpretation of Correlation Coefficient**

| S/N | Correlation coefficient | Interpretation |
|---|---|---|
| 1. | Between 0.90 and 1.00 (Between-0.90 and-1.00) | Very high positive (negative) correlation |
| 2. | Between 0.70 and 0.89 (Between-0.70 and-0.89) | *High positive (negative) correlation* |
| 3. | Between 0.50 and 0.69 (Between-0.50 and-0.69) | *Moderate positive (negative) correlation* |
| 4. | Between 0.30 and 0.49 (Between-0.30 and-0.49) | *Low positive (negative) correlation* |
| 5. | Between 0.01 and 0.29 (Between-0.01 and-0.29) | *Negligible or weak correlation* |
| 6. | When the value of the correlation = 0.00 | *Zero correlation or no correlation* |

*Source:* Hinkle, et. al. (2018) and Egozcue, et. al. (2018)

In a situation where variables of interest are continuous, thereby producing cross-sectional data sets, four prominent correlation coefficients are very popular and these include: simple, multiple, partial and canonical correlation coefficients. A correlation is said to be simple if the degree of association between only two variables is computed using either Pearson's Product Moment (PPM) technique, Spearman's Rank technique, or Kendall's Tau technique. And, when the result is positive, it means increase in one variable leads to corresponding increase in the other variable, and vice-versa. Also, whenever the coefficient of correlation is negative, it is indicates that increase in one variable results to decrease in the other, and vice-versa. When we have zero value as the correlation coefficient, it shows that no association exists, indicating that when one variable increases or decreases, the other variable remains stagnant.

A multiple correlation exists between a dependent variable and at least two independent variables. It is an index to measure how well a dependent variable behaves when correlated with two or more independent variables. In statistics, the coefficient of multiple correlation is a measure of how well a given variable can be predicted using a linear function of a set of other variables (Kynčlová, *et. al.*, 2017). It is the correlation between the variable›s values and the best predictions that can be computed linearly from the predictive variables (Langfelder and Steve, 2012).

The term 'canonical correlation', which was first introduced by Hotelling (1936), is a branch of correlation commonly found in multivariate analysis where the degree of association between a set of linear combination of variables and another set of linear combination of variables is measured. Invariably, we could say that canonical correlation analysis is a method used to identify and measure the associations between two sets of variables, since canonical itself is a statistical term for analyzing latent variables. In obtaining canonical correlation coefficient, however, at least two variables must be contained in each set of linear combination of variables.

On the other hand, we could define a partial correlation coefficient as a measurement of the strength of association between two variables, while simultaneously keeping the influence of at least one other variable constant. Specifically, this type of correlation analysis is quite different from the rest three earlier discussed. In this case, more than two random variables are involved, but these variables do not require partition unlike canonical correlation analysis. Simply, it is situation where we measure the degree of association between two variables of interest while holding the influence of all other variables constant. Partial correlation, however, quantifies linear association between two variables while adjusting for the influence of the remaining variables (Lonas, 2020).

### 2.2 Empirical Literature

Partial correlation can be used to statistically control for unwanted variables (Serlin and Harwell, 2007). In probability and statistics, partial correlation measures the degree of association between two random variables, with the effect of a set of controlling variables random variables removed. In a formal situation, the partial correlation between $x_1$ and $x_2$ given a set of *n-controlling variables* $y = \{y_1, y_2, \ldots, y_n\}$, written as $\rho_{x_1 x_2 | y}$, is the correlation between the residuals $e_{x_1}$ and $e_{x_2}$ resulting from the linear regression of $x_1$ with $y$, and of $x_2$ with $y$, respectively (Guilford and Fruchter, 1973; Lonas, 2020). The first-order partial correlation (that is, when $n$ = 1) is the difference between a correlation and the product of the removable correlations divided by the product of the coefficients of alienation of the removable correlations (Fisher, 1924; Guilford and Fruchter, 1973; Lonas, 2020).

However, partial correlations are only valid when the pattern of relationships between the variables reflects a meaningful model (Waliczek, 1996). According to Pedhazur (1982), controlling variables without regard to the theoretical considerations about the pattern of relations among them may amount to a distortion of reality and result in misleading or meaningless results. It is also important consideration is that the researcher must know exactly what dependent variable is being measured after the influence of one or more independent variable(s) has been removed.

In 1988, Horwitz and Rapoport engage in the use of method of partial correlation analysis to characterize the brain in terms of functional association among brain regions. Therefore, correlation coefficients between pairs of regional glucose metabolic rates were extensively discussed in connection with assessing patterns of association among brain regions in humans and animals. The duo argue that partial correlation coefficients (partialling out the global metabolic rate) or correlations between reference rates (regional to global metabolic rate) should be used in removing the distorting influence of systematic intrasubject differences in glucose utilization.

Also, a partial correlation analysis was used by Aloe (2013) while studying the synthesis of partial effect sizes. Three partial effect sizes for the correlation were focused on: the standardized slope, the partial correlation, and the semi-partial correlation. Out of these, Aloe reports that partial correlation was greatly employed and most useful for meta-analysis in two common situations: when primary studies reporting regression models do not report bivariate correlations, and when it is specific interest to partial out the effect of other variables.

In a paper authored by Ha and Sun (2014), a partial correlation was used to handle the problem of construction gene co-expression network. Also, Kenett, *et. al* (2015) apply partial correlation for financial marketing strategies. However, in the analysis of the dependency network methodology, the application of partial correlation has really helped to uncover dependency and, also to influence relationship between the different companies in the investigated sample.

In the recent year, Roverato and Castelo (2017) employ the use of partial correlation to replace marginal correlations, which are used to measure the degree of co-expression between genes. This is as a result of the fact that marginal correlations (Pearson or Spearman correlation coefficients) are very sensitive to indirect effects while analyzing the effects of genetic interaction on yeast. Extensively, the application of partial correlation has become widely accepted, most especially in the fields of biology, medicine, economics or accounting, engineering, and so on.

## 3. DATA AND METHODOLOGY

### 3.1 The Datasets

R has over ten thousand datasets enclosed by many packages. A package called 'Applied Econometrics with R', abbreviated as AER, houses an annual multiple time series datasets of Klein's Model I for the US economy between 1920 and 1941 inclusive. It is named '*KleinI*' in R and comprises nine variables: consumption ($x_1$), corporate profit ($x_2$), private wage bill ($x_3$), investment ($x_4$), previous year's capital stock ($x_5$), gross national product ($x_6$), government wage bill ($x_7$), government expenditure ($x_8$), and taxes ($x_9$). From these, we select only five of them for this study: $x_1$, $x_2$, $x_3$, $x_6$, and $x_7$ respectively. This is done to justify the basic assumptions involved in using partial correlation.

To gain accessibility to the datasets, the following R commands will pave way with the assurance that your system is internet connected:

*install.packages("AER")* (1)

While running the code in (1), the system will request you to choose a CRAN Mirror from a list to be displayed (appearing from the cloud), it is expected that a mirror that is very close to your location, for instance, Australia (Canberra) [https], UK (Bristol) [https], Uruguay [https], and so on, should be chosen and press 'ok' button to proceed the installation process.

After the successful installation of AER package, the next code is to call the library before its application. The command to achieve this task is written in (2) as follows:

*library (AER)* (2)

In R, a library is a repository arena where relevant commands are stored before usage. It could be recalled by running the command 'library ( )'. However, we may still define a library as location on disk where packages are installed, saved, and recalled. After library command, therefore, the next code is:

*data(KleinI)* (3)

The command in (3) is used to bring the datasets into R environment. Having done this, we may decide to view the datasets by running the command below:

*View(KleinI)* (4)

Considering the fact that '*View(KleinI)'* can neither be copied nor taken from one location to another; rather it can only be viewed on the computer screen. However, R provides a convenient command to access the datasets in Microsoft Office Excel Comma Separated Values (CSV) format. The code is:

*write.csv(KleinI, "data_for_this_study.csv")* (5)

In all operating systems, R's default directory is located at 'My Documents' or 'Documents'. After running the code in (5), we will visit the default location of R's directory and check for the file named "data_for_this_study" file; it must have been dropped in CSV format. From there, we can access the datasets as contained in Table 2.

**Table 2**
**Klein's Model I for the US Economy: 1920-1941**

| Year | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $x_8$ | $x_9$ |
|---|---|---|---|---|---|---|---|---|---|
| 1920 | 39.8 | 12.7 | 28.8 | 2.7 | 180.1 | 44.9 | 2.2 | 2.4 | 3.4 |
| 1921 | 41.9 | 12.4 | 25.5 | -0.2 | 182.8 | 45.6 | 2.7 | 3.9 | 7.7 |
| 1922 | 45 | 16.9 | 29.3 | 1.9 | 182.6 | 50.1 | 2.9 | 3.2 | 3.9 |
| 1923 | 49.2 | 18.4 | 34.1 | 5.2 | 184.5 | 57.2 | 2.9 | 2.8 | 4.7 |
| 1924 | 50.6 | 19.4 | 33.9 | 3 | 189.7 | 57.1 | 3.1 | 3.5 | 3.8 |
| 1925 | 52.6 | 20.1 | 35.4 | 5.1 | 192.7 | 61 | 3.2 | 3.3 | 5.5 |
| 1926 | 55.1 | 19.6 | 37.4 | 5.6 | 197.8 | 64 | 3.3 | 3.3 | 7 |
| 1927 | 56.2 | 19.8 | 37.9 | 4.2 | 203.4 | 64.4 | 3.6 | 4 | 6.7 |
| 1928 | 57.3 | 21.1 | 39.2 | 3 | 207.6 | 64.5 | 3.7 | 4.2 | 4.2 |
| 1929 | 57.8 | 21.7 | 41.3 | 5.1 | 210.6 | 67 | 4 | 4.1 | 4 |
| 1930 | 55 | 15.6 | 37.9 | 1 | 215.7 | 61.2 | 4.2 | 5.2 | 7.7 |
| 1931 | 50.9 | 11.4 | 34.5 | -3.4 | 216.7 | 53.4 | 4.8 | 5.9 | 7.5 |
| 1932 | 45.6 | 7 | 29 | -6.2 | 213.3 | 44.3 | 5.3 | 4.9 | 8.3 |
| 1933 | 46.5 | 11.2 | 28.5 | -5.1 | 207.1 | 45.1 | 5.6 | 3.7 | 5.4 |
| 1934 | 48.7 | 12.3 | 30.6 | -3 | 202 | 49.7 | 6 | 4 | 6.8 |
| 1935 | 51.3 | 14 | 33.2 | -1.3 | 199 | 54.4 | 6.1 | 4.4 | 7.2 |
| 1936 | 57.7 | 17.6 | 36.8 | 2.1 | 197.7 | 62.7 | 7.4 | 2.9 | 8.3 |
| 1937 | 58.7 | 17.3 | 41 | 2 | 199.8 | 65 | 6.7 | 4.3 | 6.7 |
| 1938 | 57.5 | 15.3 | 38.2 | -1.9 | 201.8 | 60.9 | 7.7 | 5.3 | 7.4 |
| 1939 | 61.6 | 19 | 41.6 | 1.3 | 199.9 | 69.5 | 7.8 | 6.6 | 8.9 |
| 1940 | 65 | 21.1 | 45 | 3.3 | 201.2 | 75.7 | 8 | 7.4 | 9.6 |
| 1941 | 69.7 | 23.5 | 53.3 | 4.9 | 204.5 | 88.4 | 8.5 | 13.8 | 11.6 |

*Source:* R Core Team (2020)

## 3.2 Theoretical Framework

There are some basic assumptions to be adhered to before making use of partial correlation. Some of these basic assumptions include linearity, normality, and outliers. The theoretical backgrounds for authenticating these assumptions are extensively discussed hereunder.

### 3.2.1 Linearity

There is need to check that variables are linearly related. This could be checked by plotting scatter diagrams. We consider the first variable as the baseline and obtain the graphical representation of the baseline variable with others. These visualizations are achieved with the help of R programming software (R Core Team, 2020).

### 3.2.2 Normality

Before obtaining partial correlation coefficients, it is expected that variables should be approximately normally distributed. This is achieved using Shapiro-Wilk test of normality, which is embedded in R.

### 3.2.3 Outlier

Recall that partial correlation is sensitive to outliers, which can have a very wide effect on the line of best fit and the correlation coefficient, thereby leading to incorrect conclusions regarding the data. It is expected that datasets should not contain significant outlier before choosing partial correlation technique for measuring the degree of association between two variables when the influence(s) of other variable(s) are held constant. Outliers are simply single data points within our datasets that do not follow the usual pattern. If the appearance of outliers is significant, it is advisable to drop the use of partial correlation technique to avoid misleading inferences. In this study, we apply box plot to check the tolerance of outlying effects on our datasets.

### 3.3 Estimation Procedures

This paper studies four different methods of computing the coefficients of partial correlation as earlier said. Each of the four methods is discussed with practical illustrations. The first, second and third order partial correlation coefficients, represented by $\ell_{12|3}$, $\ell_{12|36}$, and $\ell_{12|367}$, could be computed through *conventional approach, variance-covariance matrix method, regression residual's approach, and ordinary least square method.*

In this paper, we represent consumption by $x_1$, corporate profit by $x_2$, private wage bill by $x_3$, gross national product by $x_6$, and government wage bill by $x_7$. Therefore, for instance, the symbol $\ell_{12|3}$ means first order partial correlation coefficient between consumption and corporate profit holding the influence of private wage bill constant.

However, the theoretical procedures for each of the methods are extensively discussed with the inclusion of some R codes to serve as calculator, to avoid computational errors and to reduce computational stress.

#### 3.3.1 Computational Procedures of Conventional Approach

We discuss some of the existing formula for computing partial correlation coefficients using conventional approach. The first order partial correlation coefficient is computed by:

$$\ell_{12|3} = \frac{r_{12} - r_{13}.r_{23}}{\sqrt{\left(1 - r_{13}^2\right)\left(1 - r_{23}^2\right)}} \quad (6)$$

*for*

$$r_{12} = \frac{n\Sigma x_1 x_2 - \Sigma x_1.\Sigma x_2}{\sqrt{\left[n\Sigma x_1^2 - \left(\Sigma x_1\right)^2\right]\left[n\Sigma x_2^2 - \left(\Sigma x_2\right)^2\right]}} \quad (7)$$

$$r_{13} = \frac{n\Sigma x_1 x_3 - \Sigma x_1.\Sigma x_3}{\sqrt{\left[n\Sigma x_1^2 - \left(\Sigma x_1\right)^2\right]\left[n\Sigma x_3^2 - \left(\Sigma x_3\right)^2\right]}} \quad (8)$$

$$r_{23} = \frac{n\Sigma x_2 x_3 - \Sigma x_2.\Sigma x_3}{\sqrt{\left[n\Sigma x_2^2 - \left(\Sigma x_2\right)^2\right]\left[n\Sigma x_3^2 - \left(\Sigma x_3\right)^2\right]}} \quad (9)$$

where $r_{12}$, $r_{13}$, and $r_{23}$ are the PPM correlation coefficient between consumption and corporate profit, between consumption and private wage bill, and finally between corporate profit and private wage bill.

Also, the second order partial correlation coefficient between consumption and corporate profit holding the influence of private wage bill and gross national product constant could be obtained by the following formula:

$$\ell_{12|36} = \frac{\ell_{12|3} - \ell_{16|3} . \ell_{26|3}}{\sqrt{\left(1 - \ell_{16|3}^2\right)\left(1 - \ell_{26|3}^2\right)}} \quad (10)$$

*for*

$$\ell_{16|3} = \frac{r_{16} - r_{13}.r_{63}}{\sqrt{\left(1 - r_{13}^2\right)\left(1 - r_{63}^2\right)}} \quad (11)$$

$$\ell_{26|3} = \frac{r_{26} - r_{23}.r_{63}}{\sqrt{\left(1 - r_{23}^2\right)\left(1 - r_{63}^2\right)}} \quad (12)$$

where $\ell_{12|3}$, $\ell_{16|3}$, and $\ell_{26|3}$ are the respective first order partial correlation coefficients between consumption and corporate profit given private wage bill, between consumption and gross national product given private wage bill, and between corporate profit and gross national product given private wage bill.

As for the case of third order partial correlation coefficient between consumption and corporate profit holding the influence of private wage bill, gross national product and government wage bill constant, denoted by $\ell_{12|367}$, to the best of our knowledge, no this kind of formula has been developed to handle this correlation in the literature. This indicates that this method is restricted only to first and second partial correlation coefficients.

#### 3.3.2 Computational Procedures From Variance-Covariance Matrix Method

Given a $3 X 3$ matrix of variance-covariance:

$$S = \begin{bmatrix} S_{11} & S_{12} & S_{13} \\ S_{21} & S_{22} & S_{23} \\ S_{31} & S_{32} & S_{33} \end{bmatrix} \quad (13)$$

where each of the S can be obtained from raw datasets as follows:

$$S_{11} = \frac{1}{n-1}\left[\Sigma x_1^2 - \frac{(\Sigma x_1)^2}{n}\right] \quad (14)$$

$$S_{22} = \frac{1}{n-1}\left[\Sigma x_2^2 - \frac{(\Sigma x_2)^2}{n}\right] \quad (15)$$

$$S_{33} = \frac{1}{n-1}\left[\Sigma x_3^2 - \frac{(\Sigma x_3)^2}{n}\right] \quad (16)$$

$$S_{12} = S_{21} = \frac{1}{n-1}\left[\Sigma x_1 x_2 - \frac{(\Sigma x_1)(\Sigma x_2)}{n}\right] \quad (17)$$

$$S_{13} = S_{31} = \frac{1}{n-1}\left[\Sigma x_1 x_3 - \frac{(\Sigma x_1)(\Sigma x_3)}{n}\right] \quad (18)$$

$$S_{23} = S_{32} = \frac{1}{n-1}\left[\Sigma x_2 x_3 - \frac{(\Sigma x_2)(\Sigma x_3)}{n}\right] \quad (19)$$

Under the variance-covariance matrix approach, the first order partial correlation coefficient is computed by:

$$\ell_{12|3} = \frac{S_{12.3}}{\sqrt{(S_{11.3})(S_{22.3})}} \quad (20)$$

*for*

$$S_{12.3} = S_{12} - S_{13}.S_{33}^{-1}.S_{23} \quad (21)$$

$$S_{11.3} = S_{11} - S_{13}.S_{33}^{-1}.S_{13} \quad (22)$$

$$S_{22.3} = S_{22} - S_{23}.S_{33}^{-1}.S_{23} \quad (23)$$

It should be noted that this method could only handle first order partial correlation coefficient. Based on our knowledge, no further formula has been devised yet to handle more than first order partial correlation.

### 3.3.3 Computational Procedures of Regression Residual's Approach

The coefficients of partial correlation of any order can be numerically obtained by the regression residual's method. This approach is efficient in terms of applicability and even in terms of time. It is very easy to understand even by the non-statisticians. Given our datasets, we are expected to compute three different partial correlation coefficients. However, this method can handle all the computational situations as far as partial correlation is concerned. Each of the procedures is tailored as follows:

#### 3.3.3.1 First Order

In an attempt to compute the first order partial correlation coefficient between consumption and corporate profit holding the influence of private wage bill constant, the following steps are necessary:

**Step I:** Regress $x_1$ on $x_3$ and obtain the residual of the model ( $e_1$ )

**Step II:** Regress $x_2$ on $x_3$ again and obtain the residual of the model ( $e_2$ )

**Step III:** Thereafter, obtain the PPM between $e_1$ and $e_2$.

Result obtained as PPM becomes the first order partial correlation coefficient ( $\ell_{12|3}$ ) between consumption and corporate profit when the influence of private wage bill is held constant.

#### 3.3.3.2 Second Order

On the other hand, to obtain the second order partial correlation coefficient between consumption and corporate profit holding the influences of private wage bill and gross national product constant, the steps are tailored as follows:

**Step I:** Regress $x_1$ on $x_3$ and $x_6$ and obtain the residual of the model ( $e_1$ )

**Step II:** Regress $x_2$ on $x_3$ and $x_6$ again and obtain the residual of the model ( $e_2$ )

**Step III:** Thereafter, obtain the PPM between $e_1$ and $e_2$.

Kindly note that result obtained as PPM becomes the second order partial correlation coefficient ( $\ell_{12|36}$ ) between consumption and corporate profit when the influences of private wage bill and gross national product are removed.

#### 3.3.3.3 Third Order

Also, to obtain the third order partial correlation coefficient between consumption and corporate profit holding the influences of private wage bill, gross national product and government wage bill constant, the following steps should be followed:

**Step I:** Regress $x_1$ on $x_3$, $x_6$ and $x_7$ and obtain the residual of the model, denoted by $e_1$

**Step II:** Regress $x_2$ on $x_3$, $x_6$ and $x_7$ again and obtain the residual of the model denoted by $e_2$

**Step III:** Thereafter, compute the PPM between $e_1$ and $e_2$.

Kindly note that result obtained as PPM becomes the third order partial correlation coefficient ( $\ell_{12|367}$ ) between consumption and corporate profit when the influences of private wage bill, gross national product and government wage bill are removed.

### 3.3.4 Computational Procedures of OLS Approach

Besides regression residual's method, partial correlation coefficients could also be obtained from ordinary least square (OLS) approach. This method is also efficient in terms of application and time. Though involves little computational stress if done manually, its numerous advantages cannot be quantified. With the aid of the sophisticated programming software like R, the stress involved would be weightless. This approach handles all orders of partial correlation coefficients, be it first, second, third or more. Here, interest centers on obtaining the estimates of OLS of the first variable on others. However, the computational procedures, for instance, the first order partial correlation, $\ell_{12|3}$, are tailored below:

**Step I:** Obtain the OLS estimates of multiple linear regression model of $x_1$ on $x_2$ and $x_3$

**Step II:** Obtain the estimates of standard error with the corresponding *T-test* for each coefficient

**Step III:** Compute $\ell_{12|3}$ using the formula (24) given below

$$\ell_{12|3} = \frac{T_{x_2}}{\sqrt{T_{x_2}^2 + dfe}} \quad (24)$$

for

$$T_{x2} = \frac{\hat{\beta}_{x_2}}{SE(\hat{\beta}_{x_2})} \quad (25)$$

where $dfe$ means degree of freedom for error, which is computed as $dfe = n - p$ (Note that $p$= number of all the variables in any particular model formed).

However, the same process will be repeated when we are to obtain the second and third partial correlation coefficients. Many of these calculations would be figured out by the use of R programming software version 4.0.3 (R Core Team, 2020).

### 3.4 The R Syntax

The commands written in R language (R Core Team, 2020) to accomplish all the statistical tasks are succinctly detailed in Table 3:

**Table 3**
**The Breakdown of R Language**

| S/N | R commands | Uses | Remarks |
|---|---|---|---|
| 1. | library(AER) | *To recall the library AER before using it* | |
| 2. | data(KleinI) | To figure the dataset into R environment | |
| 3. | View(KleinI) | *To view the dataset* | |
| 4. | head(KleinI) | *To see only the first 6 elements of the data* | |
| 5. | $x_1$ = KleinI[ , 1] | *To extract the variable labeled* $x_1$ | |
| 6. | $x_2$ = KleinI[ , 2] | *To extract the variable labeled* $x_2$ | |
| 7. | $x_3$ = KleinI[ , 3] | To extract the variable labeled $x_3$ | |
| 8. | $x_6$ = KleinI[ , 6] | *To extract the variable labeled* $x_6$ | |
| 9. | $x_7$ = KleinI[ , 7] | *To extract the variable labeled* $x_7$ | |
| 10. | mydata=data.frame($x_1$, $x_2$,$x_3$,$x_6$,$x_7$) | *To combine all the variables* | |
| 11. | cor($x_1$,$x_2$, method = 'pearson') | *PPM between* $x_1$ and $x_2$ | 0.7243439 |
| 12. | cor($x_1$,$x_3$, method = 'pearson') | *PPM between* $x_1$ and $x_3$ | 0.9664623 |
| 13 | cor($x_1$,$x_6$, method = 'pearson') | *PPM between* $x_1$ and $x_6$ | 0.9604946 |
| 14. | cor( $x_2$,$x_3$, method = 'pearson') | *PPM between* $x_2$ and $x_3$ | 0.7626755 |
| 15. | cor( $x_2$, $x_6$, method = 'pearson') | *PPM between* $x_2$ and $x_6$ | 0.8409359 |
| 16. | cor($x_3$, $x_6$, method = 'pearson') | *PPM between* $x_3$ and $x_6$ | 0.9815224 |
| 17. | library(ggm) | To recall the library ggn before using it | |
| 18. | pcor(c("$x_1$", "$x_2$", "$x_3$"), var(mydata)) | *To obtain* $\ell_{12\|3}$ | -0.0767811 |
| 19. | pcor(c("$x_1$", "$x_2$", "$x_3$", "$x_6$"), var(mydata)) | *To obtain* $\ell_{12\|36}$ | -0.3984185 |
| 20. | pcor(c("$x_1$", "$x_2$", "$x_3$", "$x_6$", "$x_7$"), var(mydata)) | *To obtain* $\ell_{12\|367}$ | 0.1861561 |

*Source:* Authors' computation (2021)

## 4. RESULTS AND DISCUSSION

### 4.1 Confirmation of Assumptions

We examine the datasets by checking the validity of all the basic assumptions of partial correlation before obtaining any order of the coefficient. This examination of each assumption is detailed as follows:

#### 4.1.1 Checking Linearity Assumption

The following visualizations (Figures I-IV) explain the linearity and or otherwise of our datasets.

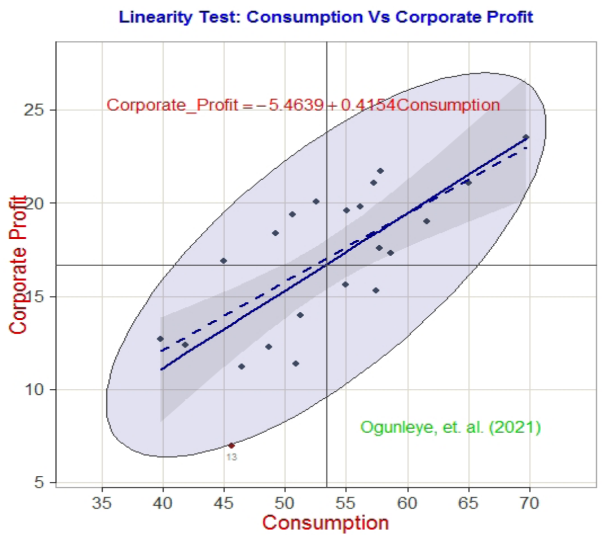


**Figure 1**
**Testing for Linearity between Consumption and Corporate Profit**

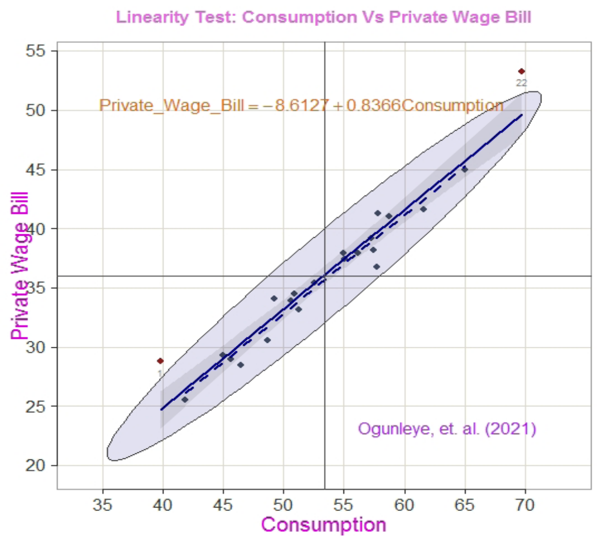


**Figure 2**
**Testing for Linearity between Consumption and Private Wage Bill**

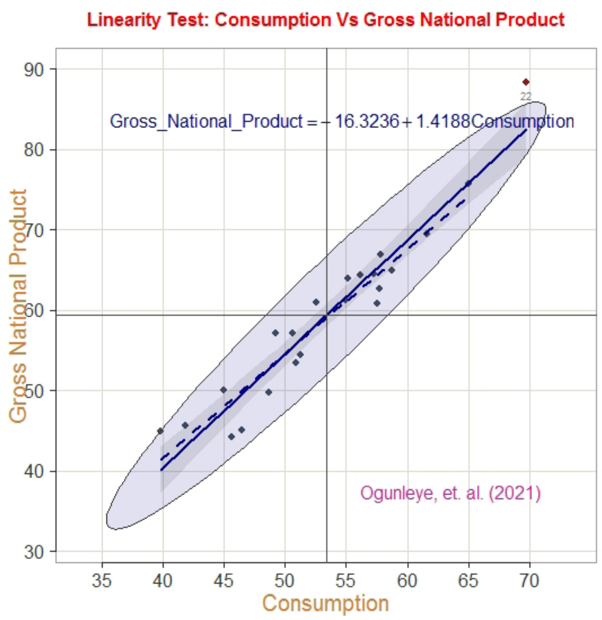


**Figure 3**
**Testing for Linearity between Consumption and Gross National Product**

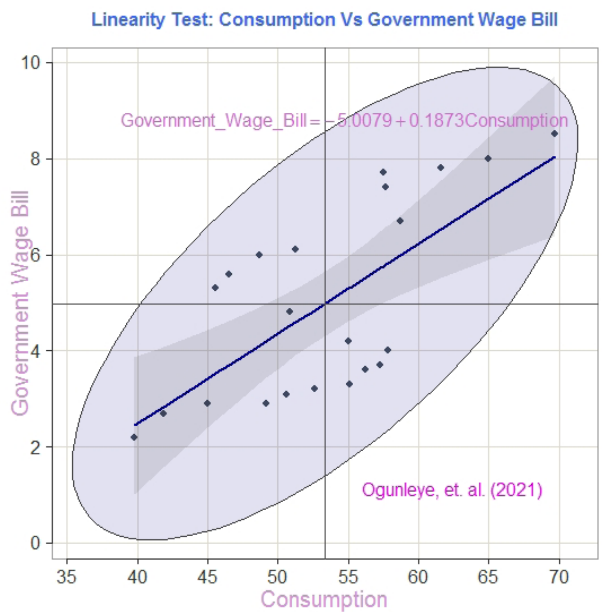


**Figure 4**
**Testing for Linearity between Consumption and Government Wage Bill**

Critical examination of the four infographics (Figures I-IV) shows that linearity assumption is not violated since almost all the data points fall within the eclipse. We therefore conclude that this assumption is upheld.

### 4.1.2 Ascertaining Normality Assumption

The following results in Table 4 show the evidences of upholding normality assumption as reported by the use of Shapiro-Wilk statistical test.

**Table 4**
**Reports of Normality Test**

| S/N | Variable | Shapiro-wilk statistic (W) | P-value | Remarks |
|---|---|---|---|---|
| 1. | Consumption | 0.98453 | 0.9713 | No violation of normality assumption |
| 2. | Corporate Profit | 0.95829 | 0.4555 | No violation of normality assumption |
| 3. | Private Wage Bill | 0.95412 | 0.3802 | No violation of normality assumption |
| 4. | Gross National Product | 0.98820 | 0.1817 | No violation of normality assumption |
| 5. | Government Wage Bill | 0.91626 | 0.0636 | No violation of normality assumption |

*Source:* Authors' computation (2021)

### 4.1.3 Checking Existence of Outliers

The third assumption is to check if outlying effect is significant in the datasets used. This could be achieved by employing the use of box plot. Report of this is presented in Figure V.

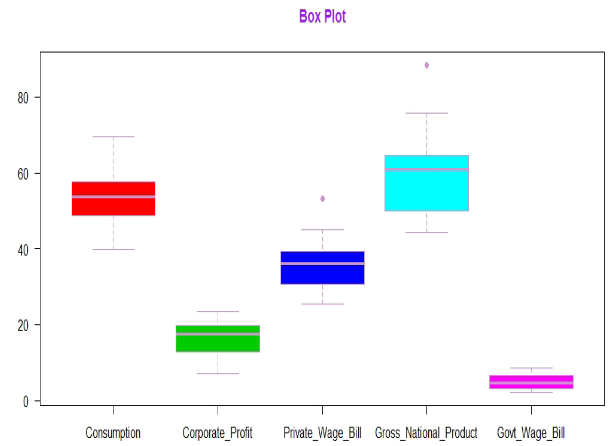


Figure V: Testing for Outliers' Effect on the Datasets

**Figure 5**
**Testing for outliers' effect on the datasets**

Careful study of the Figure V shows that the two outliers surveyed could be still be accommodated since they are scattered evenly between private wage bill and gross national product. Therefore, the outliers are said to be tolerable and the analysis by partial correlation is still perfectly okay for the datasets.

### 4.2 Computation of 1st Order Partial Correlation Coefficient

We compute the first order partial correlation coefficient by each of the four methods theoretically discussed in our earlier section. All the methods discussed in this study can handle the first order.

### 4.2.1 Computation by Conventional Approach

The following statistics are obtained from our datasets with R syntax:

$$\Sigma x_1 = 1173.700 \quad \Sigma x_2 = 367.4000 \quad \Sigma x_3 = 792.4000$$
$$\Sigma x_1^2 = 63750.67 \quad \Sigma x_2^2 = 6508.540 \quad \Sigma x_3^2 = 29390.30$$
$$\Sigma x_1 x_2 = 20071.81 \quad \Sigma x_1 x_3 = 43223.02 \quad \Sigma x_2 x_3 = 13662.37$$

The matrix of PPM correlation coefficients for $x_1$, $x_2$ and $x_3$ is obtained as follows:

$$R = \begin{pmatrix} 1 & .7243 & .9664 \\ & 1 & .7627 \\ & & 1 \end{pmatrix}$$

Therefore, the first order partial correlation coefficient is obtained as follows:

$$\ell_{12|3} = \frac{r_{12} - r_{13} \cdot r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}} = -0.0768 \equiv -0.08 \equiv -8\%$$

### 3.2.2 Computation by Variance-covariance Matrix Method

The following matrix of variance-covariance for $x_1$, $x_2$ and $x_3$ is obtained via R engine:

$$S = \begin{pmatrix} 53.9893 & 22.4295 & 45.1657 \\ 22.4295 & 17.7600 & 20.4424 \\ 45.1657 & 20.4424 & 40.4520 \end{pmatrix}$$

From $S$, compute the following:

$$S_{12.3} = S_{12} - S_{13} . S_{33}^{-1} . S_{23} = -0.394966$$

$$S_{11.3} = S_{11} - S_{13} . S_{33}^{-1} . S_{13} = 3.560533$$

$$S_{22.3} = S_{22} - S_{23} . S_{33}^{-1} . S_{23} = 7.429442$$

Therefore, the first order partial correlation coefficient is obtained as follows:

$$\ell_{12|3} = \frac{S_{12.3}}{\sqrt{(S_{11.3})(S_{22.3})}} = -0.0768 \equiv -0.08 \equiv -8\%$$

### 4.2.3 Computation by Regression Residual's Approach

Within R environment, the first and second regression models are labeled "regression.1" and "regression.2" while their residuals are tagged "residual.1" and "residual.2" respectively. Therefore, the codes are:

```
regression.1 = lm (x1 ~ x3); summary(regression.1)

regression.2 = lm (x2 ~ x3); summary(regression.2)

residual.1=resid(regression.1); write.csv(residual.1, "RES.1.csv")
residual.2=resid(regression.2); write.csv(residual.2, "RES.2.csv")
cor(residual.1, residual.2, method="pearson")
```

The results for these codes are reported in Tables 5 - 7.

**Table 5**
**Report of Linear Regression Model of $x_1$ on $x_3$**

| Parameters | Estimates | Standard Error | T-value | P-value |
|---|---|---|---|---|
| Intercept | 13.1348 | 2.42476 | 5.417 | 0.00003 |
| Coefficient of $x_3$ | 1.1165 | 0.06634 | 16.830 | 0.00000 |

Multiple R-squared = 0.934; Multiple R-squared (Adjusted) = 0.9308; F-statistic = 283.3, Overall P-value = 0.00000
The model is $x_1 = 13.1348 + 1.1165 x_3$

*Source:* Authors' computation (2021)

**Table 6**
**Report of Linear Regression Model of $x_2$ on $x_3$**

| Parameters | Estimates | Standard Error | T-value | P-value |
|---|---|---|---|---|
| Intercept | - 1.1502 | 3.50255 | - 0.4290 | 0.67300 |
| Coefficient of $x_3$ | 0.5054 | 0.09583 | 5.273 | 0.00000 |

Multiple R-squared = 0.5817; Multiple R-squared (Adjusted) = 0.5608; F-statistic = 27.81, Overall P-value = 0.00000
The model is $x_2 = -1.5017 + 0.5054x_3$

*Source:* Authors' computation (2021)

**Table 7**
**Report of Errors for Linear Regression Models**

| S/N | Residual.1 | Residual.2 | Product Moment Correlation (PPM) of the Residuals |
|---|---|---|---|
| 1. | -5.4907 | -0.3523 | |
| 2. | 0.29382 | 1.01535 | |
| 3. | -0.8490 | 3.59502 | |
| 4. | -2.0083 | 2.66935 | |
| 5. | -0.3850 | 3.77042 | |
| 6. | -0.0598 | 3.71240 | |
| 7. | 0.20716 | 2.20170 | |
| 8. | 0.74890 | 2.14903 | |
| 9. | 0.39742 | 2.79207 | $r_{e_1e_2} = \frac{n\Sigma e_1 e_2 - \Sigma e_1.\Sigma e_2}{\sqrt{\left[n\Sigma e_1^2 - (\Sigma e_1)^2\right]\left[n\Sigma e_2^2 - (\Sigma e_2)^2\right]}}$ |
| 10. | -1.4473 | 2.33084 | |
| 11. | -0.4511 | -2.05097 | |
| 12. | -0.7549 | -4.53279 | |
| 13. | 0.08598 | -6.15337 | $\therefore r_{e_1e_2} = -0.07678112 \cong -0.08 \equiv -8\%$ |
| 14. | 1.54424 | -1.70070 | |
| 15. | 1.39954 | -1.66193 | |
| 16. | 1.09657 | -1.27584 | |
| 17. | 3.47708 | 0.504909 | |
| 18. | -0.2123 | -1.91756 | |
| 19. | 1.71395 | -2.50258 | |
| 20. | 2.01776 | -0.52076 | |
| 21. | 1.62157 | -0.13895 | |
| 22. | -2.9456 | -1.93334 | |

*Source*: Authors' computation (2021)

### 4.2.4 Computation by OLS Approach

The parameter estimation of multiple linear regression model of $x_1$ on $x_2$ and $x_3$ is obtained via the use of R syntax as follows:

**Table 8**
**Report of Multiple Linear Regression Model of $x_1$ on $x_2$ and $x_3$**

| Parameters | Estimates | Standard Error | T-value | P-value |
|---|---|---|---|---|
| Intercept | 13.05497 | 2.49178 | 5.2390 | 0.00000 |
| Coefficient of $x_2$ | - 0.0532 | 0.15835 | - 0.336 | 0.74100 |
| Coefficient of $x_3$ | 1.14339 | 0.10492 | 10.8970 | 0.00000 |

Multiple R-squared = 0.9344; Multiple R-squared (Adjusted) = 0.9275; F-statistic = 135.4, Overall P-value = 0.00000
The model is $x_1 = 13.05497 - 0.0532x_2 + 1.14339x_3$

*Source:* Authors' computation (2021)

Therefore, we compute the first order partial correlation as follows:

$$\ell_{12|3} = \frac{T_{x_2}}{\sqrt{T_{x_2}^2 + dfe}} = -\frac{0.336}{\sqrt{(-0.336)^2 + 19}} = -0.07685569 \cong -0.08 \equiv -8\%$$

## 4.3 Computation of 2nd Order Partial Correlation Coefficient

Here, we obtain the second order partial correlation coefficient by all the methods except that of variance-covariance approach. Each of the remaining methods is tailored as follows:

### 4.3.1 Computation by Conventional Approach

The matrix of PPM correlation coefficients for $x_1$, $x_2$, $x_3$ and $x_6$ is obtained as follows:

$$R = \begin{pmatrix} 1 & 0.7243 & 0.9665 & 0.9605 \\ & 1 & 0.7627 & 0.8409 \\ & & 1 & 0.9815 \\ & & & 1 \end{pmatrix}$$

Therefore, the second order partial correlation coefficient is obtained as follows:

$$\ell_{12|36} = \frac{\ell_{12|3} - \ell_{16|3} \cdot \ell_{26|3}}{\sqrt{\left(1 - \ell^2_{16|3}\right)\left(1 - \ell^2_{26|3}\right)}} = -0.3984185 \cong -0.40 \equiv -40\%$$

### 4.3.2 Computation by Regression Residual's Approach

Let's represent the first and second regression models by "regression.11" and "regression.22" while their residuals could be named "residual.11" and "residual.22" within R environment. Therefore, the codes are:

```
regression.11 = lm(x1 ~ x3+ x6); summary(regression.11)

regression.22 = lm(x2 ~ x3+ x6); summary(regression.22)

residual.11=resid(regression.11); write.csv(residual.11, "RES.11.csv")

residual.22=resid(regression.22); write.csv(residual.22, "RES.22.csv")

cor(residual.11, residual.22, method="pearson")
```

The results for these codes are reported in Tables 9 - 11.

**Table 9**
**Report of Multiple Linear Regression Model of $x_1$ on $x_3$, and $x_6$**

| Parameters | Estimates | Standard error | T-value | P-value |
|---|---|---|---|---|
| Intercept | 13.3460 | 2.4216 | 5.511000 | 0.00000 |
| Coefficient of $x_3$ | 0.74830 | 0.3451 | 2.168000 | 0.04310 |
| Coefficient of $x_6$ | 0.21980 | 0.2023 | 1.087000 | 0.29060 |

Multiple R-squared = 0.9397; Multiple R-squared (Adjusted) = 0.9314; F-statistic = 143.5, Overall P-value = 0.00000
The model is $x_1 = 13.3460 + 0.74830x_3 + 0.21980x_6$

*Source*: Authors' computation (2021)

**Table 10**
**Report of Multiple Linear Regression Model of $x_2$ on $x_3$, and $x_6$**

| Parameters | Estimates | Standard error | T-value | P-value |
|---|---|---|---|---|
| Intercept | - 0.5609 | 2.39990 | - 0.2340 | 0.81778 |
| Coefficient of $x_3$ | - 1.1351 | 0.34200 | - 3.3180 | 0.00361 |
| Coefficient of $x_6$ | 0.9794 | 0.2004 | 4.88600 | 0.00010 |

Multiple R-squared = 0.8146; Multiple R-squared (Adjusted) = 0.7951; F-statistic = 41.75, Overall P-value = 0.00000
The model is $x_2 = -0.5609 - 1.1351x_3 + 0.9794x_6$

*Source:* Authors' computation (2021)

**Table 11**
**Report of Errors for the two Multiple Linear Regression Models**

| S/N | Residual.1 | Residual.2 | Product moment correlation (PPM) of the residuals |
|---|---|---|---|
| 1. | -4.96791 | 1.976717 | |
| 2. | -0.55245 | -2.7546 | |
| 3. | -1.28527 | 1.651462 | |
| 4. | -2.23798 | 1.646201 | |
| 5. | -0.66633 | 2.517124 | |
| 6. | -0.64618 | 1.100142 | |
| 7. | -0.3023 | -0.06786 | |
| 8. | 0.335618 | 0.307926 | $r_{e_1e_2} = \frac{n\Sigma e_1e_2 - \Sigma e_1.\Sigma e_2}{\sqrt{\left[n\Sigma e_1^2 - (\Sigma e_1)^2\right]\left[n\Sigma e_2^2 - (\Sigma e_2)^2\right]}}$ |
| 9. | 0.440859 | 2.985585 | |
| 10. | -1.18017 | 3.520785 | |
| 11. | -0.16086 | -0.75805 | |
| 12. | -0.00186 | -1.17811 | $\therefore r_{e_1e_2} = -0.3984185 \cong -0.40 \equiv -40\%$ |
| 13. | 0.814345 | -2.90864 | |
| 14. | 1.91261 | -0.05968 | |
| 15. | 1.5299 | -1.08119 | |
| 16. | 1.151061 | -1.03309 | |
| 17. | 3.032478 | -1.4757 | |
| 18. | 0.384017 | 0.739029 | |
| 19. | 2.180604 | -0.42371 | |
| 20. | 1.845723 | -1.28715 | |
| 21. | 1.338481 | -1.40007 | |
| 22. | -2.96439 | -2.01712 | |

*Source*: Authors' computation (2021)

### 4.3.3 Computation by OLS Approach

The parameter estimation of multiple linear regression model of $x_1$ on $x_2$ and $x_3$ is obtained via the use of R syntax as follows:

**Table 12**
**Report of Multiple Linear Regression Model of $x_1$ on $x_2$, $x_3$ and $x_6$**

| Parameters | Estimates | Standard error | T-value | P-value |
|---|---|---|---|---|
| Intercept | 13.1205 | 2.2853 | 5.7410 | 0.00000 |
| Coefficient of $x_2$ | - 0.4020 | 0.2181 | - 1.843 | 0.08190 |
| Coefficient of $x_3$ | 0.29200 | 0.4088 | 0.7140 | 0.48420 |
| Coefficient of $x_6$ | 0.61360 | 0.2863 | 2.1430 | 0.04600 |

Multiple R-squared = 0.9478; Multiple R-squared (Adjusted) = 0.9391; F-statistic = 108.9, Overall P-value = 0.00000
The model is

$$x_1 = 13.1205 - 0.4020x_2 + 0.2920x_3 + 0.6136x_6$$

*Source*: Authors' computation (2021)

Therefore, we compute the first order partial correlation as follows:

$$\ell_{12|36} = \frac{T_{x_2}}{\sqrt{T_{x_2}^2 + dfe}} = -\frac{1.843}{\sqrt{(-1.843)^2 + 18}} = -0.398430387 \cong -0.40 \equiv -40\%$$

### 4.4 Computation of 3rd Order Partial Correlation Coefficient

In handling the computation of third order partial correlation coefficient, only two methods: regression residual's approach and OLS method. The remaining two methods could not survive in this scenario. Results of the practical illustrations are presented as follows:

#### 4.4.1 Regression Residual's Approach

Let's the first and second regression models be denoted by "regression.111" and "regression.222" while their residuals could be labeled as "residual.111" and "residual.222" while using R engine. The codes are:

```
regression.111 = lm(x1 ~ x3+ x6 + x7); summary(regression.111)

regression.222 = lm(x2 ~ x3+ x6 + x7); summary(regression.222)

residual.111=resid(regression.111); write.csv(residual.111, "RES.111.csv")

residual.222=resid(regression.222); write.csv(residual.222, "RES.222.csv")

cor(residual.111, residual.222, method="pearson")
```

**Table 13**
**Report of Multiple Linear Regression Model of $x_1$ on $x_3$, $x_6$ and $x_7$**

| Parameters | Estimates | Standard error | T-value | P-value |
|---|---|---|---|---|
| Intercept | 14.9006 | 1.7872 | 8.337000 | 0.00000 |
| Coefficient of $x_3$ | 0.26830 | 0.2734 | 0.981000 | 0.33943 |
| Coefficient of $x_6$ | 0.41480 | 1.1531 | 2.710000 | 0.01436 |
| Coefficient of $x_7$ | 0.83480 | 0.1947 | 4.287000 | 0.00044 |

Multiple R-squared = 0.9693; Multiple R-squared (Adjusted) = 0.9642; F-statistic = 189.3, Overall P-value = 0.00000
The model is

$$x_1 = 14.9006 + 0.26830x_3 + 0.41480x_6 + 0.83480x_7$$

*Source*: Authors' computation (2021)

**Table 14**
**Report of Multiple Linear Regression Model of $x_2$ on $x_3$, $x_6$ and $x_7$**

| Parameters | Estimates | Standard error | T-value | P-value |
|---|---|---|---|---|
| Intercept | - 2.0636 | 1.8147 | - 1.1370 | 0.2704 |
| Coefficient of $x_3$ | - 0.6711 | 0.2776 | - 2.4180 | 0.0265 |
| Coefficient of $x_6$ | 0.79100 | 0.1554 | 5.08900 | 0.0000 |
| Coefficient of $x_7$ | - 0.8070 | 0.1977 | - 4.0820 | 0.0007 |

Multiple R-squared = 0.9037; Multiple R-squared (Adjusted) = 0.8877; F-statistic = 56.32, Overall P-value = 0.00000
The model is

$$x_1 = -2.0636 - 0.6711x_3 + 0.7910x_6 - 0.8070x_7$$

*Source*: Authors' computation (2021)

**Table 15**
**Report of Errors for the two Multiple Linear Regression Models**

| S/N | Residual.1 | Residual.2 | Product moment correlation (PPM) of the residuals |
|---|---|---|---|
| 1. | -3.2866 | 0.351488 | |
| 2. | -1.00893 | -2.31335 | |
| 3. | -0.96182 | 1.338801 | |
| 4. | -0.9944 | 0.444106 | |
| 5. | 0.333773 | 1.550375 | |
| 6. | 0.230309 | 0.252891 | |
| 7. | 0.865971 | -1.19716 | |
| 8. | 1.41548 | -0.73592 | $r_{e_1e_2} = \frac{n\Sigma e_1e_2 - \Sigma e_1.\Sigma e_2}{\sqrt{\left[n\Sigma e_1^2 - (\Sigma e_1)^2\right]\left[n\Sigma e_2^2 - (\Sigma e_2)^2\right]}}$ |
| 9. | 2.041734 | 1.438107 | |
| 10. | 0.690982 | 1.712046 | |
| 11. | 1.041809 | -1.92061 | |
| 12. | 0.588221 | -1.74852 | |
| 13. | 0.120724 | -2.23815 | $\therefore r_{e_1e_2} = 0.1861561 \cong 0.19 \equiv 19\%$ |
| 14. | 0.572634 | 1.2356 | |
| 15. | -0.03258 | 0.429174 | |
| 16. | -0.16297 | 0.237113 | |
| 17. | 0.743464 | 0.736961 | |
| 18. | 0.247026 | 0.87145 | |
| 19. | 0.663957 | 1.04235 | |
| 20. | 0.201385 | 0.302341 | |
| 21. | -0.04926 | -0.05862 | |
| 22. | -3.26091 | -1.73049 | |

*Source:* Authors' computation (2021)

### 4.3.3 Computation by OLS Approach

The parameter estimation of multiple linear regression model of $x_1$ on $x_2$ and $x_3$ is obtained via the use of R syntax as follows:

**Table 12**
**Report of Multiple Linear Regression Model of $x_1$ on $x_2$, $x_3$, $x_6$ and $x_7$**

| Parameters | Estimates | Standard error | T-value | P-value |
|---|---|---|---|---|
| Intercept | 15.2790 | 1.8707 | 8.1680 | 0.0000 |
| Coefficient of $x_2$ | 0.18330 | 0.2347 | 0.7810 | 0.4454 |
| Coefficient of $x_3$ | 0.39130 | 0.3181 | 1.2300 | 0.2354 |
| Coefficient of $x_6$ | 0.26970 | 0.2417 | 1.1160 | 0.2799 |
| Coefficient of $x_7$ | 0.98270 | 0.2732 | 3.5980 | 0.0022 |

Multiple R-squared = 0.9703; Multiple R-squared (Adjusted) = 0.9634; F-statistic = 139.1, Overall P-value = 0.00000
The model is

$$x_1 = 15.2790 + 0.1833x_2 + 0.3913x_3 + 0.2697x_6 + 0.9827x_7$$

*Source:* Authors' computation (2021)

Therefore, we compute the first order partial correlation as follows:

$$\ell_{12|367} = \frac{T_{x_2}}{\sqrt{T_{x_2}^2 + dfe}} = -\frac{0.781}{\sqrt{(0.781)^2 + 17}} = 0.186110911 \cong 0.19 \equiv 19\%$$

## 5. CONCLUSION AND RECOMMENDATION

In our research, we have extensively discussed, with detailed illustrations, the computational techniques both manually and electronically to obtain the first, second, and third partial correlation coefficients. Our datasets are obtained from R engine. To the best of our knowledge, we have demonstrated the four different methods of computing partial correlations, and recommended the best in a number of scenarios.

When our interest is centered on computing first order partial correlation coefficient, all the four methods discussed can flow but the conventional method is recommended to be least time-consuming. But when we are interested in computing the second order, only three methods: conventional, regression residual's, and OLS methods can handle such with recommendation that OLS method is the best amongst all in terms of time and computational stress.

However, from our practical illustrations, conventional method drops out of the three that reach the second stage, keeping only two: regression residual's, and OLS methods, while interested in computing higher order partial correlation coefficients. Therefore, we recommend the use of regression residual's approach and OLS method to obtain third and higher order partial correlation coefficients. Both methods are straight-forward, reliable, understandable, explicit, and detailed to account for any order of partial correlations.

## REFERENCES


Akoglu, H. (2018). User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine*, *18*(3), 91-93.

Aloe, A.M. (2013). The synthesis of partial effect sizes. *Journal of the Society for Social Work and Research, 4*(4), 390-405.

Bishara, A.J., & James B.H. (2017). Confidence intervals for correlations when data are not normal. *Behavior Research Methods, 49*(1), 294-309. Retrieved from https://doi.org/10.3758/s13428-016-0702-8.

Egozcue, J. J., Pawlowsky-Glahn, V., & Gloor, G. B. (2018). Linear association in compositional data analysis. *Austrian Journal of Statistics*, 47 (1), Pp. 3-31.

Fieller, E. C., Herman, O. H., & Egon, S. P. (1957). Tests for rank correlation coefficients I. *Biometrika, 44*(3/4), 470- 81. Retrieved from https://doi.org/10.1093/biomet/48.1-2.29.

Fisher, R.A. (1924). The distribution of the partial correlation coefficient. *Metron, 3*(3-4), 329-332.

Guilford, J.P., & Fruchter, B. (1973). Fundamental statistics in psychology and education. *McGraw-Hill*, Kogakusha, Ltd.

Ha, M.J., & Sun, W. (2014). Partial correlation matrix estimation using ridge penalty followed by threshold and re-estimation. *Biometrics*, *70*(3), 762-770.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2018). Applied statistics for the behavioral sciences (8th ed.). *Boston: Houghton Mifflin.*

Horwitz, B., & Rapoport, S. I. (1988). Partial correlation coefficients approximate the real intrasugject correlation pattern in the analysis of interregional relations of cerebral metabolic activity. *Journal of Nuclear Medicine*, *29*(3), 392-399.

Hotelling, H. (1936). Relations between two sets of variates. *Biometrika*, *28*(3-4), 321-377. doi:10.1093/biomet/28.3- 4.321. JSTOR 2333955.

Kenett, D. Y., Huang, X., Vodenska, I., Havlin, S., & Stanley, E. (2015). Partial correlation analysis: application for financial market. *Journal of Quantitative Finance*, *15*(4), 569-578.

Kynčlová, P., Hron, K., & Filzmoser, P. (2017). Correlation between compositional parts based on symmetric balances. *Mathematical Geosciences, 49*, 777-796.

Langfelder, P., & Steve, H. (2012). Fast R functions for robust correlations and hierarchical clustering. *Journal of Statistical Software, 46*(11). Retrieved from https://www.jstatsoft.org/v46/i11/.

Lonas, E. (2020). Partial correlations in compositional data analysis. *Applied Computing and Geosciences*, *6*, 100026.

Merrian-webster dictionary (2020). *Definition of correlation.* Available from https://www.merriam- webster.com/dictionary/correlation. Downloaded on 10th January, 2020.

Mukaka, M. M. (2012). A guide to appropriate use of correlation coefficient in medical research. *Malawi Medical Journal*, *24*(3), 69-71.

Pedhazur, E. J. (1982). *Multiple regression in behavioural research: Explanation and prediction* (2nd ed.). New York: Holt, Rinehart Winston.

R Core Team (2020). R: A language and environment for statistical computing Vienna, Austria: *R foundation for statistical computing.* Retrieved from https://www.R-project.org/.

Roverato, A., & Castelo, R. (2017). The networked partial correlation and its application to the analysis of genetic interactions. *Journal of the Royal Statistical Society: Applied Statistics*., Series C, *66*(3), 647-665.

Serlin, R. C., & Harwell, M. R. (2007). An empirical study of 8 tests of partial correlation coefficients. *Communication in Statistics-Simulation and Computation, 22*(2), 545-567.

Waliczec, T. M. (1996). *A primer on partial correlation coefficients.* A paper presented at the annual meeting of the Southwest Educational Research Association, New Orleans, January.

Yule, G. U. (1907). On the theory of correlation for any number of variables treated by a new system of notation. *Roy. Soc. Proc. A.* LXXIX, 182-193.