COMPARATIVE ANALYSIS ON THE DIFFERENT METHODS OF DETECTING MULTICOLLINEARITY

[1]Joshua C. Guardaquivil* and [2]Resa Mae R. Sangco

Abstract

This study explores four methods for detecting multicollinearity: Min-Max and Point-Coordinates approach, Product Moment Correlation, Eigenvalues, and Variance Inflation Factor. Since the Min-Max and Point-Coordinates approach is a new method for detecting multicollinearity, its results were compared and verified against the results obtained from the existing methods. Longley's Economic Data and Blood Pressure Data were used to evaluate the suitability of the Min-Max and Point-Coordinates approach for detecting multicollinearity. The results indicate that severe multicollinearity can significantly affect model estimation and interpretation, highlighting the importance of detecting it accurately. The study shows that the Min-Max and Point-Coordinates approach is effective in detecting multicollinearity, making it a valuable addition to the existing methods. Overall, this study provides useful insights into the importance of detecting multicollinearity and the efficacy of different methods for doing so.

Keywords: *min-max and point-coordinates, multicollinearity, product moment correlation, eigenvalues, variance inflation factor*

*Corresponding Author: Joshua C. Guardaquivil, guardaquivil.joshua@gmail.com

1.0 Introduction

Statistical modeling is used to make predictions of future values that involve one or more independent variables. In many applications, the assumption of multicollinearity in modeling is a step that some researchers neglect, which can lead to inappropriate models and erroneous conclusions. According to some literature, multicollinearity may result in insignificant parameters in the model (Belsley, 1991; Kennedy, 2002; Gujarati, 2004; Kumar, 1975). Belsley (1991) noted that multicollinearity can cause considerable changes in a model, even resulting in changes in the sign of parameter estimates. In dealing with modeling, multicollinearity is one problem encountered by researchers that violates the statistical robustness of the models. Moderate multicollinearity may not be a problem, but severe multicollinearity is a problem because it can increase the variance of coefficient estimates and make estimates sensitive to model changes, resulting in instability in coefficient estimates that can be difficult to understand (Jim, 2013).

Kock and Lynn (2012) suggest that regression models are most effective when the independent variables have a strong correlation with the dependent variable, but a minimal correlation with each other. Such models, known as "low noise" models, are statistically robust. However, multicollinearity often arises when researchers work with models that include multiple independent variables. Multicollinearity is characterized by high and significant relationships between independent variables in a model (Draper & Smith, 1981; Duncombe *et al.*, 1986). To address this issue, various methods have been developed for detecting multicollinearity, including remodeling and parsimony.

Despite the availability of various methods for detecting multicollinearity in statistical models, there is still a need to compare and evaluate the effectiveness of these methods (Kutner *et al.,* 2004). The importance of this investigation lies in the fact that the assumption of multicollinearity is often overlooked, resulting in erroneous conclusions and inappropriate models. Furthermore, while traditional methods such as Product Moment Correlation, Eigenvalues, and Variance Inflation Factor have been widely used and proven to be suitable, a new approach using a 5-steps algorithm has been introduced (Enders, 2010). It is important to establish how this new approach compares to the existing methods and to determine which approach is more efficient and user-friendly. By establishing the gap and conducting a comparative analysis, this study aims to contribute to the field of statistics by providing a comprehensive evaluation of the different methods of detecting multicollinearity and presenting a more user-friendly approach to researchers. Moreover, most of the existing methods are complex and require advanced mathematical knowledge, making them difficult for non-experts to use. This study aims to bridge this gap by comparing and evaluating the results of different methods of detecting multicollinearity. Additionally, the study applies these methods to Longley's Economic Data (Longley, 1967) and Blood Pressure Data (R Core Team, 2018) to determine the suitability of the Min-Max and Point-Coordinates Approach for detecting multicollinearity. These datasets include a single outcome variable and multiple predictor variables, making them ideal for testing for multicollinearity.

2.0 Methodology

Multicollinearity detection methods such as Product Moment Correlation, Eigenvalue, and Variance Inflation Factor have been commonly used by researchers in various fields, including financial models, market price fluctuations, and survival analysis (Abidemi *et al.,* 2016). In this study, we introduce the Min-Max and point-coordinates method as an alternative approach to detect multicollinearity.

Min-Max and Point-Coordinates

The proposed method's brief algorithm is as follows:

1. Rearrange the observations in ascending order.
2. Attach serial numbers to the observations.
3. Locate minimum and maximum observations of variables.
4. Plot scatter diagram of least serial and maximum serial numbers against min-max of variables.
5. Prove that the lines are parallel using coordinates of points.

The gradient is equivalent to the ratio of range of variables if minimum and maximum points are used in the computation of gradient, that is,

$$\text{Gradient} = \frac{\Delta X}{\Delta Y} = \frac{Range\ (x)}{Range\ (y)}$$

where y is the dependent variable and x is the independent variable and range is computed using the formula Maximum – Minimum (Statistics How To, 2018; Math is Fun, n. d.; Protter & Protter, 1988).

One of the assumptions of the Min-Max and point -coordinates method is that when two lines have the same slope, they are considered parallel. Additionally, if the product of the slopes of two lines is equal to -1, then they are considered perpendicular. These assumptions are based on the principles of geometry (Math Centre, 2009).

Product Moment Correlation

Correlation is a method to determine the strength of the connection between two variables, and it is measured by the correlation coefficient, which ranges from -1 to +1. The closer the correlation coefficient to 1, the stronger the relationship between the variables. On the other hand, the closer it is to 0, the weaker the relationship. The Pearson r correlation is commonly used to calculate the degree of the relationship between two sets of data, especially when they are linearly related. The following

[1]*Mindanao State University*
[2]*Northeastern Mindanao State University*

ISSN: 2244-6990 (Print) 2408-3577 (Online)

formula is used to calculate the Pearson's *r* correlation:

$$r = \frac{n\left(\sum_{i,j}^{n} x_i y_j\right) - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{j=1}^{n} y_j\right)}{\sqrt{\left[n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right]\left[n\sum_{j=1}^{n} y_j^2 - \left(\sum_{j=1}^{n} y_j\right)^2\right]}}$$

where i=1,2,…,n and j=1,2,…,n. The number of observations in the data set is denoted by n, r for Pearson's correlation coefficient.

The correlation coefficient, which is a statistic that ranges from -1 to +1, measures the strength of the association between two variables. A correlation coefficient value around 1 suggests a perfect degree of relationship between the two variables, while a value approaching 0 indicates a weaker association. The Pearson *r* correlation is widely used to measure the degree of the linear relationship between two sets of data in statistics. However, several assumptions need to be considered, including normal distribution of both variables, linearity, and homoscedasticity.

Linearity assumes a straight-line relationship between the variables, and homoscedasticity assumes that data is normally distributed about the regression line. A correlation coefficient between .10 and .29 represents a small association, while coefficients between .30 and .49 represent a medium association, and coefficients above .50 represent a large relationship. The results range between -1 and +1, and the closer the value of *r* is to zero, the greater the variation of the data points around the line of best fit (Statistics Solutions, 2015; Chen & Popovich, 2002).

Eigenvalues

Eigenvalues, which are also referred to as proper values, characteristic values, or latent roots, are scalar quantities associated with linear systems of equations, specifically matrix equations. Multicollinearity can be detected using eigenvalues as well. If the predictor variables have multicollinearity, one or more of the eigenvalues will be small, approaching zero. Let $\lambda_1, \lambda_2, \ldots, \lambda_p$ be the eigenvalues of the correlation matrix. Compute the eigenvalues of X'X in solving for λ's by,

$$|X'X - \lambda I| = 0.$$

the condition number or condition indices of correlation matrix is defined as

$$K = \sqrt{\left(\frac{\lambda_{max}}{\lambda_{min}}\right)} \;;$$

$$K = \sqrt{\left(\frac{\lambda_{max}}{\lambda_j}\right)}, j = 1, 2, \ldots, p.$$

To detect multicollinearity, one can examine the eigenvalues associated with the linear system of equations (i.e., a matrix equation). If one or more of the eigenvalues are close to zero, it indicates the presence of multicollinearity among the predictor variables (Strang, 2010). Additionally, the condition number or condition indices can be examined. A condition index greater than 15 indicates a possible problem with collinearity, while a value greater than 30 suggests a serious problem (Joshi, 2012).

Variance Inflation Factor

A common and formal technique to detect multicollinearity is through Variance Inflation Factors (VIF). VIF is used to measure how much the estimated regression coefficient variances inflate when predictor variables are linearly related. This method is widely accepted and recommended in literature (Kutner *et al.,* 2005).

For the multiple regression model with p predictors, X_1,i=1,…,p VIFs are the diagonal elements r^{ii} of the inverse of the correlation matrix R_{pxp} of the p predictors. The VIF for the ith predictor variable can be expressed by:

$$VIF_i = r^{ii} = \frac{1}{1-R_i^2}, i = 1, \ldots, p,$$

where R_i^2 is the multiple correlation coefficient of the regression between *xi* and the remaining *p-1* predictors.

Tolerance is just the inverse of the VIF which is expressed as,

$$Tolerance = \frac{1}{VIF_I}$$

The Variance Inflation Factor (VIF) is a widely accepted method for detecting multicollinearity in statistical analysis. VIF measures the extent to which the variance of estimated regression coefficients is inflated when the predictor variables are linearly related. It is important to note that VIF and tolerance are reciprocals of each other. A tolerance of 0.20 corresponds to the rule of 5, while a tolerance of 0.10 corresponds to the rule of 10. However, analysts often rely on informal rules of thumb applied to VIF. According to these rules, multicollinearity is present if the largest VIF is greater than 10 (some analysts use a more conservative threshold value of 30) or if the mean of all VIF's is considerably larger than 1 (Brien, 2007; Bock, n.d.). Kennedy (1992) notes that harmful collinearity is indicated by VIFi > 10 for standardized data.

3.0 Results and Discussion

In applying the different methods of detecting multicollinearity, Longley's Economic data and Blood Pressure data was used.

A. Longley's Economic Data

For Longley's Economic data, there were 7 economic variables, observed yearly from 1947 to 1962 (n=16). The independent variables are GNP price product (GNP deflator), Gross National Product (GNP), number of unemployed (Unemployed), number of people in the armed forces (Armed Forces), 'noninstitutionalized' population > 14 years of age (Population), and year. The independent variable of this data is the serial number.

Table 1. Longley's economic data

GNP deflator	GNP	Unemployed	Armed Forces	Population	Year	Employed
83.00	234.289	235.6	159.0	107.608	1947	60.323
88.50	259.426	232.5	145.6	108.632	1948	61.122
88.20	258.054	368.2	161.6	109.773	1949	60.171
89.50	284.599	335.1	165.0	110.929	1950	61.187
96.20	328.975	209.9	309.9	112.075	1951	63.221
98.10	346.999	193.2	359.4	113.270	1952	63.639
99.00	365.385	187.0	354.7	115.094	1953	64.989
100.00	363.112	357.8	335.0	116.219	1954	63.761
101.20	397.469	290.4	304.8	117.388	1955	66.019
104.60	419.180	282.2	285.7	118.734	1956	67.857
108.40	112.769	293.6	279.8	120.445	1957	68.169
110.80	444.546	468.1	263.7	121.950	1958	66.513
112.60	482.704	381.3	255.2	123.366	1959	68.655
114.20	502.601	393.1	251.4	125.368	1960	69.564
115.70	518.173	480.6	257.2	127.852	1961	69.331
116.90	554.894	400.7	282.7	130.081	1962	70.551

Min-Max and Point-coordinates Approach

To apply the procedures using Min-Max and Point-coordinates Approach, one must follow the 5 steps – algorithm that we introduced. The procedure are as follows:

Step 1: Rearrange the observations in ascending order.

Step 2: Attach serial numbers to the observations. and let y = Serial Number, x_1 = GNP deflator, x_2 = GNP, x_3 = Unemployed,

x_4 = Armed Forces, x_5 = Population, and x_6 = Year.

Table 2. With serial numbers attached

y	x_1	x_2	x_3	x_4	x_5	x_6
10	83.00	234.289	187.0	145.6	107.608	1947
20	88.50	259.426	193.2	159.0	108.632	1948
30	88.20	258.054	209.3	161.6	109.773	1949
40	89.50	284.599	232.5	165.0	110.929	1950
50	96.20	328.975	235.6	251.4	112.075	1951
60	98.10	346.999	282.2	255.2	113.270	1952
70	99.00	365.385	290.4	257.2	115.094	1953
80	100.00	363.112	293.6	263.7	116.219	1954
90	101.20	397.469	335.1	279.8	117.388	1955
100	104.60	419.180	357.8	282.7	118.734	1956
110	108.40	112.769	368.2	285.7	120.445	1957
120	110.80	444.546	381.3	304.8	121.950	1958
130	112.60	482.704	393.1	309.9	123.366	1959
140	114.20	502.601	335.0	335.0	125.368	1960
150	115.70	518.173	354.7	354.7	127.852	1961
160	116.90	554.894	359.4	359.4	130.081	1962

Now from the original data arrangement, the table above is now rearranged according to increasing/ascending order and serial numbers are attached to the observations.

Step 3: Locate the minimum and maximum observations for variables.

Based on the updated data arrangement shown from the table in Step 2, the minimum and maximum values are as follows:

Table 3. Minimum and maximum values

	Minimum Values	Maximum Values
Serial Number (y)	1	16
GNP Deflator (x_1)	83.00	116.90
GNP (x_2)	234.289	559.894
Unemployed (x_3)	187.0	480.3
Armed Forces (x_4)	145.6	359.4
Population (x_5)	107.608	130.081
Year (x_6)	1947	1962

Step 4: Plot scatter diagram of least serial and maximum serial numbers against min-max of variables.

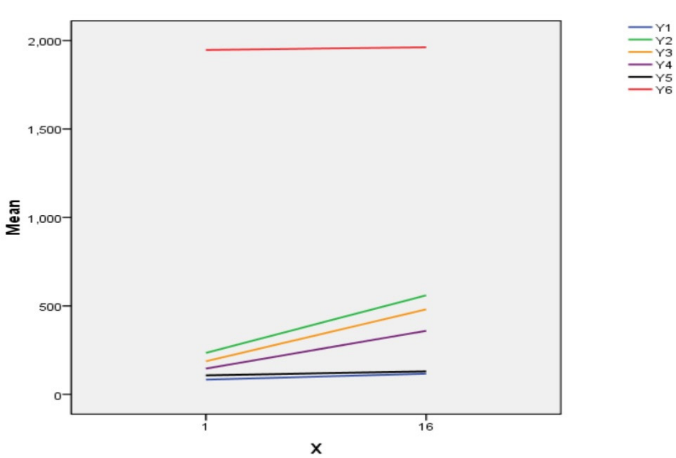

Figure 1. Scatter Plot

The scatter plot of the Point-Coordinates shows that there are lines drawn that are not parallel to one other which implies that there is a significant relationship among the variables.

Step 5: Prove that the lines are parallel using coordinates of points.

$$Gradient\ (y^1) = \frac{Range\ (x^1)}{Range\ (y)}$$

$$= \frac{116.90 - 83.00}{16-1} = \frac{33.90}{15} = 2.69$$

$$Gradient\ (y^2) = \frac{Range\ (x^2)}{Range\ (y)}$$

$$= \frac{559.894 - 234.289}{16-1} = \frac{320.605}{15} = 21.3737$$

$$Gradient\ (y^3) = \frac{Range\ (x^3)}{Range\ (y)}$$

$$= \frac{480.6 - 187.0}{16-1} = \frac{293.60}{15} = 19.573$$

$$Gradient\ (y^4) = \frac{Range\ (x^4)}{Range\ (y)}$$

$$= \frac{359.4 - 145.6}{16-1} = \frac{231.80}{15} = 14.253$$

$$Gradient\ (y^5) = \frac{Range\ (x^5)}{Range\ (y)}$$

$$= \frac{130.081 - 107.608}{16-1} = \frac{22.473}{15} = 1.4982$$

$$Gradient\ (y^6) = \frac{Range\ (x^6)}{Range\ (y)}$$

$$= \frac{1962 - 1947}{16-1} = \frac{15}{15} = 1.0$$

Table 4. Gradients Values

Variables	Gradient
GNP Deflator (x_1)	0.226
GNP (x_2)	0.213737
Unemployed (x_3)	1.957
Armed Forces (x_4)	1.425
Population (x_5)	0.14982
Year (x_6)	0.10

The result shows the values of the gradients of the independent variables of interest. By comparing the values of their gradients, we can observe that all of the gradients are not equal which implies that the lines are not parallel. Since all the gradient values are not equal and the lines are not parallel, the variables are significantly related. Thus, multicollinearity exists.

Product Moment Correlation

Table 5 shows the Pearson's *r* product moment correlation among the variables. As depicted in the table the *r* coefficients of (x_1 and x_2), (x_1 and x_5), and (x_1 and x_6) are equal to .992, .979, and .991, respectively, which means that (x_1 and x_2), (x_1 and x_5), and (x_1 and x_6) are strongly correlated at α=0.01. Since their *p*-value are all equal to 0.000 which is less than α=0.01, this implies that there is a significant relationship between the variables (x_1 and x_2), (x_1 and x_5), and (x_1 and x_6). The *r* coefficients of (x_2 and x_5) and (x_2 and x_6) is equal to .991 and .995, respectively, which is also means that (x_2 and x_5) and (x_2 and x_6) are strongly correlated at α=0.01 and also significant. The *r* coefficient of (x_5 and x_6) is .994 which implies that (x_5 and x_6) are strongly correlated and significant (α=0.01). Since all of their *p*-values are equal to 0.000 which is less than α=0.01, therefore, it implies that there is significant relationship between the variables. Thus, multicollinearity exists.

Table 5. A product moment correlation result

Pearson Correlation	Y	X_1	X_2	X_3	X_4	X_5	X_6
Y	1	.971	.984	.502	.457	.960	.971
		.000	.000	.024	.037	.000	.000
X_1	.971	1	.992	.621	.465	.979	.991
	.000		.000	.005	.035	.000	.000
X_2	.984	.992	1	.604	.446	.991	.995
	.000	.000		.007	.042	.000	.002
X_3	.502	.621	.604	1	-.117	.687	.668
	.024	.005	.007		.225	.002	.024
X_4	.457	.465	.446	-.117	1	.364	.417
	.037	.035	.042	.225		.083	.054
X_5	.960	.979	.991	.687	.364	1	.994
	.000	.000	.000	.002	.083		.000
X_6	.971	.991	.995	.688	.417	.994	1

Eigenvalue

The result shows the eigenvalues and condition indices of the independent variables of interest. According to the result, x_4 has an eigenvalue equal to .000 with condition index of 230.424, x_5 has an eigenvalue equal to 6.246^{-6} with condition index of 1048.080, and x_6 has an eigenvalue equal to 3.664^{-9} with condition index of 43275.043. Since their eigenvalues fall near to 0 and their condition indices are all greater than 15 according to the rule of thumb, then, we can say that there are strongly correlated variables and thus, multicollinearity exists.

Table 6. Eigenvalues and condition numbers

	Eigenvalues	Condition Index
X_1	.082	9.142
X_2	.046	12.256
X_3	.011	25.337
X_4	.000	230.424
X_5	.000006246	1048.080
X_6	.000000003664	43275.043

Variance Inflation Factor

The result in Table 7 shows the VIF and tolerance of the independent variables in detecting multicollinearity. According to the rule of thumb, if VIF value is greater than 10 then, it implies a high correlation, and that multicollinearity exists. From the results below, the VIF values of x_1=135.532,x_2=1 788.513,x_3=33.619,x_5=399.151, and x_6=758.981. It also shows that their VIF values are all greater than 10, which implies that there are highly correlated variables. Therefore, we conclude that multicollinearity exists.

Table 7. Variance inflation factor result

	Tolerance	VIF
X_1	.007	135.532
X_2	.001	1788.513
X_3	.030	33.619
X_4	.279	3.589
X_5	.003	399.151
X_6	.001	758.981

Min-Max and Point-Coordinates approach and the existing methods namely, the Product Moment Correlation, Eigenvalues, and Variance Inflation Factor show similar result that there are highly and significantly related independent variables and that multicollinearity in the Longley's Economic Data exist.

B. Blood Pressure Data

For the Blood Pressure Data, there were 20 individuals with high blood pressure and the variables are blood pressure *(y = BP, in mm Hg)*, age *(x_1=Age, in years)*, body surface area (x_2=BSA, in sq m), duration of hypertension (x_3 = Dur, in years), and basal pulse (x_4 = Pulse, in beats per minute).

Table 8. Blood pressure data

BP	Age	BSA	Dur	Pulse
105	47	1.75	6.1	64
115	49	2.10	5.2	70
116	49	1.98	8.2	72
117	50	2.01	5.8	73
112	51	1.89	7.0	72
121	48	2.25	9.3	71
121	49	2.25	5.2	69
110	47	1.90	6.2	66
110	49	1.83	7.1	69
114	48	2.07	5.6	64
114	47	2.07	5.3	74
115	49	1.98	5.6	71
114	50	2.05	10.2	68
106	45	1.92	5.6	67
125	52	2.19	10.0	73
114	46	1.98	7.4	69
106	46	1.87	6.3	65
113	46	1.90	8.6	70
110	48	1.88	9.0	71
122	56	2.09	7.0	74

Min-Max and Point-Coordinates approach

Step 1: Rearrange the observations in ascending order.

Step 2: Attach serial numbers to the observations and let:

y = serial number; x_1 = Age; x_2 = BSA; x_3 = Dur; x_4 = Pulse

Table 9. With serial number attached

y	x1	x2	x3	x4
1	46	1.75	5.2	64
2	46	1.83	5.2	64
3	46	1.87	5.3	65
4	47	1.88	5.6	66
5	47	1.89	5.6	67
6	47	1.90	5.8	68
7	48	1.90	6.1	69
8	48	1.92	6.2	69
9	48	1.98	6.3	69
10	49	1.98	7.0	70
11	49	1.98	7.0	70
12	49	2.01	7.0	71
13	49	2.05	7.1	71
14	49	2.07	7.4	71
15	50	2.07	8.2	72

ISSN: 2244-6990 (Print) 2408-3577 (Online)

Table 9. *cont.* With serial number attached

y	x1	x2	x3	x4
16	50	2.09	8.6	72
17	51	2.10	9.0	73
18	52	2.19	9.3	73
19	53	2.25	10.0	74
20	56	2.25	10.2	74

The table above is now rearranged according to increasing order and serial numbers are attached to the observations.

Step 3: Locate the minimum and maximum observations for variables.

Table 10. Minimum and maximum values

	Minimum	Maximum
y	1	20
x_1	46.0	56.0
x_2	1.75	2.25
x_3	5.20	10.20
x_4	64.0	74.0

Step 4: Plot scatter diagram of least serial and maximum serial numbers against min-max of variables.

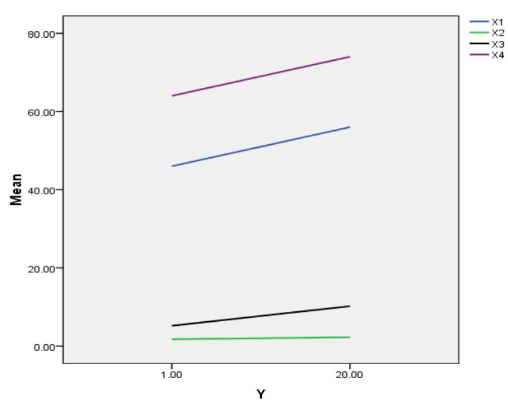

Figure 2. Scatter Plot diagram.

The scatter plot of the Point-Coordinates shows that the lines drawn do not meet which implies that there is no significant relationship among the variables.

Step 5: Prove that the lines are parallel using coordinates of points.

$$Gradient\ (y1) = \frac{Range\ (x1)}{Range\ (y)}$$

$$= \frac{56-46}{20-1} = \frac{10}{19} = 0.5263$$

$$Gradient\ (y2) = \frac{Range\ (x2)}{Range\ (y)}$$

$$= \frac{2.25-1.75}{20-1} = \frac{.50}{19} = 0.0263$$

$$Gradient\ (y3) = \frac{Range\ (x3)}{Range\ (y)}$$

$$= \frac{10.2-5.2}{20-1} = \frac{5}{19} = 0.2632$$

$$Gradient\ (y4) = \frac{Range\ (x4)}{Range\ (y)}$$

$$= \frac{74-64}{20-1} = \frac{10}{19} = 0.5263$$

Table 11. Gradient Values

Variables	*Gradient*
Age (x1)	0.5263
BSA (x2)	0.0263
Dur (x3)	0.2632
Pulse (x4)	0.5263

The table shows the result of gradient of the variables. It shows that x1 and x4 have a common gradient which is equal to 0.5263 implies that their lines are parallel. For x2 and x4, though they have different gradients, their lines drawn which, if extended, do not meet the lines of other variables. Thus, multicollinearity does not exist.

Product Moment Correlation

Table 12. Product moment correlation result

Pearson Correlation	Y	X_1	X_2	X_3	X_4
Y	1.00	.659	.866	.296	.646
		.001	.000	.102	.001
X_1	.656	1.00	.378	.216	.569
	.001		0.50	.180	.004
X_2	.866	.378	1.00	.125	.404
	.000	0.50		.299	.039
X_3	.296	.216	.125	1.00	.243
	.102	.180	.299		.151
X_4	.646	.569	.404	.243	1.00

The table above shows the Pearson Correlation coefficients and significant values. The result shows that x_1 and x_4 has highest value of r coefficient equal to .569 and a significant value of .004 which indicates a moderate correlation. The remaining pair of variables shows us significantly low correlation. Since there is no highly and significantly correlated variables, therefore, it implies that multicollinearity does not exist.

Eigenvalue

Table 13 shows the eigenvalue and condition index results of the variables. As we can see, all of the independent variables have eigenvalues that fall near to zero and condition index greater than 15 which implies that multicollinearity exists. Even though eigenvalue shows different result we have another test to check the multicollinearity.

Table 13. Eigenvalue and Condition Index

	Eigenvalue	**Condition Index**
x1	.038	11.434
x2	.003	43.101
x3	.001	63.226
x4	.001	76.143

Variance Inflation Factor

The table shows the VIF and tolerance results. It shows that the VIF values of all independent variables x_1 = 1.553,x_2= 1.244,x_3= 1.073,and x_4= 1.608 are less than 10 and their tolerance values .644, .804, .932 and .622 respectively are all greater than 0.10 (corresponds to a rule of 10) which indicates that multicollinearity does not exists.

Table 14. VIF and Tolerance Result

	Tolerance	VIF
x1	.644	1.553
x2	.804	1.244
x3	.932	1.073
x4	.622	1.608

The computation of Min-Max and Point-Coordinates Approach shows us that multicollinearity does not exist in the data. In the computation of the existing method, Eigenvalue gives us a different result compared to Product Moment Correlation and Variance Inflation Factor which shows us that multicollinearity does not exist. Since the majority of the methods show similar result, it is enough evidence that multicollinearity does not exist in the Blood Pressure Data.

4.0 Conclusion

The application on detecting the presence of multicollinearity using the Longley's Economic Data, Min-Max and Point-Coordinates approach, by performing 5-algorithm gives us the result that there exists highly, and significantly related independent variables based on the parallelism principle and that there is the presence of multicollinearity. The existing methods such as Product Moment Correlation, Eigenvalues, and Variance Inflation Factor share similar results in detecting the presence of multicollinearity where multicollinearity exists. We also used the Blood Pressure Data where multicollinearity does not exist. In this case, Min-Max and Point-Coordinates Approach, Product Moment Correlation and Variance Inflation Factor give us the same result where multicollinearity does not exist in the data. Therefore, Min-Max and Point-Coordinates Approach is suitable in detecting the presence of multicollinearity.

Based on the findings, it is recommended that researchers consider using the Min-Max and Point-Coordinates Approach in detecting multicollinearity. This approach was found to be user-friendly, requires less mathematical rigor, and provides comparable results to existing methods. The limitations of the study and recommendations for future research provide opportunities for further exploration and advancement in this area.

References

Abidemi, A. K., Bright, A. F., & Uzoma, U. E. (2016). Detection of multicollinearity using min-max and point-coordinates approach. *American Journal of Theoretical and Applied Statistics, 4(6),* 640-643. doi: 10.11648/j.ajtas.20150406.36.

Belsley, D. (1991). *Conditioning diagnostics: Collinearity and weak data in regression.* New York: Wiley. ISBN 0-471-52889-7

Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity, 41(5),* 673. doi: 10.1007/s11135-006-9018-6.

Bock, T. (n.d.). *Display R. variance inflation factor.* https://www.displayr.com/variance-inflation-factors-vifs/

Chen, P. Y., & Popovich, P. M. (2002). C*orrelation: Parametric and nonparametric measures.* Thousand Oaks, CA: Sage Publications.

Draper, N. R., & Smith, H. (2003). *Applied regression analysis* (3rd ed.). Wiley, New York.

Duncombe, P., Kenward, M., Kollerstorm, J., & Wetherill, G. B. (1986) *Regression analysis with application* (1st ed.). Chapman and Hall, New York.

Enders, W. (2010). *Applied econometric time series.* John Wiley & Sons.

Gujrati, D. N. (2004). *Basic econometrics.* 3rd edition, Tata McGraw-Hill, New Delhi.

Jim F. (2013). *What are the effects of multicollinearity and when can I ignore them?* http://blog.minitab.com/blog/adventures-in-statistics/what-are-the-effects-of-multicollinearity-and-when-can-i-ignore-them.

Kennedy, P. E. (2002). More on venn diagrams for Regression. *Journal of Statistics Education, 10(1).* www.amstat.org/publications/jse/v10n1/kennedy.html.

Kock, N., & Lynn, G. S. (2012). Lateral collinearity and misleading results invariance-based SEM: An illustration and recommendation. J*ournal of the Association for Information Systems, 13(7),* 546-580.

Kumar, T. K. (1975). Multicollinearity in regression analysis. *Review of Economics and Statistics, 57(3),* 365-366.

Kutner, M. H., Li, W., Nachtsheim, C. J., & Neter, J. (2005). A*pplied linear statistical models* (5th ed.). Irwin McGraw-Hill, New York. ISBN 0-07-238688-6.

Longley, J. W. (1967). An appraisal of least-squares programs from the point of view of the user. J*ournal of the American Statistical Association, 62,* 819-841.

Math Centre. (2009). *Equations of straight lines.* www.mathcentre.ac.uk/resources/uploaded/mc-ty-strtlines-2009-1.pdf.

Math is Fun. (n.d.). *Gradient (slope) of a straight line.* https://www.mathsisfun.com/gradient.html.

Protter, M. H., & Protter, P. E. (1988). *Protter and Protter: Calculus with analytic geometry* (4th ed.).

R Core Team. (2018). *Datasets distributed with R.* https://cran.r-project.org/doc/manuals/r-release/R-data.html.

Statistics How To. (2018). *Range of Set of data in math and statistics.* http://www.statisticshowto.com/probability-and-statistics/statistics-definitions/range-statistics/.

Statistics Solutions. (2015). *Correlation (Pearson, Kendall, Spearman).* http://www.statisticssolutions.com/correlation-pearson-kendall-spearman/.

Strang, G. (2010). I*ntroduction to linear algebra* (4th edition). Wellesley-Cambridge Press.