ISBN: 978-93-90847-07-5

13. Multicollinearity and Multivariate Linear Regression: The Consequence, Diagnostics and Remedies

Dipen Roy

Professor of Commerce,
University of North Bengal,
India.

13.1 Introduction:

In the socio-economic perspective, almost every variable is subject to influence of multiple other variables. For example, simply weight of a group high-school students, studying in a class, does not depend on age alone. Besides age, it depends on heredity, height, family income, nutrition, physical exercise, etc. Likewise, the relationship of a macroeconomic variable with its predictors is not a simple bi-variate relation. Contrary to this, many of the economic relationships are multi-variate relationships. For example, GDP is a function of multiple explanatory variables such as size of FDI (X_1) inflow, domestic capital formation (X_2), foreign trade (X_3), governmental expenditure (X_4), foreign aids (X_5) and many other variables. Thus, a researcher begins his research with a hypothetical relationship, expressed as below:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + e_t$$

Where Y is the dependent variable and X_is are multiple predictors (explanatory variables) explaining increase of Y. The coefficient β_i is the sensitivity of the variable i. This initiates the beginning of a multivariate analysis.

The relationship stated above stands valid only when all the assumptions of OLS are duly satisfied. One of the assumptions of OLS requires that the explanatory variables X_1, X_2, X_3, X_4, etc., should not be linearly correlated with each other. If the explanatory variables X_1, X_2, X_3, X_4, etc., appear linearly correlated with each other, the phenomenon is defined as multicollinearity, which takes a researcher to a paradoxical situation [Zikmund, W.G., et al (2016)]. In this paper we propose to unfold the paradoxical situations confronted by a researcher with simple numerical examples. This is supposed to enable the researchers to take appropriate precautions before drawing the final inference from a model.

In short, the explanatory variables should be orthogonal, meaning not related with each other. However, in practice, many of the explanatory variables (predictors) move together in an inter-related fashion. Agricultural production, industrial production, consumption, income tax collection and many other macro-variables move together, at least, in terms of direction.

In other words, when there is up (boom), there is up; when there is down (recession), there is down. This makes many macro-variables become inter-correlated with each other. This is the root cause behind the multicollinearity of time-series data.

Multi-collinearity is not necessarily limited to time-series data. It can occur in cross-section data, as well, when the explanatory variables have inherent causal relationship.

13.2 Consequences of Multicollinearity:

While analyzing multivariate data, researchers focus lies on the determination of the estimates of the coefficients β_1 β_2, β_3, etc.

Theoretically, the presence of perfect linear relationship between the explanatory variable makes the parameters β_1 β_2, β_3 etc undefined. Using a simple hypothetical relation, consisting of two explanatory variables X_1 and X_2, we show how the coefficients become undefined.

Say the hypothetical relationship is written as below:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

The formula of β_1 is

$$\beta_1 = \frac{(\sum x_1 y)(\sum x_2^2) - (\sum x_2 y)(\sum x_1 x_2)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$$

We propose to examine how β_1 becomes undefined. We presume that there is a linear relationship between X_1 and X_2. The said relationship is represented by $X_2 = kX_1$. While writing kx_1 for x_2 in above equation, the formula gets transformed as below:

$$\beta_1 = \frac{k^2(\sum x_1 y)(\sum x_1^2) - k^2(\sum x_1 y)(\sum x_1 x_1)}{k^2(\sum x_1^2)(\sum x_1^2) - k^2(\sum x_1 x_1)2}$$

$$\beta_1 = \frac{k^2(\sum x_1 y)(\sum x_1^2) - k^2(\sum x_1 y)(\sum x_1^2)}{k^2(\sum x_1^2)^2 - k^2(\sum x_1^2)^2} = \frac{0}{0}$$

That is, presence of multi-collinearity renders the coefficients β_1 undefined. This is true in case of β_2 as well. It makes multivariate analysis becomes meaningless. However, the escape is that, in reality, the correlation coefficient between two explanatory variables is neither perfect ($r = 1$), nor zero.

It lies between the two, i.e., $0 \le r \le 1$. As the degree of correlation between explanatory variables increases the regression coefficients continues to become unstable.

In the words of Field, Andy (2016) presence of multicollinearity makes the coefficients almost unreliable. In the words of Koutsoyiannis, A. (1996) the coefficients become unstable, as sample size is increased or more correlated variables are incorporated into the model.

Field Andy (2016) observes the consequences of multicollinearity on three different aspects of multivariate analysis. These are value of the coefficients β_1 β_2, β_3 etc., measure of goodness of fit R^2 and sensitivity of the predictors. Firstly, the author points that the coefficients β_1 β_2, β_3 etc. appear unreliable. Secondly, the goodness of fit, R^2 remains unaffected even when the extra predictors are added to the model or withdrawn from the model. These two points have been explained with a case of numerical example in the following paragraphs.

13.3 Advertisement and R&D Expenditure: A Case of Multicollinearity

Consider the data on 'Sales Revenue, R&D and Advertisement' given in the table below (**Table 1**). We begin with a belief that advertisement and R&D add to the sales revenue of the firm. Therefore, advertisement and R&D are predictors, while sales revenue is the dependent variable. For computing the regression coefficients, it requires us to assume that there is no relationship between R&D and Advertisement. However, we have taken a data-set, where the predictors, advertisement and R&D, are linearly correlated. The purpose of this example is to bring to light the consequences of multicollinearity.

Table 13.1: Measuring the Impact of R&D and Advertisement on Sales Revenue

	Figures are in $ billion									
Advertisement	10	12	11	10	13	15	19	21	22	26
R&D Outlay	3	4	3	3	4	5	6	7	7	8
Sales Revenue	45	50	47	54	50	59	62	65	66	73

We begin with estimating the relationship of sales with individual predictors, advertisement and R&D, separately. When multicollinearity is fully absent, the coefficient of a predictor obtained from simple regression becomes almost equal to the coefficient of the predictor obtained from a multivariate regression. However, presence of multicollinearity gives a misleading result. We propose to verify this from the analysis of subsequent paragraphs. Definitely, our calculation is based on the data given above in **Table 1**.

Table 13.2 Shows the Coefficient Table of Simple Relationship of Sales with Advertisement.

Table 13.2: Coefficients[a]						
Model		**Unstandardized Coefficients**		**Standardized Coefficients**	**t**	**Sig.**
		B	**Std. Error**	**Beta**		
1	(Constant)	32.341	2.917		11.088	.000
	ADV	1.557	.174	.954	8.967	.000
a. Dependent Variable: SALES						

The coefficient of advertisement is statistically significant. It means that advertisement has strong influence on sales. Goodness of fit of the model is 91%. In short, the equation of sales on advertisement can be written as below

$$\text{SALES} = 32.34 + 1.557(\text{ADV}) \qquad R^2 = 0.91$$
$$\textbf{\textit{t}}\text{ - Statistic} \quad (11.08) \quad (8.96)$$

While computing simple regression equation of sales on R&D, we find the following output as shown in Table 13.3

Table 13.3: Coefficients[a]						
Model		**Unstandardized Coefficients**		**Standardized Coefficients**	**t**	**Sig.**
		B	**Std. Error**	**Beta**		
1	(Constant)	33.506	2.744		12.209	.000
	R&D	4.719	.517	.955	9.131	.000
a. Dependent Variable: SALES						

The result as shown in **Table 13.3** reflects that the coefficient of R&D is statistically significant. In other words, there is strong influence R&D on sales. Goodness of fit is 91%. In short, the equation of sales on R&D can be written as:

$$\text{SALES} = 33.50 + 4.719(\text{R\&D}) \qquad R^2 = 0.91$$
$$\textbf{\textit{t}}\text{ - Statistic} \quad (12.21) \quad (9.13)$$

In practice, it is found that every success of R&D requires the company to make an additional advertisement to address the audience about the new features of the product.

So, a higher the level of expenditure on R&D, higher is the level of advertisement expenditure.

That is, two explanatory variables are linearly correlated with each other. Actual measure of correlation between two predictors, advertisement and R&D is 99%.

Now we propose to incorporate both the predictors in a single model and study the consequences of multicollinearity. On the line of the relationship such as $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$, we initially write that

$$\text{SALES} = \beta_0 + \beta_1 (\text{ADV}) + \beta_2 (\text{R\&D})$$

The SPSS output of multiple regression is briefly stated as below:

$$\text{SALES} = 33.50 + 0.63\text{ADV} + 2.82(\text{R\&D}) \qquad R^2 = 0.91$$
$$\textbf{\textit{t}}\text{ - Statistic} \quad (10.45) \quad (0.45) \quad (0.67)$$

Table 13.4 shows the ANOVA table and Table 5 presents the table of the coefficient. These are obtained as elements of SPSS output.

Table 13.4: ANOVA[a]

Model		**Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**
1	Regression	714.492	2	357.246	37.657	.000[b]
	Residual	66.408	7	9.487		
	Total	780.900	9			

a. Dependent Variable: SALES

b. Predictors: (Constant), ADV, R&D

Table 13.5: Coefficients[a]

Model		**Unstandardized Coefficients**		**Standardized Coefficients**	**t**	**Sig.**
		B	**Std. Error**	**Beta**		
1	(Constant)	32.937	3.151		10.453	.000
	ADV	.633	1.391	.387	.455	.663
	R&D	2.821	4.210	.571	.670	.524

a. Dependent Variable: SALES

The findings (Table 13.4 and Table 13.5) are really puzzling, because the regression is statistically significant, because ANOVA table indicates that F is statistically significant (much better than 1% level); however, the coefficients of the predictors, advertisement and R&D, are statistically insignificant. Now on the basis of above three outputs, we can summarize the observations as below:

a) Firstly, the coefficients of the predictors, advertisement and R&D, obtained from multivariate regression are much lower (See Table 13.5) compared to the corresponding values obtained from simple regression analysis.
b) The coefficients of the predictors in the coefficient table of multiple regression (see Table 13.5) are statistically insignificant, while the corresponding coefficients were significant in the outputs of simple regression, where we assessed the effect of individual predictor on sales separately.
c) Thirdly, goodness of fit, R^2 is same in all three outputs. It shows that if there is perfect multicollinearity, addition of linearly correlated variables does not improve the goodness of fit.
d) Standard Errors of the coefficients obtained in multiple regression analysis are several times higher than their original Standard Errors. This point is discussed in greater details in the following paragraph.

13.4 Multicollinearity and Standard Error of Coefficients:

Theoretically, presence of multicollinearity magnifies the size of Standard Errors of the coefficients. The same can be proved as below:

$$\text{Var}(\beta_1) = \frac{(\sum x_2^2)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$$

$$= \frac{k^2 \sum x_1^2}{k^2 (\sum x_1^2)^2 - k^2 (\sum x_1^2)^2} = \frac{k^2 \sum x_1^2}{0} = \infty$$

Standard Error (β_1) = $\sqrt{Var(\beta_1)} = \infty$

Table 13.6: Multi-collinearity and the Stand Error of the Coefficients:

Simple Regression		**Multiple Regression**	
Predictor	Standard Error(β_i)	Predictor	Standard Error(β_i)
Advertisement	0.174	Advertisement	1.39
R&D	0.517	R&D	4.21

Table 13.6 reflects that the Standard Errors of the coefficients shown under multiple regression column are several times higher than the corresponding standard errors obtained in simple regression analysis. In an ideal situation, while there would exists no multicollinearity, the coefficients of the predictors and their corresponding Standard Errors, as obtained from simple regression analysis would remain identical with their corresponding values computed from multiple regression model. In other words, in the absence of multicollinearity the results of simple regression analysis and those of the multiple regression analysis would be identical. Unfortunately, this did not happen in our analysis. Existence of Multicollinearity results in exaggerated value of Standard Error of the regression coefficients, which leads to abnormally lower value of the computed t-statistic. This leads the researcher to infer that regression coefficients are statistically insignificant, even though the goodness of fit of the regression is quite satisfactory (*accepted through F-Test*).

Table 13.7: Collinearity and Goodness of Fit

Influence of	**Regression Equation**	**Goodness of Fit**
ADV on SALES	SALES = 32.34 + 1.557(ADV)	$R^2 = 0.91$
R&D on SALES	SALES = 33.50 + 4.719(R&D)	$R^2 = 0.91$
ADV and R&D on SALES	SALES = 33.50 + 0.63ADV+ 2.82(R&D)	$R^2 = 0.91$

Table 13.7 compiles R^2 values obtained in different trials of assessing the effects of Advertisement and R&D on sales. Surprisingly, R^2, the goodness fit, remains unaffected ($R^2 = 0.91$) in all three trials. It means multicollinearity does not reduce or increase the goodness of fit. Addition or withdrawal of a similarly correlated variable leaves no effect on goodness of fit.

13.5 The Gravest Consequence of Multicollinearity:

Now we propose to take another example with time-series data. The table (**Table 13.8**) contains data relating to five variables. To note, here X_1 is dependent variable and remaining other variables X_2, X_3, X_4 and X_5 are predictors. This example will draw attention to the gravest consequence of multicollinearity.

Table 13.8: Linearly Correlated Predictors: Time-series Data

Year	X_1	X_2	X_3	X_4	X5
1	23.31	15.05	30.38	25.21	18.88
2	32.98	16.24	27.15	22.6	16.71
3	10.37	21.92	31.48	11.66	17.07
4	48.48	32.69	39.99	25.55	10.69
5	20.17	37.02	45.19	34.68	29.68
6	-17.6	17.48	23.59	22.27	22.14
7	6.11	10.31	11.72	1.27	7.79
8	19.06	8.29	10.36	3.24	7.26
9	18.65	6	6.38	0.12	2.83
10	1.87	9.93	10.36	12.61	6.41
11	-6.58	6.85	6.36	10.4	6.5
12	3.36	2.58	1.53	3.96	-2.31
13	1.07	6.87	7.33	13.87	10.46
14	36.51	14.75	18.4	20.93	10.08
15	50.98	18.56	24.27	23.05	19.34
16	41.36	25.12	30.96	18.5	22.46
17	35.37	26.73	32.66	35.2	32.39
18	10.46	27.82	31.64	21.96	30.57
19	27.1	17.38	19.92	7.67	14.83
20	11.26	18.95	19.94	15.25	20.09
21	18.17	14.14	14.47	16.34	22.3
22	6.45	10.07	10.16	1.36	18.11
23	12.79	7.92	6.1	3.67	14.24

Year	X_1	X_2	X_3	X_4	X5
24	-3.25	8.08	7.58	7.5	7.47
25	13.77	6.73	5.92	2.41	5.3
26	-3.09	12.12	11.21	5.23	-2.03
27	-4.91	8.86	8.97	15.65	7.4
28	6.96	6.08	5.14	1.28	7.34
29	12.48	8.39	7.79	21.74	11.66

One of the methods of detecting multicollinearity is scanning the correlation matrix. Looking at the correlation matrix [See Table 13.9], it is observed that the variables X_2, X_3, X_4 and X_5 are strongly correlated with each other. Hence, this analysis is undertaken with the advance information that there is multicollinearity. Our objective is to point to the worst consequence that multicollinearity can produce. We adopt the approach called management by exception, i.e., learning from the mistakes.

Table 13.9: Correlation Matrix

		X_1	X_2	X_3	X_4	X_5
X_1	Pearson Correlation	1	.521**	.556**	.462**	.349*
	Sig. (1-tailed)		.002	.001	.006	.032
	N	29	29	29	29	29
X_2	Pearson Correlation	.521**	1	.962**	.771**	.749**
	Sig. (1-tailed)	.002		.000	.000	.000
	N	29	29	29	29	29
X_3	Pearson Correlation	.556**	.962**	1	.805**	.742**
	Sig. (1-tailed)	.001	.000		.000	.000
	N	29	29	29	29	29
X_4	Pearson Correlation	.462**	.771**	.805**	1	.723**
	Sig. (1-tailed)	.006	.000	.000		.000
	N	29	29	29	29	29
X_5	Pearson Correlation	.349*	.749**	.742**	.723**	1
	Sig. (1-tailed)	.032	.000	.000	.000	
	N	29	29	29	29	29

We want to begin with a simple regression equation of two variables, one predicted and the other a predictor. We take X_2 is a predictor of X_1. Table 10 shows the coefficient table obtained from regression output of X_1 on X_2.

Based on the entries of Table 13.10, we construct the following equation.

$$X_1 = 0.178 + 1.013\, X_2 \qquad R^2 = 0.27$$

t - Statistic (.033) (3.173)

Table 13.10: Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.178	5.394		.033	.974
	X2	1.013	.319	.521	3.173	.004

a. Dependent Variable: X1

The result shows that X_2 has a positive influence on the value of X_1. As per indications of ANOVA table (not shown here), the regression is statistically significant (at 1% level).

Subsequently, we add X_3 and X_4 to the model. Table No 11 shows the Coefficient Table. It shows a negative coefficient for X_2.

This coefficient was positive in simple regression output, shown in Table 13.10. But with addition of new variables, it turns negative. This is the gravest consequence of multicollinearity. Table No 12 shows the Coefficient Table, when all four collinear variables are introduced.

Table 13.11: Three Predictors' Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.817	5.740		.316	.754
	X2	-.335	1.176	-.172	-.285	.778
	X3	.977	.921	.689	1.060	.299
	X4	.068	.468	.041	.145	.886

a. Dependent Variable: X1

Table 13.12: Four Predictors' Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.407	5.899		.408	.687
	X2	-.181	1.219	-.093	-.148	.883
	X3	.972	.933	.685	1.041	.308
	X4	.166	.502	.099	.331	.744
	X5	-.302	.505	-.162	-.598	.556

a. Dependent Variable: X1

This is needless to say that in case of each output, in the light of ANOVA Table data, regression appeared significant. However, the paradox is that when we read all three tables [Table 13.10, Table 13.11 and Table 13.12] simultaneously, the coefficient of X_2 changes from 1.013 in Table 10 to - 0.335 in Table 13.11. It again changes to – 0.181 in Table 13.12 in the third trial.

In short, the coefficient of same predictor X_2 is fluctuating from one trial to another trial. Fluctuation is occurring to the coefficients of all other variables too (such as X_3 and X_4).

Koutsoyiannis (1996) points to this fact and states that the values of the coefficients become unstable as additional collinear variables are added to the model.

The very crucial drawback of multicollinearity is that the effect of an individual variable on the predicted variable gets eliminated by the influence of other linearly correlated variables.

As a result, measuring the marginal effect of changing a variable by one unit becomes thoroughly useless.

In other words, presence of multicollinearity makes it very difficult to assess the relative importance of a variable.

13.6 Multicollinearity Diagnostic:

Scholars can construct the correlation matrix to trace the presence of multicollinearity.

The presence strong correlation between the predictors is taken as the indication of the presence of multicollinearity. Second alternative is to look at the Standard Errors of the coefficients.

However, none of these criteria is a satisfactory indicator multicollinearity (Koutsoyiannis, A., 1996).

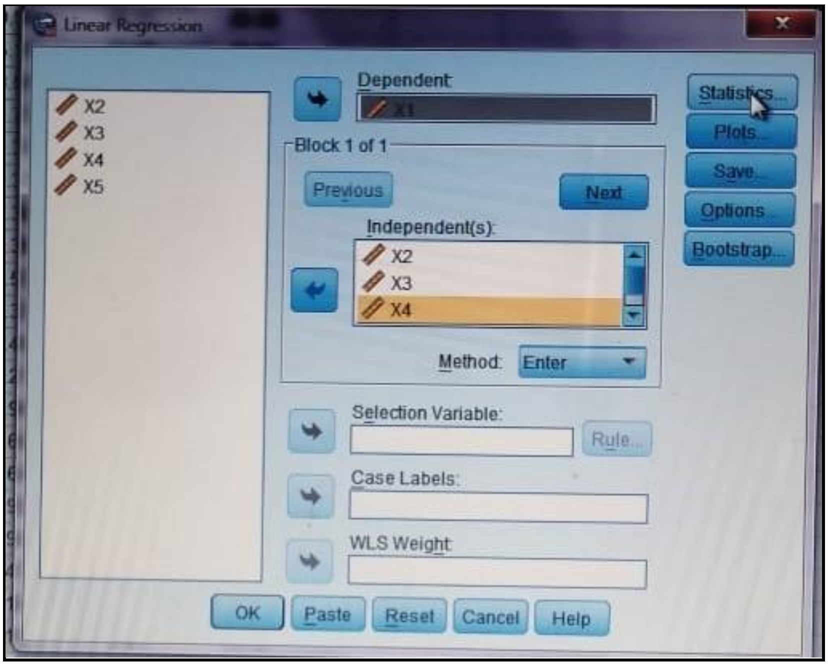
SPSS software provides two multicollinearity diagnostics called Variance Inflation Factor (VIF) and Tolerance; these are quite efficient as well as a globally recognized index for measuring collinearity. This comes to the great help of the researchers today.

The VIF indicates whether a predictor has strong linear relationship with other predictors.

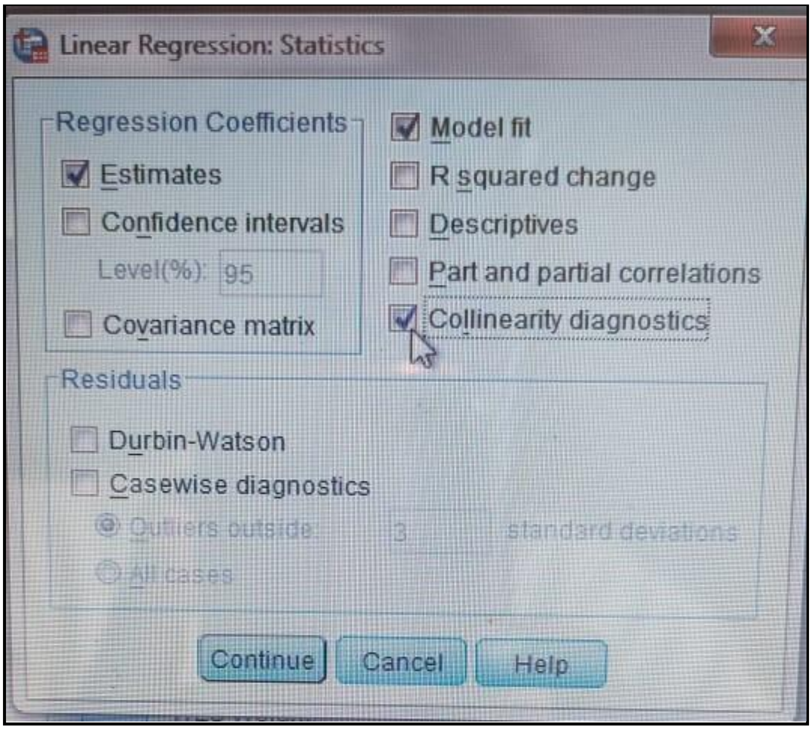
Given the computer is already loaded with SPSS, we demonstrate the steps involved in the process of conducting collinearity test.

While the data-set is ready, click on '**analyze**' located on upper menu bar, choose '**Regression**' from the list of analysis; then chose **'Linea**r' [In short, Analyse →Regression → Linear].

This will open the dialogue box of Linear Regression, a picture of which is given below:



This dialogue box requires a researcher to enter dependent variable and independent variables (predictors) in identified boxes. He is supposed to enter X_1 as dependent variable, remaining other predictors X_2, X_3, X_4 and X_5 as independent variable. Now making a click on OK is enough to get regression output. If the researcher wants to examine multicollinearity, he has to make a click on statistics, shown at the top right corner of the dialogue box above; it opens a dialogue box shown below:



Given the above dialogue box, the researcher is select collinearity diagnostics and run the regression model. This gives the following outcome (see Table 13.13) with full VIF indices corresponding to each coefficient.

Table 13.13: Coefficients [a]							
Model	Unstandardized Coefficients		Standardized Coefficients	T	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
(Constant)	2.504	5.916		.423	.676		
X2	-.198	1.222	-.102	-.162	.873	.072	13.924
X3	1.028	.931	.725	1.104	.281	.066	15.198
X4	.079	.483	.048	.164	.871	.338	2.963
X5	-.273	.504	-.146	-.543	.592	.390	2.561
a. Dependent Variable: X1							

Table 13.13 shows two different indices for measuring multicollinearity. These are VIF and Tolerance. There is no consensus as to which value of VIF can be taken as the index of a serious degree of multicollinearity. Myers (1990) states that VIF equal to 10 is a limit of great concern. When VIF is less than 10, multi-collinearity is not likely to be hazardous. However, Hair J. F. et al (2010) states that VIF above 5 means there is multicollinearity. The rule of thumb is to take VIF equal to 5 as the cut-off point. It means if VIF is less than 5, the problem of multicollinearity is not severe.

Tolerance Level, which is defined as 1/VIF, can also be used to measure severity of multicollinearity.

Menard (1995) suggests that Tolerance Level below 0.2 is a matter of concern. It follows that as a rule of thumb we can accept VIF equal to 5 or Tolerance level 0.20 as the cut-off level. The both refers to the same cut-off level.

We propose to apply VIF equal to 5 as the cut-off limit to the output shown in Table 13. While we look at Table 13, the table of coefficients, we notice that variable X_2 and X_3 have VIF more than 5. While predictors X_3 and X_4 have VIF below 5. It means the predictors X_2 and X_3 (with VIF more than 5) should be dropped from the model, while predictors X_3 anmed X_4 can be retained. Some experts point to the question of relevance. If X_2 and X_3 are predictors of primary importance in the model, as price in a demand function, the problem becomes more difficult to handle. In that case, the researchers have to rely on other methods, rather than depending on Multiple Regression.

13.7 Remedies of Multicollinearity:

If VIF is within the cut-off limit, the limited degree of multicollinearity is not likely to affect the findings of the model. While some of the unimportant predictors have strong multicollinearity, those predictors can be dropped from the model without affecting the spirit of the model.

However, if multicollinearity has serious effects on the coefficients of the predictors of primary importance some solutions are to be explored. There are many methods of handling multicollinearity. Some of the simple and useful remedies have been enlisted below:

a. Increase the sample size: As sample size is increased, standard error of the coefficients continues to reduce; this enhances the values of t-statistic of the coefficients.
b. Principal Component Regression: This process involves transformation of group of linearly correlated predictors into an orthogonal factor, say F_1. Hence from a set of predictors a researcher may find two or three factors, which would be subsequently combined in a revised regression model, i.e., $Y = \alpha + \beta F_1 + \lambda F_2 + \gamma F_3$.
c. Using Additional Equation: Multicollinearity can be avoided by introducing an additional equation to the model by explaining the way predictors are related with each other. Say, $X_1 + 2X_3 + 0.5X_3 = 10$

Besides above, many other remedies are there. Interested readers can check textbook of Econometrics enlisted below.

13.8 References:

1. Bowerman, B. L. & O'Connell R. T. (1990). Linear Statistical Models: An Applied Approach, (2nd Ed) Belmont, CA: Duxbury.
2. Field, Andy (2009) Discovering Statistics using SPSS. Sage. Thousand Oaks, California.
3. Hair J. F., W. C. Black, B. J. Babin and R. Anderson (2010). Multivariate Data Analysis. Upper Saddle River, NJ: Prentice Hall, [VIF above 5 is a problem.].
4. Koutsoyiannis, a (1996). Theory of Econometrics. ELBS Macmillan, Hampshire, UK
5. Menard, S. (1995). Applied Logistic Regression Analysis. Sage University Paper Series on Quantitative Applications in Social Science. Pp.07-106, Thousand Oaks CA: Sage.
6. Myers, R. (1990). Classical and Modern Regression with Applications. (2nd Ed) Boston, MA: Duxbury.
7. Zikmund, W.G., Babin, B. J. Carr, J. C. Adhikari, A. Griffin, M. (2016) Business Research Methods: A South Asian Perspective, Cengage Learning India Private Ltd, Delhi.