UNIT 10 MULTICOLLINEARITY*

Structure

10.0 OBJECTIVES

After going through this unit, you should be able to

- explain the concept of multicollinearity in a regression model;
- comprehend the difference between the near and perfect multicollinearity;
- describe the consequences of multicollinearity;
- [1]explain how multicollinearity can be detected; and
- describe the remedial measures of multicollinearity; and
- explain the concept of ridge regression.

10.1 INTRODUCTION

The classical linear regression model assumes that there is no perfect multicollinearity. Multicollinearity means the presence of high correlation between two or more explanatory variables in a multiple regression model.

* Dr. Pooja Sharma, Assistant Professor, Daulat Ram College, University of Delhi

Absence of multicollinearity implies that there is no exact linear relationship among the explanatory variables. The assumption of no perfect multicollinearity is very crucial to a regression model since the presence of perfect multicollinearity has serious consequences on the regression model. We will discuss about the consequences, detection methods, and remedial measures for multicollinearity in this Unit.

10.2 TYPES OF MULTICOLLINEARITY

Multicollinearity could be of two types: (i) perfect multicollinearity, and (ii) imperfect multicollinearity. Remember that the division is according to the degree or extent of relationship between the explanatory variables. The distinction is made because of the nature of the problem they pose. We describe both types of multicollinearity below.

10.2.1 Perfect Multicollinearity

In the case of perfect multicollinearity, the explanatory variables are perfectly correlated with each other. It implies the coefficient of correlation between the explanatory variables is 1. For instance, suppose want to derive the demand curve for a good Y. We assume that quantity demanded (Y) is a function of price (X_2) and income (X_3). In symbols,

$Y = f(X_2, X_3)$ where X_2 is price of good Y and X_3 is the weekly consumer income.

Let us consider the following regression model (population regression function):

$$Y_i = A_1 + A_2X_{2i} + A_3X_{3i} + u_i \quad \ldots (10.1)$$

In the above equation, suppose

A_2 is < 0. This implies that prices are inversely related do demand.

$A_3 > 0$. This indicates that as income increases, demand for the good increases.

Suppose there is a perfect relationship between X_2 and X_3 such that

$$X_{3i} = 300 - 2X_{2i} \quad \ldots (10.2)$$

In the above case, if we regress X_3 on X_2 we obtain the coefficient of determination $R^2 = 1$.

If we substitute the value of X_3 from equation (10.2), we obtain

$$\begin{aligned} Y_i &= A_1 + A_2X_{2i} + A_3(300 - 2X_{2i}) + u_i \\ &= A_1 + A_2X_{2i} + 300A_3 - 2A_3X_{2i} + u_i \\ &= (A_1 + 300A_3) + (A_2 - 2A_3)X_{2i} + u_i \quad \ldots (10.3) \end{aligned}$$

Let $C_1 = (A_1 + 300A_3)$ and $C_2 = (A_2 - 2A_3)$. Then equation (10.3) can be written as:

$$Y_i = C_1 + C_2X_{2i} + u_i \quad \text{....(10.4)}$$

Thus if we estimate the regression model given at (10.4), we obtain estimators for C_1 and C_2. We do not obtain unique estimators for A_1, A_2 and A_3.

As a result, in the case of perfect linear relationship or perfect multicollinearity among explanatory variables, we cannot obtain unique estimators of all the parameters. Since we cannot obtain their unique estimates, we cannot draw any statistical inferences (hypothesis testing) about them. Thus, in case of perfect multicollinearity, estimation and hypothesis testing of individual regression coefficients in a multiple regression are not possible.

10.2.2 Near or Imperfect Multicollinearity

In the previous section, the presence of perfect multicollinearity indicated that we do not get unique estimators for all the parameters in the model. In practice, we do not encounter perfect multicollinearity. We usually encounter near or very high multicollinearity. In this case the explanatory variables are approximately linearity related.

High collinearity refers to the case of "near" or "imperfect" multicollinearity. Thus, when we refer to the problem of multicollinearity we usually mean "imperfect multicollinearity''

Let us consider the same demand function of good Y. In this case we however assume that there is imperfect multicollinearity between the explanatory variables (in order to distinguish it from the earlier case, we have changed the parameter notations). The following is the population regression function:

$$Y_i = B_1 + B_2X_{2i} + B_3X_{3i} + u_i \quad \text{....(10.5)}$$

Equation (10.5) refers to the case when two or more explanatory variables are not exactly linear. For the above regression model, we may obtain an estimated regression equation as follows:

Equation (10.5):	$\hat{Y}_i = 145.37$	$-$	$2.7975X_{2i}$	$-$ $0.3191X_{3i}$
Standard Error:	(120.06)		(0.8122)	(0.4003)
t-ratio:	(1.2107)		(–3.4444)	(–0.7971)

$R^2 = 0.97778$ ... (10.6)

Since the explanatory variables are not exactly related, we can find estimates for the parameters. In this case, regression can be estimated unlike the first case of prefect multicollinearity. It does not mean that there is no problem with our estimators if there is imperfect multicollinearity. We discuss the consequences of multicollinearity in the next section.

Check Your Progress 1

1) What is meant by perfect multicollinearity?

.....

.....

.....

.....

.....

2) What do you understand by imperfect multicollinearity?

.....

.....

.....

.....

.....

3) Explain why it is not possible to estimate a multiple regression model in the presence of perfect multicollinearity.

.....

.....

.....

.....

.....

10.3 CONSEQUENCES OF MULTICOLLINEARITY

We know from Unit 4 that the ordinary least squares (OLS) estimators are the Best Linear Unbiased Estimators (BLUE). It implies they have the minimum variance in the class of all linear unbiased estimators. In the case of imperfect multicollinearity, the OLS estimators still remain BLUE. Then what is the problem? In the presence of multicollinearity, there is an increase in the variance and standard error of the coefficients. As a result, very few estimators are statistically significant.

Some more consequences of multicollinearity are given below.

(a) The explanatory variables may not be linearly related in the population (i.e., in the population regression function), but they could be related in a particular sample. Thus multicollinearity is a sample problem.

(b) Near or high multicollinearity results in large variances and standard errors of OLS estimators. As a result, it becomes difficult to estimate true value of the estimator.

(c) Multicollinearity results in wider confidence intervals. The standard errors associated with the partial slope coefficients are higher. Therefore, it results in wider confidence intervals.

$$P_r\left[b_2 - t_{\alpha/2}SE(b_2) \leq \beta_2 \leq b_2 + t_{\alpha/2}SE(b_2)\right] = 1 - \alpha \quad \ldots.(10.7)$$

Since the values of standard errors have increased the interval reflected in expression in (10.7) has widened.

(d) Insignificant t ratios: As pointed out above, standard errors of the estimators increase due to multicollinearity. The t-ratio is given as $= \frac{b_2}{SE(b_2)}$. Therefore, the t-ratio is very small. Thus we tend to accept (or do not reject) the null hypothesis and tend to conclude that the variable has no effect on the dependent variable.

(e) A high R^2 and few significant t-ratios: In equation (10.6) we notice that the R^2 is very high, about 98% or 0.98. The t-ratios of both the explanatory variables are not statistically significant. Only the price variable slope coefficient has significant t-value. However, using F-test while testing overall significance $H_0: R^2 = 0$, we reject the null hypotheses. Thus there is some discrepancy between the results of the F-test and the t-test.

(f) The OLS estimators are mainly partial slope coefficients and their standard errors become very sensitive to small changes in the data. If there is a small change in data, the regression results change substantially.

(g) Wrong signs of regression coefficients: It is a very prominent impact of the presence of multicollinearity. In the case of the example given at equation (10.6) we find that the coefficient of the variable income is negative. The income variable has a 'wrong' sign as economic theory suggests that income effect is positive unless the commodity concerned is an inferior good.

10.4 DETECTION OF MULTICOLLINEARITY

In the previous section we pointed out the consequences of multicollinearity. Now let us discuss how multicollinearity can be detected.

(h) High R^2 and Few Significant t-ratios

This is the classic symptom of multicollinearity. If R^2 is high (greater than 0.8), the null hypothesis that the partial slope coefficients are jointly or simultaneously equal to zero $\left[H_0 : \beta_2 = \beta_3 = 0\right]$ is rejected in most cases (on the basis of F-test). But the individual t-tests will reflect that none or very few partial slope coefficients are statistically different from zero. This suggests very few slope coefficients are statistically significant.

(ii) High Pair-wise Correlations among Explanatory Variables

Due to high correlation among the independent variables, the estimated regression coefficients have high standard errors. But this is not necessarily true as demonstrated below. Even low correlation among the independent variables can lead to the problem of multicollinearity.

Let r_{23}, r_{24} and r_{34} represent the pair-wise correlation coefficients between X_2 and X_3 and X_4 respectively. Suppose $r_{23} = 0.90$, reflecting high collinearity between X_2 and X_3. Let us consider partial correlation coefficient $r_{23.4}$ that indicates correlation between X_2 and X_3 (while keeping the influence of X_4 constant). Suppose we find that $r_{23.4} = 0.43$. It indicates that partial correlation between X_2 and X_3 is low reflecting the absence of high collinearity. Therefore, pair-wise correlation coefficient when replaced by partial correlation coefficients does not indicate the presence of multicollinearity. Suppose the true population regression is given by equation (10.8)

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_2 X_{3i} + \beta_4 X_{4i} + u_i \quad \ldots (10.8)$$

Suppose the explanatory variables are perfectly correlated with each other as shown in equation (10.9) below

$$X_{4i} = \lambda_2 X_{2i} + \lambda_3 X_{3i} \quad \ldots (10.9)$$

X_4 is an exact linear combination of X_2 and X_3

If we estimate the coefficient of determination by regressing X_4 on X_2 and X_3, we find that

$$R^2_{4.23} = \frac{r^2_{42} + r^2_{43} - 2r_{42}\, r_{43}\, r_{23}}{1 - r^2_{23}} \quad \ldots (10.10)$$

Suppose, $r_{42} = 0.5, r_{43} = 0.5, r_{23} = -0.5$. If we substitute these values in equation (10.10), we find that $R^2_{4.23} = 1$. An implication of the above is that all the correlation coefficients (among explanatory variables) are not very high but still there is perfect multicollinearity.

(iii) Subsidiary or Auxiliary Regressions

Suppose one explanatory variable is regressed on each of the remaining variables and the corresponding R^2 is computed. Each of these regressions is referred to as subsidiary or auxiliary regression. For example, in a regression model with seven explanatory variables, we regress X_1 on X_2, X_3, X_4, X_5, X_6 and X_7 and find out the R^2_1. Similarly, we can regress X_2 on X_1, X_3, X_4, X_5, X_6 and X_7 and find out the R^2_2. By examining the auxiliary regression models we can find out the possibility

of multicollinearity. We take the rule of thumb that multicollinearity may be troublesome if R_i^2 obtained from auxiliary regression is greater than overall R^2 of the regression model.

A limitation of this method is that we have to compute R_i^2 several times, which is cumbersome and time consuming.

(iv) Variance Inflation Factor (VIF)

Another indicator of multicollinearity is the variance inflation factor (VIF). The R_i^2 obtained from auxiliary regressions may not be a reliable indicator of collinearity. In VIF method we modify the formula of variance of the estimators as follows; (b_2) and (b_3)

$$\text{var}\,(b_2) = \frac{\sigma^2}{\sum x_{2i}^2(1-R_2^2)} = \frac{\sigma^2}{\sum X_{2i}^2}.\left(\frac{1}{1-R_2^2}\right) \quad \text{... (10.11)}$$

In equation (10.11), you should note that R_2^2 is the auxiliary regression discussed earlier.

Compare the variance of b_2 given in equation (10.11) with the usual formula for variance of an estimator given in Unit 4. We find that

$$var(b_2) = \frac{\sigma^2}{\sum x_{2i}^2}\ VIF \quad \text{... (10.12)}$$

where VIF $= \left(\frac{1}{1-R_2^2}\right)$

Similarly, $var(b_3) = \frac{\sigma^2}{\sum x_{3i}^2}(VIF)$

Note that as R_i^2 increases the VIF also increases. This inflates the variance and hence standard errors of b_2 and b_3

If $R_i^2 = 0$, $VIF = 1 \Rightarrow V(b_2) = \frac{\sigma^2}{\sum x_{2i}^2}$ and $V(b_3) = \frac{\sigma^2}{\sum x_{3i}^2}$

Therefore, there is no collinearity.

On the other hand,

if $R_i^2 = 1$, $VIF = \infty \Rightarrow V(b_2) \to \infty, V(b_3) \to \infty$

If R_i^2 is high, however $V(b_2)$ tends to ∞.

Note that $var(b_2)$ depends not only on R_i^2, but also on σ^2 and $\sum x_{2i}^2$. It is possible that R_i^2 is high (say, 0.91) but $var(b_2)$ could be lower due to low σ^2 or high $\sum x_{2i}^2$. Thus $V(b_2)$ is still lower resulting in high *t* value. Thus R_i^2 obtained from auxiliary regression is only a superficial indicator of multicollinearity.

Check Your Progress 2

1) Bring out four important consequences of multicollinearity.

.....

.....

.....

.....

.....

2) Explain how multicollinearity can be detected using partial correlations.

.....

.....

.....

.....

.....

3) Describe the method of detection of multicollinearity using the variance inflation factor (VIF).

.....

.....

.....

.....

.....

10.5 REMEDIAL MEASURES OF MULTICOLLINEARITY

Multicollinearity may not necessarily be an "evil'' if the goal of the study is to forecast the mean value of the dependent variable. If the collinearity between the explanatory variables is expected to continue in future, then the population regression function can be used to predict the relationship between the dependent variable Y and other collinear explanatory variables.

However, if in some other sample, the degree of collinearity between the two variables is not that strong the forecast based on the given Regression is of little use.

On the other hand, if the objective of the study is not only prediction but also reliable estimations of the individual parameters of the chosen model then serious collinearity may be bad, since multicollinearity results in large standard errors of estimators and therefore widens confidence interval. Thus, resulting in accepting null hypotheses in most cases. If the objective of the study is to estimate a group

of coefficients (i.e., sum or difference of two coefficients) then this is possible even in presence of multicollinearity. In such a case multicollinearity may not be a problem.

$$Y_i = C_1 + C_2 X_{2i} + u_i \qquad \ldots(10.13)$$

$$C_1 = A_1 + 300A_3, \qquad C_2 = A_2 - 2A_3$$

Running the above regression in equation (10.2), as presented in earlier section 10.2, one can easily estimate C_2 by using OLS method, although neither A_2 nor A_3 can be estimated individually. There can be situation when in spite of inflated S.E., the individual coefficients happened to be numerically significant since the true value itself is so large even or estimate on the downside still shows up a significant test.

Certain remedies prescribed for reducing the severity of collinearity problem which can be listed as OLS estimators can still retain BLUE property despite of near collinearity. Further, one or more regression coefficients can e individually statistically significant or some of them with wrong signs.

10.5.1 Dropping a Variable from the Model

The simplest solution may be to drop one or more of the collinear variables. However, dropping a variable from the model may lead to model specification error. In other words, when we estimate the model without the excluded variable, the estimated parameters of the reduced model may turn out to be biased. Therefore, the best practical advice is not to drop a variable from a model that is theoretically sound. A variable which has *t* value of its coefficient greater than 1, then than variable should not be dropped as it will result in a decrease in $\bar{R}^2$.

10.5.2 Acquiring Additional Data or New Sample

Acquiring additional data implies increasing the sample size. This is likely to reduce the severity of the multicollinearity problem. As we know from equation (10.11),

$$\text{var}\,(b_2) = \frac{\sigma^2}{\sum x_{2i}^2 (1 - R_2^2)}$$

Given σ^2 and R_2^2, if the sample size of X_2 increases, there is an increase in $\sum x_{2i}^2$. It will lead a decrease in var (b_2) and its standard error.

10.5.3 Re-Specification of the Model

It is possible that some important variables are omitted from the model. The functional form of the model may also be incorrect. Therefore, there is a need of looking into the specification of the model. Many times, taking log form of a model leads to solving the problem of multicollinearity.

10.5.4 Prior Information about Certain Parameters

Estimated values of certain parameters are available in existing studies. These values can be used as prior information. These values give us some tentative idea on the plausible value of the parameters.

10.5.5 Transformation of Variables

Transformation of the variables would minimize the problem of collinearity.

10.5.6 Ridge Regression

The ridge regressions are another method of resolving the problem of multicollinearity. In the ridge regression, the first step is to standardize the variables both dependent and independent by subtracting the respective means and dividing by their standard deviations. This mainly implies that the main regression is run by transforming both dependent and explanatory variables into the standardized values.

It is observed that in the presence of multicollinearity, the value of variance inflation factor is substantially high. This is mainly due to a high value of coefficient of determination. The ridge regression is applied when the regression equations are in the form of matrix involving large number of explanatory variables.

The ridge regression proceeds by adding a small value, k, to the diagonal elements of the correlation matrix. The reason that the diagonal of ones in the correlation matrix could be considered as a ridge, this is the reason such regression is referred as ridge regression.

10.5.7 Other Remedial Measures

There are several other Remedies suggested such as combining time series and cross-sectional data, factor or principal component analysis and ridge regressions.

Polynomial Regression Models

Let us consider total cost of production (TC) as a function of output as well as marginal cost (MC) and Average Cost (AC)

$$Y_i = \beta_1 + \beta_2 X_1 + \beta_3 X_i^2 + \beta_4 X_i^3 \qquad \text{…….(10.12)}$$

The cost function is defined as Cubic function for cost as a third-degree polynomial of variable X. This model in equation (10.12) is linear in parameters β^s, therefore satisfy assumption of CLRM of linear Regression Model and can be estimated by usual OLS method. However, one needs to worry about problem of collinearity since it is not linear in variables and at the same time X^2 and X^3 are non-linear function of X and do not violate the assumptions of no perfect collinearity i.e., no perfect linear relationship between variables. The estimated results are presented in equation (10.13).

$$\hat{Y}_i = 141.7667 + 63.4776X_i - 12.9615X_i^2 + 0.9396\ X_i^3 \qquad \text{….(10.13)}$$

Se (6.3753) (4.7786) (0.9857) (0.0591)

$$R^2 = 0.9983$$

$$AC = \frac{RC}{X_i} = \frac{141.7667}{X_i} + 63.4776 - 12.96X_i + (0.9396)X_i^2$$

$$AC_i = 63.4776 - 12.9615X_i + 141.7667X_i + 0.9396X_i^2$$

$$MC = \frac{\partial TC}{\partial X_i} = 63.4776 - 2X(12.9615)X_i + 3 \times 0.9396X_i^2$$

If the cost curves are U-shaped Average Marginal cost curves then the theory suggests that the coefficient should satisfy following

1) β_1, β_2 and $\beta_4 > 0$
2) $\beta_3 < 0$
3) $\beta_3^2 < 3\beta_2\beta_4$

Check Your Progress 3

1) Define two significant methods to rectify the problem of multicollinearity?

.....

.....

.....

.....

.....

2) Describe the method of ridge regression.

.....

.....

.....

.....

.....

10.6 LET US SUM UP

This unit presents a clear understanding of the concept of multicollinearity in the regression model. The unit also presents a clear distinction of near and perfect multicollinearity. The unit familiarizes the consequences of presence of multicollinearity in regression model. The method of detection of multicollinearity has been highlighted in the unit. Finally various techniques that provide remedial measures including the concept of ridge regression have been explained in the unit.

10.7 ANSWERS/ HINTS TO CHECK YOUR PORGRESS EXERCISES

Check Your Progress 1

1) The case of perfect multicollinearity mainly reflects the situation when the explanatory variables and perfectly correlated with each other implying the coefficient of correlation between the explanatory variables is 1.

2) This refers to the case when two or more explanatory variables are not exactly linear this reinforces the fact that collinearity can be high but not perfect. "High collinearity" refers to the case of "near" or imperfect" or high multicollinearity. Presence of multicollinearity implies "imperfect multicollinearity''

3) In the case of perfect multicollinearity it is not possible to obtain estimators for the parameters of the regression model. See Section 10.2 for details.

Check Your Progress 2

1) (i) In case of imperfect multicollinearity, some of the estimators are statistically not significant. But OLS estimates still retain their BLUE property that is, Best Linear Unbiased Estimators. Therefore, imperfect multicollinearity does not violate any of the assumptions, OLS estimators retain BLUE property. Being BLUE with minimum variance does not imply that the numerical value of variance will be small.

(ii) The R^2 value is very high but very few estimators are significant (t-ratios low). The example mentioned in earlier section where the demand function of good Y we computed using the earnings of individuals, reflects the situation where R^2 is quite high about 98% or 0.98 but only price variable slope coefficient has significant t-value. However, using F-test while testing overall significance $H_0 : R^2 = 0$, we reject the hypotheses that both prices and earnings have no effect on the demand of Y.

(iii) The ordinary least square OLS estimators mainly partial slope coefficients and their standard errors become very sensitive to small changes in the data, i.e. they then to be rentable. A small charge of data, the regression results change quite substantially as in case example of near or imperfect multicollinearity mentioned above, the standard errors go down and t-ratios have increased in absolute values.

(iv) Wrong signs of regression coefficients. It is a very prominent impact of presence of multicollinearity. In case of example where earnings of individuals were used in deriving demand curve of good Y, the earning

variable has the 'wrong' sign for the economic theory since the income effect usually positive unless it is case of inferior good.

2) Examining partial correlations: In case of three explanatory variables X_2, X_3 and X_4 very high or perfect multicollinearity between X_4 and X_2, X_3.

Subsidiary or auxiliary regressions: When one explanatory variables X is regressed on each of the remaining X variable and the corresponding R^2 is computed. Each of these regressions is referred as subsidiary or auxiliary regression. A regression Y on X_2, X_3, X_4, X_5, X_6 and X_7 with six explanatory variables. If R^2 comes out to be very high but few significant t-ratios or very few X coefficients are individually statistically significant then the purpose is to identify the source of the multicollinearity or existent of perfect or near perfect linear combination of other X^s.

For this we Regress X_2 on remaining X^s and obtain R_2^2 or also written as $R^2_{2.34567}$

Regress X_3 on remaining X^s, and obtain R_3^2 coefficient of determination also written as $R^2_{3.24567}$ each R_i^2 obtained will lie between 0 and 1. By testing the null hypothesis $H_0 : R_i^2 = 0$ by applying F-test. Let r_{23}, r_{24} and r_{34} represent pairwise correlation between X_2 and X_3, X_2 and X_4, X_3 and X_4 respectively suppose $r_{23} = 0.90$, reflecting high collinearity between X_2 and X_3. Considering partial correlations coefficient $r_{23.4}$ that indicators correlations coefficient between X_2 and X_3, Adding the influence of X_4 constant. If $r_{23.4} = 0.43$. Thus, partial correlation between X_2 and X_3 is low reflecting no high collinearity or low degree of collinearity. Therefore, pairwise correlation when replaced by partial correlation coefficients does not provide indicator of presence of multicollinearity.

3) Variance Inflation Factor (VIF): R^2 obtained variables auxiliary regression may not be completely reliable and is not reliable indicator of collinearity. In this method we modify the formula of var (b_2) and (b_3)

$$\text{var}(b_2) = \frac{\sigma^2}{\sum x_{2i}^2 \left(1 - R_2^2\right)}$$

$$= \frac{\sigma^2}{\sum X_{2i}^2} \cdot \left(\frac{1}{1 - R_2^2}\right)$$

$$\text{VIF} = \left(\frac{1}{1 - R_2^2}\right) \qquad \therefore V(b_2) = \frac{\sigma^2}{\sum x_{2i}^2} \text{V.I.F.}$$

Similarly, $V(b_3) = \frac{\sigma^2}{\sum x_{3i}^2}(VIF)$

VIF is variance inflation factor. As R^2 increases VIF $\frac{1}{1-R^2}$ increased thus inflating the variance and hence standard errors of b_2 and b_3

If $R^2 = 0$, $VIF = 1 \Rightarrow V(b_2) = \frac{\sigma^2}{\sum x_{2i}^2}$ and $V(b_3) = \frac{\sigma^2}{\sum x_{3i}^2}$

$\Rightarrow$ No collinearity

If $R^2 = 1$, $VIF = \infty \Rightarrow V(b_2) \to \infty, V(b_3) \to \infty$

If R^2 is high, however $var(b_2) \to \infty, var(b_3)$ does not only depend on R^2 (auxiliary coefficient of determination) or VIF. If also depends on σ^2 and $\sum x_{2i}^2$ it is possible that R_i^2 is high 0.91 but $var(b_2)$ could be lower due to low σ^2 or high $\sum x_{2i}^2$ thus $V(b_2)$be still lower resulting in high t value not showing any low t end thus defeating the indicator of multicollinearity. Thus R^2 obtained from and binary regression is only a surface indicator of multicollinearity.

Check Your Progress 3

1) (i) Dropping a variable from the Model: The simplest solution might seem to be to drop one or more of the collinear variables. However, dropping a variable from the model may lead to model specification error in either words, where we estimate the model without that variable, the estimated parameters of reduced model may turn out to be biased. Therefore, the best practical advice is not to drop or variable from an economically variable model first because the collinearity problem is serious. A variable which has t value of its coefficient greater than 1, then than variable should not be dipped as it will result in decrease in adjusted $\overline{R}^2$

(ii) Acquiring Additional Data or new sample: Acquiring additional data implies increasing the sample size can reduce the severity of collinearity problem.

$$V(b_2) = \frac{\sigma^2}{\sum x_{3i}^2 (1 - R_2^2)}$$

Given σ^2 and R^2, if the sample size of X_3 increases $\Rightarrow \sum x_{3i}^2$ will increase as a result $V(b_3)$will tend to decrease and standard error b_3 will also.

2) In ridge regression we first standardise all the variables in the model. Go through Sub-Section 10.5.6 for details.