[Academic Script]

Prediction in Linear Models and Multicollinearity

Subject:	Business Economics
Course:	B. A. (Hons.), 5th Semester, Undergraduate
Paper No. & Title:	Paper – 531 Elective Paper Q1 – Advanced Econometrics
Unit No. & Title:	Unit – 1 Relaxing the Assumptions of The Classical Linear Model
Lecture No. & Title:	Lecture – 3 Prediction in Linear Models and Multicollinearity

Academic Script

1. Prediction in Linear Models

Welcome friends today we want to study prediction in linear models and Multicollinearity. As you know there are many statistical applications by means of statistical models and stee theories, they will be studied unless we get applications it will be meaningless. In this applications also it is the last part that is predictions in the linear models. So we will study certain things connected to two variables models and k variable models. Next we will have the problem of multicollinearity. It is also an important factor when you want to analyse data. Here we will know how multicollinearity arises, how we can detect them and what are the remedial actions for multicollinearity.

While dealing with linear model, it is equally important to know about prediction based upon fitted model. This is essentially very useful for forecasting and also decision making excercises. We shall deal with this topic in two parts-(1.1) case of two variables model and (1.2) case of K variate model.

(1.1) Two Variables Case

Suppose that we may be interested to know about the effect of raising petrol (or diesel) price on consumption by its users. We can have the analysis in two parts (i) Individual prediction (ii) Mean prediction

(i) Individual Prediction

We have the model Y =α+ βX_i + U_i

Here X = Price of petrol

Y = Consumption

At a given price $X_i = X_0$, we can have the actual consumption given by

$Y_0 = \alpha + \beta X_0 + U_0$

Where U_0 = Disturbance term

When we find OLSE of α and β then we get prediction $= E(Y|X_0)$ $= \hat{Y}_0 = \hat{\alpha} + \hat{\beta}X_0$

Here $e_0 = Y_0 - \hat{Y}_0$ = Prediction Error

$= Y_0 - (\hat{\alpha} + \hat{\beta}X_0)$

Hence $E(e_0) = 0 \Rightarrow E(\hat{Y}_0) = Y_0$ which means that Least Squares predictor is unbiased.

$V(e_0) = E({e_0}^2) = V(U_0) + V(\hat{\alpha}) + V(\hat{\beta}){X_0}^2$

Which gives $V(e_0) = \sigma^2 \left[1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum {x_i}^2}\right]$

This shows that $V(e_0)$ is minimum at $X_0 = \bar{X}$ and it is given by

$V(e_0) = \sigma^2 \left(1 + \frac{1}{n}\right)$

S.E. of prediction is given by

S. E. $= \sqrt{V(e_0)}$

$= \sigma \sqrt{1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum {x_i}^2}}$

As before $\hat{\sigma} = \sqrt{\sum {e_i}^2 / (n-2)} = S$

We can apply t test for significance of individual prediction $\hat{Y}_0$ given by the formula

$t = \frac{Y_0 - \hat{Y}_0}{S.E. of\ \hat{Y}_0}$

Which has student's t distribution with $(n-2)$ degrees of freedom

Also $100(1-\alpha)$ % confidence interval for Y_0 is given by the formula

$\hat{Y}_0 \pm t_{\alpha/2}$. S. $\sqrt{1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum {x_i}^2}}$

(e.g. 95% confidence interval for Y_0 will be

$(\hat{\alpha} + \hat{\beta}X_0) \pm t_{0.025,(n-2)}$.(S.E. of e_0)

Where $t_{0.025,(n-2)}$ is tabulated value of t for significance level 0.025 and d. f. $(n-2)$

(ii)Mean Value Prediction

Many times we may be interested about mean prediction instead of individual prediction and practically it makes more sense to consider such prediction.

i. e. To consider $E(Y_0)$ rather than Y_0 itself

We have$E(Y_0) = \alpha + \beta X_0$, and here the prediction error defined as $e_0{}^*$ will be given by $e_0{}^* = E(Y_0) - Y_0$

$$= (\alpha + \beta X_0) - (\hat{\alpha} + \hat{\beta} X_0)$$

Then $E(e_0{}^*) = 0 \Rightarrow$ linear predictor is unbiased.

Here $V(e_0{}^*) = E(e_0{}^*)^2$ and it is given by

$$V(e_0{}^*) = \sigma^2 \left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i{}^2}\right]$$

Which indicates that it is minimum at $X_0 = \bar{X}$, thus $V(e_0{}^*) = \frac{\sigma^2}{n}$

As before we have t statistic for testing of this prediction (σ^2 replaced by $\hat{\sigma}^2 = S^2 = \sum e_i{}^2 / (n-2)$) and 100(1–α) % confidence

interval for Mean prediction will be obtained from

$$\left(\hat{\alpha} + \hat{\beta} X_0\right) - S.\, t_{\alpha/2,n-2} \leq E(Y_0) = \hat{\alpha} + \hat{\beta} X_0 \leq \left(\hat{\alpha} + \hat{\beta} X_0\right) + S.\, t_{\alpha/2,n-2}$$

(As for example, 99% confidence interval for Mean value prediction is given by the expression.

$$\left(\hat{\alpha} + \hat{\beta} X_0\right) + S.\, t_{0.005,n-2} \leq E(Y_0|X_0) \leq \left(\hat{\alpha} + \hat{\beta} X_0\right) + S.\, t_{0.005,n-2}$$

Here $t_{0.005,(n-2)}$is the tabulated value of t statistics for d. f. $(n-2)$ and significance Level 0.005

(1.2) K variables case

(1) Individual prediction

We write the linear model as

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} + U_i \text{ (i = 1, 2… n)}$$

Now suppose that the explanatory Variables X_2 ,X_3 ,. . . . . , X_k have some known (fixed) values denoted by X_{20} ,X_{30} ,. . . . . ,X_{k0} and we want to know about predicted value Y_0 on the basis of OLS estimation of the model

Then $Y_0 = \beta_1 + \beta_2 X_{20} + \beta_3 X_{30} + . \ldots + \beta_k X_{k0} + U_0$

$\hat{Y}_0 =$ Prediction of $Y = \hat{\beta}_1 + \hat{\beta}_2 X_{20} + \hat{\beta}_3 X_{30} + . \ldots + \hat{\beta}_k X_{k0}$

We can write $Y_0 = \underline{X}'_0 \underline{\beta} + U_0$

$$\hat{Y}_0 = \underline{X}'_0 \underline{\hat{\beta}}$$

Where $\underline{X}'_0 = [1, X_{20}, X_{30}, ... , X_{k0}] : 1 \times K$

$\underline{\hat{\beta}} = (\hat{\beta}_1 , \hat{\beta}_2, ... , \hat{\beta}_k) : 1 \times k$, $\underline{\beta} = (\beta_1, \beta_2, \beta_3, ... , \beta_k) : 1 \times k$

Then prediction error $= d_0 = Y_0 - \hat{Y}_0 = U_0 - \underline{X}'_0 (\underline{\hat{\beta}} - \underline{\beta})$

$$E(d_0) = 0 \Rightarrow V(d_0) = \sigma^2 \left[1 + \underline{X}'_0 (X' X)^{-1} \underline{X}_0\right]$$

σ^2 is estimated by $S^2 = \sum e_i^2 \Big/ (n-k)$

Upon normality assumption of disturbances, that is $U_i \sim I_N (0, \sigma^2)$ i = 1, 2… n

We have $\hat{Y}_0 \sim N\left[\underline{X}'_0 \underline{\hat{\beta}}, \sigma^2 \left\{1 + \underline{X}'_0 (X' X)^{-1} \underline{X}_0\right\}\right]$

Hence as usual we can carry out t test by using the formula

$$t = \frac{d_0}{S.E. of\ d_0}$$

S.E. of $d_0 = S \sqrt{1 + \underline{X}'_0 (X' X)^{-1} \underline{X}_0}$ with degrees of freedom (n$-k$)

Also 100(1$-\alpha$) % confidence interval for prediction Y_0 will be given by the formula

$$\hat{Y}_0 - S \sqrt{1 + \underline{X}'_0 (X' X)^{-1} \underline{X}_0} \; . t_{\alpha/2, n-K} \leq Y_0 \leq \hat{Y}_0 + S \sqrt{1 + \underline{X}'_0 (X' X)^{-1} \underline{X}_0} \; . t_{\alpha/2, n-K}$$

Where $t_{\alpha/2,n-K}$ is the value of t for significance level $\alpha/2$ and degrees of freedom (n$-k$)

(2)Mean Prediction

Concept of mean prediction is more reasonable than individual prediction. We extend this for k variate General Linear Model.

We have $Y_0 = \underline{X'}_0 \underline{\beta} + U_0$

Then Mean Prediction $= E(Y|\underline{X}_0) = \underline{X'}_0 \underline{\hat{\beta}}$

Here Prediction error $= d^*_0 = Y_0 - E(Y|\underline{X}_0)$

$$= U_0 - \underline{X'}_0 \left(\underline{\hat{\beta}} - \underline{\beta}\right)$$

$$E(d_0^*) = 0$$

$$V(d_0^*) = \sigma^2 \left(\underline{X'}_0 (X'X)^{-1} \underline{X}_0\right)$$

Upon normality assumption we have

$E(Y|\underline{X}_0) \sim N\left[\underline{X'}_0 \underline{\beta}, \sigma^2 \underline{X'}_0 (X'X)^{-1} \underline{X}_0\right]$

σ^2 is estimated by $\sum e_i^2 / (n-k)$.

We have t statistics given by

$$t = \frac{d_0^*}{S.E.of\ d_0^*} = \frac{Y_0 - E(Y|\underline{X}_0)}{\sqrt{V(d_0^*)}}$$

Which gives $t = \frac{Y_0 - E(Y|\underline{X}_0)}{S\sqrt{\underline{X'}_0 (X'X)^{-1} \underline{X}_0}}$

having student's t distribution with $(n-k)$ degrees of freedom, apply t test for significance of mean prediction.

Also 100(1$-\alpha$)% confidence interval for mean prediction will be given by

$$\underline{X'}_0 \underline{\hat{\beta}} \pm S\sqrt{\underline{X'}_0 (X'X)^{-1} \underline{X}_0}\ .\ t_{0.005,(n-k)}$$

Where $t_{0.005,(n-K)}$ is the value of t statistics for significance level 0.025 and degrees of freedom $(n-k)$.

2. Multicollinearity

1 what is multicollinearity?

In linear models, multicollinearity occurs if one (or more) explanatory variable can be expressed as a linear combination of other variables. This concept was given by Ragnar Frisch and it corresponds to the concept of linear dependence of vectors.

In linear models, we have one basic assumption that all the explanatory variables are linearly independent. Due to his in k variate model, with data matrix $X: n \times k,$ if there is no multicollinearity, $|X'X| \neq 0$ and thus Rank of $(X'X) = \rho\,(X'X) = k,$ so that matrix $X'X$ is non-singular and $(X'X)^{-1}$ exists

If multicollinearity arises then $\rho\,(X'X) < k$ and $X'X$ is not a full rank matrix

In particular if $|X'X| = 0,$ then $X'X$ is a singular matrix and regular inverse $(X'X)^{-1}$ does not exist. This is the case of perfect multicollinerarity.

Diagrammatic presentation is as under

(Here variable Y is shown as dependent on X_2 ,X_3 and X_4)

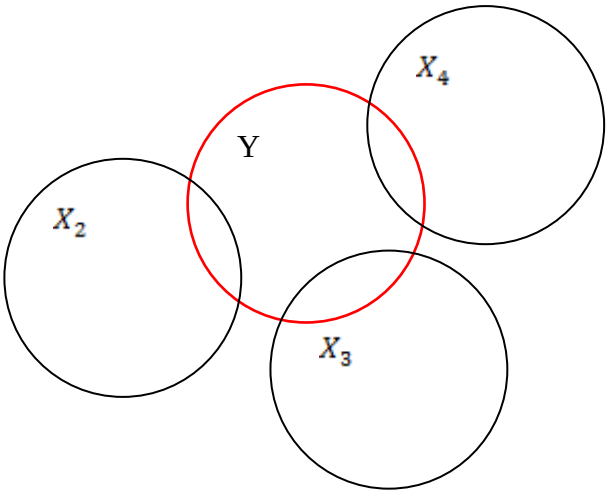

Figure: 1 No multicollinearity between X_2 ,X_3 andX_4

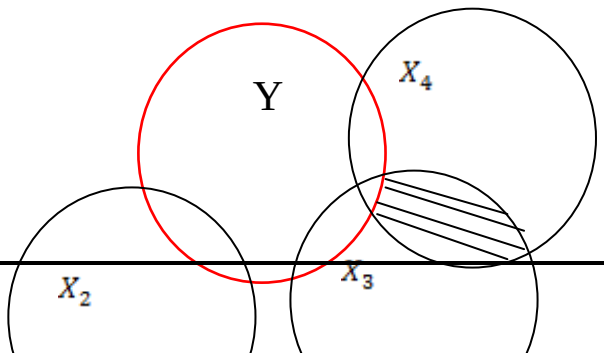

Figure :2 Here X_2 is Independent with X_3 and X_4, but X_3 and X_4 are related thus there is multicollinearity between X_3 and X_4.

The case of multicollinearity is thus violation of one of the basic assumptions of classical linear model. (It is violation of rank condition)

To understand this concept let us take a very simple illustration as under

Variable X_2	Variable X_3	Variable X_4
10	50	52
15	75	75
18	90	92
24	120	129
30	150	152

We find that $X_{3i} = 5X_{2i}$ and here correlation coefficient between X_2 and X_3, that is $r_{23} = 1$.

Also $X_{4i} = 5X_{2i} + \vartheta_i$ where ϑ_i is stochastic error.

Here $r_{24} = 0.9959$ Thus there is perfect multicollinearity between X_2 and X_3 ($X_3 = \lambda X_2$) but there is imperfect multicollinearity between X_2 and X_4 ($X_4 = \lambda X_2 + \vartheta$)

We give below two actual applications to understand the concept

(1) $Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + U$

Where Y = consumption

X_2 = Income

X_3 = Wealth

Here X_2 and X_3 can be multicollinear as income can generate wealth and vice versa.

(2)

(GDP)=Constant+$\beta_2(Population) + \beta_3(Population\ Density) + \beta_3(time) + U$

We cannot take both explanatory variables population as well as population density. This generates multicollinearity.

3. Practical Consequences of multicollinearity

We have two cases (as seen above) for perfect multicollinearity and imperfect multicollinearity. (Less than perfect multicollinearity)

(I) In the case of perfect multicollinearity, the OLS estimators are indeterminate and also their variances (and hence standard errors) are indeterminate.

(II)In the case of severe but Less than perfect multicollinearity, we have the following consequences

(1)OLS estimators can be obtained but their variances (and hence standard errors) become very large (more in the case of high degrees of multicollinearity)

(2)Due to the large standard errors the confidence intervals also become larger and do not remain meaningful.

(3)Due to high degree of multicollinearity, the probability of accepting a false hypothesis increases

(4)Standard errors become much more sensitive due to slight change in the data in the presence of multicollinearity

(5)Under high degree of multicollinearity a higher value of R^2 is obtained but none or very few estimated regression coefficients are found to be statistically significant.

To illustrate this, we consider 3 variables model as shown below

$Y_i = \hat{\beta}_2\, x_{2i} + \hat{\beta}_3\, x_{3i} + e_i$

Here all the variables are measured from their means. (Thus $y_i = Y_i - \bar{Y}$, $x_{2i} = X_{2i} - \overline{X_2}$, $x_{3i} = X_{3i} - \overline{X_3}$)

OLS estimators are given by the formula

$$\hat{\beta}_2 = \frac{(\sum y_i x_{2i})(\sum x_{3i}^2) - (\sum y_i x_{3i})(\sum x_{2i} x_{3i})}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2}$$

$$\hat{\beta}_3 = \frac{(\sum y_i x_{3i})(\sum x_{2i}^2) - (\sum y_i x_{2i})(\sum x_{2i} x_{3i})}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2}$$

Let us assume that $X_{3i} = \lambda X_{2i} \Rightarrow x_{3i} = \lambda x_{2i}$

(λ is an arbitrary constant)

Then $\hat{\beta}_2 = \frac{0}{0}$ = Indeterminate

$\hat{\beta}_3 = \frac{0}{0}$ = Indeterminate (Due to perfect multicolllinearity)

$V(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_{2i}^2 (1 - r_{23}^2)}$ σ^2 = Variance of disturbance term

$V(\hat{\beta}_3) = \frac{\sigma^2}{\sum x_{3i}^2 (1 - r_{23}^2)}$ $r_{23} = corr(X_2, X_3)$

If $r_{23} = 1$, $V(\hat{\beta}_2) = \infty$, $V(\hat{\beta}_3) = \infty$

Hence they are indeterminate. Larger the value of r_{23} (that means higher degree of multicollinearity), larger will be the variances and hence standard errors will also be larger.

How to detect multicollinearity?

Multicollinearity is a sample phenomenon arising due to the largely non experimental data collected.

We do not have one unique method for its detection. There are several rules of thumb which are commonly used for its detection.

(1) High R^2 but few significant t ratios

Collinearity is often suspected when R^2 is high (that is from 0.8 to 1) and when zero order correlations are also high but none or very few of the partial regression coefficient are significant

(2) High pairwise correlation among Regressors

One must check for the partial correlation coefficients also.

For example, for regression of Y on X_2 ,X_3 andX_4

If we find that R^2 is very high but the partial correlation coefficients denoted by $r^2{}_{12\cdot34}$, $r^2{}_{13\cdot24}$ and $r^2{}_{14\cdot23}$ are comparatively low, we can suspect multicollinearity between X_2 ,X_3 andX_4 , so that at least one of these variables is superfluous.

(3) Auxilliary Regressions

Since multicollinearity arises due to one or more of the explanatory variables having exact or approximate linear combinations, of the other explanatory variables, one way of finding out which X variable is related to other X variables is to regress each X_i on the other remaining X variables and compute the value of R^2 in each case. We denote it by R_i^2

Each such regressions are called auxilliary regressions. On the basis of R_i^2, we compute the ratio

$$F_i = \frac{R_i^2/(k-2)}{(1-R_i^2)/(n-k+1)}$$

Which has F distribution with $(k-2)$ and $(n-k+1)$ degrees of freedom. If F_i is significant it shows that X_i is collinear with other $X's$. if is not significant, we can retain that variable in the model.

(4) Klein's rule of thumb

Let R^2 = Multiple coefficient of determination when Y is regressed against all $X's$

And R_i^2 = Multiple coefficient of determination for auxiliary regression when X_i is regressed against other $X's$

Then if $R^2{}_i > R^2$, we may conclude that multicollinearity may be a troublesome problem.(However this is only an approximate rule)

(5) Eigen values and condition Index

We have $|X'X| = \prod_{i=1}^{k} \lambda_i = \lambda_1, \lambda_2 \ldots \lambda_k$

Where λ_i is the i^{th} eigen value (characteristic root) of the matrix $X'X$

We can obtain λ_{max} = Maximum Eigen value

And λ_{min} = Minimum Eigen value

Then define condition number K by

$K = \frac{\lambda_{max}}{\lambda_{min}}$ and further condition Index is given by CI $= \sqrt{K} = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}}$

If K lies between 100 and 1000 there is moderate to strong multicollinearity and if K exceeds 1000, there is severe multicollinearity.

Thus if CI lies between 10 to 30, there is moderate to strong multicollinearity, but it is severe if CI exceed 30.

(6) Tolerance and Variance Inflation Factor

For k variate GLM, V $(\hat{\beta}_i) = \frac{\sigma^2}{\sum {x_i}^2}\left(\frac{1}{1-{R^2}_i}\right)$

$$= \left(\frac{\sigma^2}{\sum {x_i}^2}\right)(VIF)_i$$

Where $\hat{\beta}_i$ = estimated partial regression coefficient

R_i^2 = Multiple coefficient of determination for auxilliary regression of X_i on other $X's$

$(VIF)_i$ = Variance Inflation Factor

$= \frac{1}{(1-{R^2}_i)}$

VIF is indicator for multicollinearity. Larger the value of VIF, more troublesome is multicollinearity

As a rule of thumb, if VIF exceeds 10, there is very high multicollinearity.

Other measure is tolerance which is defined as

$(TOL)_i = \left(1 - R_i^2\right) = \frac{1}{(VIF)_i}$

$(TOL)_i$ lies between 0 and 1

Clearly $(TOL)_i$ is equal to 1, if $R_i = 0$, and here X_i is not related to other $X's$ and there is no multicollinearity.

If $(TOL)_i$ is equal to 0, if $R_i = 1$ and here X_i is perfectly linearly related to other $X's$ and there is perfect multicollinearity.

4. How to takle the problem of multicollinearity?

(Remedial Measures for multicollinearity)

There are several methods as the remedial measures for tackling the problem of multicollinearity

1. By Using A priori information

On the knowledge of some parameter values beforehand we can deal with the problem of multicollinearity. Consider consumption, Income wealth model as

$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + U$

Where $Y =$ consumption

$X_2 =$ Wealth

$X_3 =$ Income

Due to nature of X_2 and X_3 , there is multicollinearity

Now suppose that $\beta_3 = 0.20\beta_2$ (that is about 20% of income is generated from wealth)

Then $Y = \beta_1 + \beta_2 X_2 + 0.20\beta_2 X_3 + U$

$= \beta_1 + \beta_2(X_2 + 0.20X_3) + U$

$= \beta_1 + \beta_2 X_4 + U$

Where $X_4 = X_2 + 0.20X_3$

This helps in resolving the problem of multicollinearity. We have used a priori information which is collected from past data or experience or analysis or intuition or from relevant economic theory etc.

2 By Combining Cross Sectional and time series data

Sometimes a combination of time series and cross sectional data can be useful in overcoming the problem of multicollinearity

Suppose we have the model

$$\log Y = \beta_1 + \beta_2 \log P + \beta_3 \log I + U$$

Where Y = Number of cars sold,

P = Average price

I = Income

Obviously P and I have multicollinearity. We may consider cross sectional data for this model first and we may obtain estimate of β_3 (denoted by $\hat{\beta}_3$). In cross section data the variation over a point in time is not too much.

We use this estimate $\hat{\beta}_3$ to run the time series model.

Thus $Y_i^* = \beta_1 + \beta_2 \log P_i + U_i$

Where $Y_i^* = \log Y_i - \hat{\beta}_3 \log I_i$

This resolves the problem of multicollinearity such method is also called pooling of time series and cross section data.

3 By dropping variable and specification bias

This method is simplest in the sense that one of the superfluous variables can be dropped from regression.

As for illustration, let $Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + U$

We suspect multicollinearity between X_2 and X_3.

We may drop X_2 or X_3 and run the model.

Suppose we drop X_3 then

$Y = \beta_1 + \beta_2^* X_2 + U$

Here $E(\beta_2^*) = \beta_2 + \beta_3 E(X_3)$

Hence β_2^* will be a biased estimate of β_2 as long as $E(X_3)$ is not zero.

This method appears to be simple but we commit specification error and specification bias and it can mislead to the true values of the parameters.

4 By transforming the variables

Since in multicollinearity, the variables tend to move in the same direction, sometimes it may be helpful for transformation of the variables by means of taking successive differences.

Thus Let $Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + U_t$

Then $\Delta Y_t = \beta_2 \Delta X_{2t} + \beta_3 \Delta X_{3t} + v_t$

Where $\Delta Y_t = Y_t - Y_{t-1}$

$\Delta X_{2t} = X_{2t} - X_{2,t-1}$

$\Delta X_{3t} = X_{3t} - X_{3,t-1}$

$\vartheta_t = U_t - U_{t-1}$

We can run regression based upon new form of the model.

Here by taking differences we may loose some observations and thus degrees of freedom also decreases which are very important for small samples. The assumption of linear model will be violated as U_t will be serially correlated. In the case of cross section data, this method is not appropriate.

5 Reducing Collinearity in Polynomial Regression

Suppose that we have a cubic cost function given by

$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + U$

Where $Y =$ cost, $X =$ output

We may write $X^2 = z_1$ and $X^3 = z_2$ and proceed further, but z_1 and z_2 have multicollinearity. One way to deal with the issue is to use observations deviated from their means, which reduces multicollinearity substantially.

Here techniques like orthogonal polynomials can be used.

6 Using additional or new data

Here when we suspect multicollinearity, we can use additional observations on the variables and augment data matrix X further. This increases sample size and helps in reducing standard errors of the estimates. But choosing additional observations may not be that easy as it seems.

7 Klein's Eigen value approach

We know that for k variate GLM, $|X'X| = \lambda_1, \lambda_2 \ldots \lambda_k$ Where λ_i is the i^{th} Eigen value of the matrix $X'X$.

If $|X'X|$ is very near to zero, we suspect the presence of multicollinearity.

Suppose that i^{th} Eigen value λ_i is very small. Then we choose a constant d and write the new augmented data matrix as X^*

Where $X^* = \begin{pmatrix} X \\ d\,\underline{\vartheta}'_i \end{pmatrix} \begin{matrix} n \\ 1 \end{matrix}$
k

$X^k : (n+1) \times k$

Here $\underline{\vartheta}'_i$ is the eigen vector corresponding to eigen value λ_i

It can be shown that V $(\hat{\beta}_i)$ is given by

$$V(\hat{\beta}_i) = \sigma^2 \left[\frac{\vartheta_{i1}^2}{\lambda_1} + \frac{\vartheta_{i2}^2}{\lambda_2} + \cdots + \frac{\vartheta_{ii}^2}{(\lambda_i + d^2)} + \cdots + \frac{\vartheta_{ik}^2}{\lambda_k}\right]$$

Where $\vartheta_{i1}, \vartheta_{i2}, \ldots, \vartheta_{ik}$ are the elements of $\underline{\vartheta_i}$. By proper choice of the constant d, we can reduce the i^{th} component in the above formula and thus V $(\hat{\beta}_i)$ can be reduced. Choice of constant d can be obtained by considering many advanced approaches.

8 By using Ridge Regression

This is rather a mechanical and purely numerical method. Ridge regression estimator $\underline{\hat{\beta}_R}$ for $\underline{\beta}$ is defined as

$$\underline{\hat{\beta}_R} = (X'X + CI)^{-1} X'\underline{y}$$

Where $C > 0$ is an arbitrary constant

Here $E\left(\underline{\hat{\beta}_R}\right) = (X'X + CI)^{-1} X'X\underline{\beta}$

This shows that Ridge Regression Estimate (RRE) is not unbiased for $\underline{\beta}$

However $V\left(\underline{\hat{\beta}_R}\right) = \sigma^2[(X'X + CI)^{-1} X'X(X'X + CI)^{-1}]$

From which it can be shown that the variances of RRE Estimators are less than OLS estimators. Hence this method can be beneficial to resolve the problem of multicollinearity.

Then question is how to select constant C. one approach for choosing C to run Ridge Regression is to find C such that $E\left[\left(\underline{\hat{\beta}}_R - \underline{\beta}\right)' \left(\underline{\hat{\beta}}_R - \underline{\beta}\right)\right]$ is minimum. that is choose constant C such that the sum of MSE's of the ridge estimators is minimum. Ridge technique essentially consists of an arbitrary numerical adjustment to the sample data to tackle the problem of multicollinearity.

From this we have Generalized Ridge Regression Estimator (GRRE) of $\underline{\beta}$ given by

$$\underline{\hat{\beta}}_{GR} = (X'X + AI)^{-1} X'\underline{y}$$

Where A: n×n is an arbitrary matrix then

$$E\left(\underline{\hat{\beta}}_{GR}\right) = (X'X + AI)^{-1} X'X\underline{\beta}$$

Which shows that GRRE is biased estimator for $\underline{\beta}$.

As before $E\left(\hat{\beta}_{GR}\right) = \sigma^2[(X'X + AI)^{-1} X'X(X'X + AI)^{-1}]$

Which can be made smaller by properly choosing matrix A