TOPIC: MULTICOLLINEARITY FOURTH SEMESTER CC410(ECONOMETRICS)

MONICA,
ASSISTANT PROFESSOR,
DEPT. OF ECO.,
PATNA WOMEN'S COLLEGE.

INTRODUCTION

•The Term was coined by Ragnar Frisch.

•Originally ,It denotes the presence of perfect or exact linear relationship among some or all explanatory variables in a regression model.
For example, we may have $\lambda_0 + \lambda_1 X_{1i} + \lambda_2 X_{2i} + \ldots\ldots + \lambda_k X_{ki} = 0$

•When there is an intercorrelation between the explanatory variables, it is difficult to separate the effects attributed by these variables on Y.
For Eg. Assume that consumption expenditure of an individual depends on his income and liquid assets

REASONS FOR THE PROBLEM OF MULTICOLLINEARITY

•Growth and Trend factors in time series are main causes of multicollinearity. For Eg. In periods of Boom or rapid economic growth the basic economic magnitudes grow i.e Income,Consumption,prices,employment,etc.

•Use of Lagged values of some explanatory variables as separate independent factors in the relationship.
For Eg.In Consumption Function, lagged income(Y_{t-1}) is also taken as one of the independent variables i.e $C_t=f(Y_t,Y_{t-1})$

CONSEQUENCES OF MULTICOLLINEARITY

- If intercorrelation between explanatory variable is perfect

a) Estimates of Coefficients are indeterminate.

b) Standard errors of these estimates become infinitely large.

Proof:- $Y=\beta_0+\beta_1 X_1+B_2 X_2+u$

$$\hat{\beta}_1=\frac{(\sum x_1 y)(\sum x_2^2)-(\sum x_2 y)(\sum x_1 x_2)}{(\sum x_1^2)(\sum x_2^2)-(\sum x_1 x_2)^2} \qquad \hat{\beta}_2=\frac{(\sum x_2 y)(\sum x_1^2)-(\sum x_1 y)(\sum x_1 x_2)}{(\sum x_1^2)(\sum x_2^2)-(\sum x_1 x_2)^2}$$

Substituting kX_1 for X_2 ,we obtain:-

$$\hat{\beta}_1=\frac{k^2(\sum x_1 y)(\sum x_1^2)-(\sum x_1 y)(\sum x_1^2)}{k^2(\sum x_1^2)-k^2(\sum x_1^2)^2}=0$$

and $\beta_2= 0$ which shows estimates of coefficients are indeterminate.

Similarly , $Var(\beta_1)=\frac{\sigma^2\sum x_{2i}^2}{(\sum x_{1i}^2)(\sum x_{2i}^2)-(\sum x_{1i}x_{2i})^2}$ and $Var(\beta_2)=\frac{\sigma^2\sum x_{1i}}{(\sum x_{1i}^2)(\sum x_{2i}^2)-(\sum x_{1i}x_{2i})^2}$

Substituting $X_{2i}=\lambda X_{1i}$ then we get $Var(\beta_1)=\infty$ and $Var(\beta_2)=\infty$

TEST FOR DETECTING MULTICOLLINEARITY

Frisch's Confluence Analysis(Bunch Map Analysis)

1. Regress the dependent variable on each one of the explanatory variables separately.
2. We obtain all the elementary regressions and we examine their results on the basis of a priori and statistical criteria.
3. We gradually insert additional variable and we examine their effects on the Individual coefficients , on their standard errors and on the overall R^2

Farrar -Glauber Test

Chi-Square Test

Detection of existence and severity of multicollinearity

F-Test

Locating which variables are multicollinear

T-Test

Finding out pattern of multicollinearity i.e which variables are responsible for appearance of multicollinear variables

REMEDIAL MEASURES

- **A Priori Information** : It is a method of restricted least squares.

For Eg. $Yi=\beta 0+ \beta 1X1i+ \beta 2X2i+ui$

Suppose that , it is a known priori that $\beta 2=0.1\beta 1$.For this model ,if Y=Consumption,X1=Income,X2=Wealth, then identity means that the rate of change of consumption with respect to wealth is one tenth of the corresponding rate with respect to income. So, we get $Y_i=\beta_0+ \beta_1X_{1i}+0.1\beta_1X_{1i}+u_i$; $Yi=\beta_0+ \beta_1X_i+ui$, Where $Xi=X1i+0.1X2i$

- **Combining Cross –Sectional and Time –Series Data**: (pooling data)For Eg. If we wish to study the demand for a particular brand of car in chennai and assume that we have time series data on the number of car sold, average price of car and consumer income. If we estimate using OLS,we have the problem of multicollinearity as the variables will be highly collinear in time series data.So,Tobin suggested we should estimate it using the cross –sectional data

- **Increase of the Size of the Sample:** Now as the sample size increases,$\sum x_2$ will increase. For any given r^2,the variance β_2 will decrease, thus decreasing standard error which enables to estimate β_2 precisely.

- Dropping Variable: The simplest method when we face high multicollinearity is to drop one of the collinear variables from the model. We may be committing a significant specification bias or specification error, which arises from incorrect specification of the model.

THANK YOU !