[Academic Script]

Multiple Regression Model and Extensions

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 – 2 Multiple Regression Model and Extensions

Academic Script

Hello friends nice to meet you. In the first lecture we had introduced some concepts about two variables and three variables models. This models are extremely useful for applications. However in actual life there can be more than 2 or 3 variable. In general there can be many variables. A number of variables which are explanatory can be effective upon the dependent variable. So it is necessary to get extension of our studies. For that purpose we shall now introduced what is known as classical k variate linear regression models. This model is stated with its basic assumptions. Two main basic assumptions of the model are one is Homoscadasticity and other is no multicollinearity. This model will be explained with its properties and applications then you will also know about certain mathematical forms of the model connected with linear regression. This models are Log-Log model, Log-lin model, Lin-Log model, reciprocal model etc. After that you will have a brief study of what is known as bird eye view of econometric methods. This will give you the idea of what we are going to study further. In the words of great philosopher Confusious " When you know, what you have to know, you have already cross the half way of knowledge. Now let us begin our lecture 2.

1. **k Variate General Linear Model**We want to extend now the ideas of two and three variables model for k variables.

Illustration

Agricultural production for a particular crop depends upon certain input factors like – rainfall (or irrigation), fertilizer, pesticides, seedlings for the crop, Labourforce participation, humidity, capital investments etc. This dependency relationship can be expressed by means of a linear model which expresses multiple regression between the concerned variables.

Let Y = Dependent (endogenous) variable (crop products)

X_2 ,X_3 , ... , X_k are the independent explanatory variables (rainfall, fertilizer, etc.) we express the equation for the model as

$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + . \; . . . . + \beta_k X_k +$ U (PRF) . . . . . (1)

Hereβ_1, β_2,β_3, ... , β_k are the k unknown parameters (which are regression coefficients) and U is the disturbance term.

We can write equation (1) as

$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + . \; . . . . + \hat{\beta}_k X_{ki} + e_i$ (SRF) . ... . (2)

Here $Y_i = i^{th}$ observation of Y

$X_{2i} = i^{th}$observation of X_2

⋮

$X_{ki} = i^{th}$observation of X_k

and $e_i = i^{th}$ error term (i = 1, 2,…, n)

$\hat{\beta}_1$, $\hat{\beta}_2$, ... , $\hat{\beta}_k$ denote the estimated values of the population parameters$\beta_1, \beta_2, \beta_3$, ... , β_k.

The expression given in (2) above is a system of n homogenous equations in which we know all n values of Y and X_2 ,X_3 , ... , X_k but beta coefficients are unknown.

2. Matrix Form

Matrix presentation for equations in (1) is as under

$\underline{Y} = X\underline{\beta} + \underline{U}$ (PRF) . . . . (3)

Where $\underline{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}_{n\times 1}$,

$$\underline{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}_{k\times 1},$$

$$\underline{U} = \begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_k \end{pmatrix}_{n\times 1}$$

$$X = \begin{pmatrix} 1 & X_{21} & X_{31} & \cdots & X_{k1} \\ 1 & X_{22} & X_{32} & \dots & X_{k2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{2n} & X_{3n} & \cdots & X_{kn} \end{pmatrix}_{n\times k}$$

Correspondingly equations in (2) will become

$\underline{Y} = X\underline{\hat{\beta}} + \underline{e}$ (SRF) . . . . (5)

Where $\underline{\hat{\beta}} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{pmatrix}_{k\times 1}$,

$\underline{e} = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}_{n\times 1}$ . . . . (6)

The above model is called k variate General Linear Model under the following **basic assumptions**

(1) $E(\underline{U}) = \underline{0} \Rightarrow E(U_i) = 0$ for all i = 1,2, ... ,n

(2) $V(\underline{U}) = \sigma^2 I_n \Rightarrow V(U_i) = \sigma^2$ for all i = 1,2, ... , n

and COV $(U_i, U_j) = 0$ for all i = 1,2, ... , n and j = 1,2, ... , n (i $\neq$j)

This assumption about uniformly of variance is called <u>Homoscadasticity.</u>

(3) Data matrix X is fixed and non−stochastic (This assumption states that there is no distributional pattern in the variables X_2 ,X_3 , ... , X_k .

(4) Rank of matrix $X = \rho(X)$ = k < n

Hence $\rho(X'X) = \rho(X)$ = k $\Rightarrow |X'X| \neq 0$

And thus columns of X are not linearly dependent.

$(X' X)^{-1}$ exists as $|X'X| \neq 0$.

(This assumption states that there is no multicollinearity among X variables).

Once when we estimate this model, then prediction about Y's on the basis of given (known) values of X's can be obtained by means of $\underline{\hat{Y}} = X \underline{\hat{\beta}}$

Problem now is how to estimate Beta Coefficients.

This is done by least squares method.

3. OLS Estimation

We use the least squares principle to find estimates of Beta Coefficients.

Thus minimize $\underline{e}' \underline{e} = \sum_{i=1}^{n} e_i^2$ to obtain unknown Betas, so that we get

$\underline{\hat{\beta}} = (X' X)^{-1} X' \underline{Y}$

and V $(\underline{\hat{\beta}}) = (X' X)^{-1} \sigma^2$.

$\underline{\hat{\beta}}$ Obtained above are called ordinary least squares estimators. [OLSE]

By **Gauss – Markov** theorem $\underline{\hat{\beta}}$ is BLUE for $\underline{\beta}$.

(BLUE → Best Linear Unbiased Estimator

Linear as it is linear function of observations on Y, unbiased because E $(\underline{\hat{\beta}}) = \underline{\beta}$ and best because V $(\underline{\hat{\beta}})$ is minimum).To compute V $(\underline{\hat{\beta}})$ we need $\hat{\sigma}^2$.

Unbiased estimator of σ^2 is given by

$$\hat{\sigma}^2 = \underline{e}' \underline{e} / (n-k) = \sum_{i=1}^{n} e_i^2 / (n-k)$$

$[Note\ that\ matrix\ M = [I_n - (X' X)^{-1} X']_{n \times n}\ is\ an\ Idempotent\ matrix\ with\ its\ rank = \rho(M) = (n-k)]$

4. MLE of $\underline{\beta}$ and σ^2

We would also like to find out Maximum Likelihood Estimators for $\underline{\beta}$ and σ^2.

This will be based upon normality assumptions.

Thus if $\underline{U} \sim I_N\,(\underline{0}, \sigma^2 I_n)$

(Means each U_i is independently normally distributed with mean zero and variance σ^2)

Then MLE of $\underline{\beta}$ and σ^2 are obtained by

MLE of $\underline{\beta}$ = $\underline{\tilde{\beta}}$ = $(X'\,X)^{-1}X'\underline{Y}$ = $\underline{\hat{\beta}}$ = OLSE of $\underline{\beta}$

Thus MLE and OLSE of $\underline{\beta}$ are the same.

MLE of $\sigma^2 = \tilde{\sigma}^2 = {\underline{e}'\,\underline{e}}/{n} = {\sum_{i=1}^{n} e_i^{\,2}}/{n}$

Now OLSE of $\sigma^2 = \hat{\sigma}^2 = {\underline{e}'\,\underline{e}}/{(n-k)}$

So that (n-k) $\hat{\sigma}^2 = \underline{e}'\,\underline{e}$ = n $\tilde{\sigma}^2$

Hence $\tilde{\sigma}^2 = \left(\frac{n-k}{n}\right)\hat{\sigma}^2$

i.e. MLE of $\sigma^2 = \left(\frac{n-k}{n}\right)$OLSE of σ^2

However for large n, $\tilde{\sigma}^2 \equiv \hat{\sigma}^2$

(They are identical for large n)

Note Under normality assumptions, that is

$\underline{U} \sim I_N\,(\underline{0}, \sigma^2 I_n)$ it can be easily verified that $\underline{\hat{\beta}} \sim N\left[\underline{\beta}, (X'\,X)^{-1}\,\sigma^2\right]$

5. Coefficient of Determination R^2

As in the earlier case, here also we can define Multiple Coefficient of Determination denoted by R^2 by the following formula

$$R^2 = \frac{Explained\ sum\ of\ squares}{Total\ sum\ of\ squares} = \frac{ESS}{TSS}$$

Extending the earlier formula, we have

$$R^2 = \frac{\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i} + \cdots + \hat{\beta}_k \sum y_i x_{ki}}{\sum y_i^2}$$

(Where the variables are measured from their means)

This relation in matrix form, becomes

$$R^2 = \frac{\underline{\hat{\beta}}' X' \underline{Y} - n\bar{Y}^2}{\underline{y}' \underline{y} - n\bar{Y}^2}$$

(Here R^2 is also shown by symbol $R^2{}_{1\cdot 23..k}$)

6. Testing Problems

We have correlation matrix P representing pair wise ordinary or zero-order correlation coefficients

$$P = \begin{pmatrix} 1 & r_{12} & r_{13} & \cdots & r_{1k} \\ r_{21} & 1 & r_{23} & \cdots & r_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ r_{k1} & r_{k2} & r_{k3} & \cdots & 1 \end{pmatrix}_{k \times k}$$

r_{ij} = correlation coefficient between X_i and X_j (i, j = 1, 2 . . . k)

And similar results exist for partial correlation coefficients in terms of above correlation coefficients.

(I) Testing for Individual Beta Coefficients

To test $H_0 : \beta_i = \beta_{i0}$

Vs $H_1 : \beta_i \neq \beta_{i_0}$

Here β_{i_0} is some specified value of β_i

We have t = $\frac{\hat{\beta}_i - \beta_{i_0}}{\sqrt{V(\hat{\beta}_i)}}$ which has students t distribution with (n-k) degrees of freedom.

Hence test can be carried out as usual and confidence intervals can also be obtained.

(II) Testing for Overall Significance of Regression

We can have ANOVA as under

ANOVA Table				
Source	**D.F**	**S.S**	**M.S.S**	**F**
Due to regression	$k-1$	$\underline{\hat{\beta}}'\, X'\underline{y} - n\bar{Y}^2 = \text{A}$	$\frac{A}{K-1} = A'$	$\frac{\text{A}'}{\text{B}'}$
Due to residuals	n−k	$\underline{y}'\,\underline{y} - \underline{\hat{\beta}}'\, X'\underline{y} = \text{B}$	$\frac{B}{n-k} = B'$	−
Total	n−1	$\underline{y}'\,\underline{y} - n\bar{Y}^2$	−	−

Thus we test the hypothesis,

$H_0 : \beta_2 = \beta_3 = \ldots = \beta_k = 0$ (i.e. All Betas Zero)

Vs $H_1 : \beta_2 \neq \beta_3 \neq \ldots \neq \beta_k \neq 0$ (i.e. All Betas non-Zero)

As above we have $F = \frac{\text{A}'}{\text{B}'} = \frac{\left(\underline{\hat{\beta}}'\, X'\underline{y} - n\bar{Y}^2\right)/(k-1)}{\left(\underline{y}'\,\underline{y} - \underline{\hat{\beta}}'\, X'\underline{y}\right)/(n-k)}$

Which has F distribution with $\upsilon_1 = (k-1)$ and $\upsilon_2 = (n-k)$ d. f.

We expect F to be significant for the Validity of Betas (i.e. Validity of the model)

(III) Testing for $\underline{R^2}$

$H_0 : \rho^2 = 0$ (i.e. $\rho = 0$) $\Rightarrow \rho = 0$

Vs $H_1 : \rho^2 \neq 0$ (i.e. $\rho \neq 0$) $\Rightarrow \rho \neq 0$

(ρ = Population Multiple Correlation Coefficient)

This is also provided by the following ANOVA table

ANOVA table				
Source	**D.F**	**S.S**	**M.S.S**	**F**
Regression	k−1	$R^2\left(\underline{y}'\,\underline{y} - n\bar{Y}^2\right) = \text{A}$	$\frac{A}{K-1} = A'$	$\frac{\text{A}'}{\text{B}'}$
Error	$n-k$	$(1-R^2)\left(\underline{y}'\,\underline{y} - n\bar{Y}^2\right) = \text{B}$	$\frac{B}{n-k} = B'$	−
Total	n−1	$\underline{y}'\,\underline{y} - n\bar{Y}^2$	−	−

Thus $F = \frac{\text{A}'}{\text{B}'} = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}$

Which has F distribution with $\upsilon_1 = (k-1)$ and $\upsilon_2 = (n-k)$ d. f.

We expect F to be significant for the Validity of the model (that is significance of R^2).

Illustration 1

Consider the following model

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

Where Y = Production of onion crop in India

X_2 = Cultivation Area

X_3 = time in years

U = Disturbance term

The estimated model is worked out as under with other details

$$\begin{array}{lllll} Y & = & 2.33 & + 1.02X_2 + 0.04\,X_3 & \\ t & = & (8.74)^{**} & (3.98)^{**}\ (1.61) & \\ R^2 & = & 0.9615 & F = 281.3^{**} & n = 13 \end{array}$$

Interpret the model

Here $\hat{\beta}_1$ = 2.33, $\hat{\beta}_2$ = 1.02, $\hat{\beta}_3$ = 0.04

Per year, (when cultivation area is fixed) there is about 4% increase in crop production.

Also irrespective of the time component, for unit cultivation area there is about 1.02 units increase in production.

The partial regression coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$ are highly significant at 1% level of significance. Coefficient $\hat{\beta}_3$ is not significant. About 96.15% of the variation is explained by the model and F is found to be highly significant, thus claiming for suitability of the model considered.

Illustration 2

Two different forms of regression models are fitted with the given data

<u>Model I</u>

Y $=\beta_1 + \beta_2 X_2 + \beta_3 X_3 + U_1$

Model II

$Y = \beta'_1 + \beta'_2 X_2 + \beta'_3 X_3 + \beta'_4 X_4 + U_2$

For these models, $\hat{R}_1^{\,2} < \hat{R}_2^{\,2}$ (numerically)

If both the models have significant values for R^2, which model would you recommend?

$\hat{R}_1^{\,2}$ And $\hat{R}_2^{\,2}$ are the adjusted values of $R^2$for these models. Since $\hat{R}_1^{\,2} < \hat{R}_2^{\,2}$, we consider second model as better one as compared to the first model.

(Note that here the dependent variable is the same in both the models, however explanatory variables may change.)

Thus decision can be taken on the basis of larger value of adjusted R^2.

7. Functional Forms of Regression Models

As an extension of the linear models, we can consider the functional forms of the models. These are classified as

(1)Long- linear (or Double Log model)

(2) Semi log models.

(3) Reciprocal models.

Semi log models are also classified as log Lin and Lin Log models.

We can consider two variables relationship to illustrate them and can be extended further for more variables.

The following tabular form gives in nutshell the broad categories of the above models.

	Model	Function	Slope $(=\frac{dy}{dx})$	Elasticity = $(\frac{dY}{dX})$ (X/Y)
1	Linear	$Y = \beta_1 + \beta_2 X$	β_2	β_2 (X/Y)
2	Log-Log (Double Log)	$\log Y = \beta_1 + \beta_2 \log X$	β_2 (Y/X)	β_2
3	Log- lin	$\log Y = \beta_1 + \beta_2 X$	$\beta_2\,(Y)$	$\beta_2\,(X)$
4	Lin-log	$Y = \beta_1 + \beta_2 \log X$	β_2 (1/X)	β_2 (1/Y)
5	Reciprocal	$Y = \beta_1 + \beta_2 (1/X)$	$-\beta_2\,(1/X^2)$	$-\beta_2\,(\frac{1}{XY})$

Note that double- log linear model has constant elasticity, whereas all other models have elasticity variable, depending upon the nature of the corresponding observations on the variables concerned.

(We can define new variable for reciprocal or logarithmic value and run regression.

e.g. In model (5) put $Z =$ 1/X,

In model (2), put $\log Y = Z$, $\log X = P$ etc.

Conclusions based upon transformed model are to be viewed on the basis of the original model also)

8. Broad Generalization of Linear Models

(BIRD EYE VIEW for Econometric methods)

We have stated GLM with some specific assumptions.

If these assumptions are violated one by one then we have very interesting cases which represents different features for economic methods for linear models. We express them briefly as under.

(1) $E(\underline{U}) = \underline{0}$

When this assumption is violated we can write $E(U_i) = \mu_i \Rightarrow E(U_i - \mu_i) = 0$

for i = 1, 2... , n

This means $E(\underline{w}) = \underline{0}$ where $\underline{w} = \underline{U} - \underline{\mu}$

It is like change of origin. This can be dealt with in the usual manner, but here there is Specification Problem and Specification Bias which should be considered.

(2) $V(\underline{U}) = \sigma^2 I_n$

This shows that $V(U_i) = \sigma^2$ for $i = 1,2\ldots,n$

If this assumption is violated we may write $V(U_i) = \sigma_i^2$ (i = 1, 2..., n)

And $COV(U_i, U_j) \neq 0$ i, j = 1,2,...,n (i≠j)

Thus variances are not uniform and any pair of covariance is non-zero.

This is the case of Heteroscadasticity.

Here instead of OLS estimators, Generalized Least Square Estimators are used.

For this case, in particular we can use Weighted Least Squares estimators. Under certain stated assumptions we have Two stage Least Squares methods for estimation.

(3) Autocorrelation

In particular, under the heteroscadastic situations if the disturbance terms have a first order autoregressive relation given by the following

$U_i = \rho\, U_{i-1} + \epsilon_i$ (i = 1, 2..., n)

Where ρ = Autocorrelation coefficient.

ϵ_i are the stochastic error terms with assumptions similar to classical model.

Here the explanatory variables may be related with disturbances i.e. $COV(X, \underline{U}) \neq 0$

These phenomena of autocorrelation can be detected by means of Durbin Watson (DW) test and resolved by means of using methods like Cochran Orcutt Iterative procedure.

(4) $\rho(X) = k < n$

If we assume that $\rho(X) < k$ then the assumption$(X'X)$ to be non-singular is violated. Here $|X'X|$ may be equal to or very near to zero. In this case, the columns of X are linearly related. This is called Multicollinearity Problem which can be dealt with by means of different approaches.

(5) Stochastic Process

If Data matrix X is Stochastic instead of non-stochastic, the variables in X have some distributional aspects. In this case we have the problem of Stochastic Regressors and some methods like Instrumental variables etc. can be used.

(6) Lag model

We may have lag in time relationship like

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3(i-1)} + U_i$$

i.e. one (or more) explanatory variables may be related in terms of the earlier period values. This is lag structure in the models. Special methods are needed to resolve these problems.

(7) Measurement Errors

Actual measurements in variables contain some error of measurement. This can be dealt with by special techniques.

(8) Simultaneous Equations System Models

Here in place of a single matrix equation we have a system of several linear equations and special feature about this system is that a particular variable in one equation which is explanatory variable may occur as an endogenous variable in other equation and so on.

This system of simultaneous equations needs separate approach. First is the problem of Identification. The solution procedure uses the methods like Indirect Least Squares, Two Stage Least Squares etc.

(9) Non-normality assumptions

Instead of using the normality assumption for disturbances, we may have the situations where other than normal distribution holds good for disturbances. This needs special specific treatment. In some situations, the problem can be resolved by taking large samples.

(10) Dummy Variables

If the explanatory variables are qualitative in nature then we can replace them by giving numerical values. This is Dummy Variables approach. There are very interesting applications in this case.

In particular, when the dependent variable is qualitative then we have Linear Probability Model (LPM) tackling the situation.

A specific approach here is by using Logit and Probit techniques.

(11) Time Series Models

We have forecasting for time series using AR, MA, ARIM, VAR models etc.