UNIT 8 MULTIPLE LINEAR REGRESSION MODEL: INFERENCES*

Structure

8.0 OBJECTIVES

After going through this unit, you should be able to

- explain the need for the assumption of normality in the case of multiple regression;
- describe the procedure of testing of hypothesis on individual estimators;
- test the overall significance of a regression model;
- test for the equality of two regression coefficients;
- explain the procedure of applying the Chow test;
- make prediction on the basis of multiple regression model;
- interpret the results obtained from the testing of hypothesis, both individual and joint; and
- apply various tests such as likelihood ratio (LR), Wald (W) and Lagrange Multiplier Test (LM).

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

8.1 INTRODUCTION

In the previous unit we discussed about the interpretation and estimation of multiple regression models. We looked at the assumptions that are required for the ordinary least squares (OLS) and maximum likelihood (ML) estimation. In the present Unit we look at the methods of hypothesis testing in multiple regression models.

Recall that in Unit 3 of this course we mentioned the procedure of hypothesis testing. Further, in Unit 5 we explained the procedure of hypothesis testing in the case of two variable regression models. Now let us extend the procedure of hypothesis testing to multiple regression models. There could be two scenarios in multiple regression models so far as hypothesis testing is concerned: (i) testing of individual coefficients, and (ii) joint testing of some of the parameters. We discuss the method of testing for structural stability of regression model by applying the Chow test. Further, we discuss three important tests, viz., Likelihood Ratio test, Wald test, and Lagrange Multiplier test. Finally, we deal with the issue of prediction on the basis of multiple regression equation.

One of the assumptions in hypothesis testing is that the error variable u_i follows normal distribution. Is there a method to test for the normality of a variable? We will discuss this issue also. However, let us begin with an overview of the basic assumptions of multiple regression models.

8.2 ASSUMPTIONS OF MULTIPLE REGRESSION MODELS

In Unit 7 we considered the multiple regression model with two explanatory variables X_2 and X_3. The stochastic error term is u_i.

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \quad \text{… (8.1)}$$

8.2.1 Classical Assumptions

There are seven assumptions regarding the multiple regression model. Most of these assumptions are regarding the error term. We discussed about these assumptions in the previous Unit. Let us briefly mention those assumptions again.

a) The regression model is linear in parameters and variables.

b) The mean of error terms is zero. In other words, the expected value of error term conditional upon the explanatory variables X_{2i} and X_{3i} is zero.

   $E(u_i)= 0$ or $E(u_i|X_{2i}, X_{3i}) = 0$

c) There is no serial correlation (or autocorrelation) among the error terms. The error terms are not correlated. It implies that the covariance between the error term associated with ith observation u_i and the error term associated with jth observation, u_j is zero.

$$\text{cov}(u_i, u_j) = 0$$

d) Homoscedasticity: The assumption of homoscedasticity states that the error variance is constant throughout the population. The variance of the error term associated at each observation has the same variance.

$$var(u_i) = \sigma^2$$

e) Exogeneity of explanatory variables: There is no correlation between the explanatory variables and the error term. This assumption is also called exogeneity, because the explanatory variables are assumed to be exogenous (given from outside; X is not determined inside the model). In contrast, Y is determined within the model. When the explanatory variable is correlated with the error term, it is called endogeneity problem. In order to avoid this problem, we assume that the explanatory variables are kept fixed across samples.

f) Independent variables are not linear combination of one another. If there is perfect linear relationship among the independent variables, the explanatory variables move in harmony and it is not possible to estimate the parameters. It is also called multicollinearity problem.

g) The error variable is normally distributed. This assumption is not necessary in OLS method for estimation of parameters. It is required for construction of confidence interval and hypothesis testing. In the maximum likelihood method discussed in the previous Unit, in order to estimate the parameters we assumed that the error term follows normal distribution.

8.2.2 Test for Normality of the Error Term

As pointed out earlier, we look into the assumption of normality of the error term. In order to test for normality of the error term we apply the Jarque-Bera test (often called the JB test). It is an asymptotic or large sample test. We do not know the error terms in a regression model; we know the residuals. Therefore, the JB test is based on the OLS residuals. Recall two concepts from statistics: skewness and kurtosis. A skewed curve (i.e., asymmetric) is different from a normal curve. A leptokurtic or platykurtic curve (i.e., tall or short in height) is different from a normal curve. The JB test utilises the measures of skewness and kurtosis.

We know that for a normal distribution S = 0 and K = 3. A signification deviation from these two values will confirm that the variable is not normally distributed.

Jarque and Bera constructed the J-statistic given by

$$JB = \frac{n}{6}\left[S^2 + \frac{(K-3)^2}{4}\right] \quad \ldots (8.2)$$

where

n = sample size

S = measure of skewness ($\frac{\mu_3}{\sigma^3}$)

K = measure of kurtosis ($\frac{\mu_4}{\mu_2^2}$)

Skewness and kurtosis are measured in terms of the moments of a variable. As you know from BECC 107, Unit 4, the formula for calculating the r[th] moment of variable X_i is

$$\mu_r = \frac{1}{n}\sum_{i-1}^{n} f_i (X_i - \bar{X})^r \qquad \ldots (8.4)$$

Variance is the second moment μ_2.

In equation (8.2) the JB statistic follows chi-square distribution with 2 degrees of freedom, $\sim\chi^2_{(2)}$.

Let us find out the value of the JB statistic if a variable follows normal distribution. For the normal distribution, as mentioned above S = 0 and K = 3. By asubstitutingthese values in equation (8.2) we obtain

$$JB = \frac{n}{6}[0 + 0] = \frac{n}{6} \times 0 = 0 \qquad \ldots (8.3)$$

For a variable not normally distributed JB statistics will assume increasingly large values. The null hypothesis is

H_0: The random variable follows normal distribution.

We draw inferences from the JB statistic as follows:

a) If the calculated value of JB statistic is greater than the tabulated value of χ^2 for 2 degrees of freedom, we reject the null hypothesis. We infer that the random variable is not normally distributed.

b) If the calculated value of the JB statistic is less than the tabulated value of χ^2 for 2 degrees of freedom, we do not reject the null hypothesis. We infer that the random variable is normally distributed.

Check Your Progress 1

1) List the assumptions of multiple regression models.

....

....

....

....

2) State the Jarque-Bera test for normality.

....

....

....

....

8.3 TESTING OF SINGLE PARAMETER

The population regression function is not known to us. We estimate the parameters on the basis of sample data. Since we do not know the error variance σ^2, we should apply *t*-test instead of *z*-test (based on normal distribution).

Let us consider the population regression line given at equation (8.1).

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

The sample regression line estimated by ordinary least squares (OLS) method is

$$\hat{Y}_i = b_1 + b_2 X_{2i} + b_3 X_{3i} \qquad \ldots (8.4)$$

where b_1, b_2 and b_3 are estimators of β_1, β_2 and β_3 respectively. The estimator of error variance σ^2 is given by $\hat{\sigma}^2 = \frac{RSS}{n-k}$.

There are two approaches to hypothesis testing: (i) test of significance approach, and (ii) confidence interval approach. We discuss both the approaches below.

8.3.1 Test of Significance Approach

In this approach we proceed as follows:

(i) Take the point estimate of the parameter that we want test, viz., b_1, or b_2 or b_3.

(ii) Set the null hypothesis. Suppose we expect that variable X_2 has no influence on *Y*. It implies that β_2 should be zero. Thus, null hypothesis is $H_0: \beta_2 = 0$. In this case what should be alternative hypothesis? The alternative hypothesis is $H_A: \beta_2 \neq 0$.

(iii) If $\beta_2 \neq 0$, then β_2 could be either positive or negative. Thus we have to apply two-tail test. Accordingly, the critical value of the t-ratio has to be decided.

(iv) Let us consider another scenario. Suppose we expect that β_3 should be positive. It implies that our null hypothesis is $H_0: \beta_3 > 0$. The alternative hypothesis is $H_A: \beta_3 \leq 0$.

(v) If $\beta_3 > 0$, then β_3 could be either zero or negative. Thus the critical region or rejection region lies on one side of the *t* probability curve. Therefore, we have to apply one-tail test. Accordingly the critical value of t-ratio is to be decided.

(vi) Remember that the null hypothesis depends on economic theory or logic. Therefore, you have to set the null hypothesis according to some logic. If you expect that the explanatory variable should have no effect on the dependent variable, then set the parameter as zero in the null hypothesis.

(vii) Decide on the level of significance. It represents extent of error you want to tolerate. If the level of significance is 5 per cent ($\alpha = 0.05$), your decision on the null hypothesis will go be wrong 5 per cent times. If you take 1 per cent level of significance (α = 0.01), then your decision on the null hypothesis will be wrong 1 per cent times (i.e., it will be correct 99 per cent times).

(viii) Compute the t-ratio. Here the standard error is the positive square root of the variance of the estimator. The formula for the variance of the OLS estimators in multiple regression models is given in Unit 7.

$$t = \frac{b_2 - \beta_2}{se(b_2)} \quad \ldots (8.5)$$

(ix) Compare the computed value of the t-ratio with the tabulated value of the t-ratio. Be careful about the two issues while reading the t-table: (i) level of significance, and (ii) degree of freedom. Level of significance we have mentioned above. Degree of freedom is (n–k), as you know from the previous Unit.

(x) If the computed value of t-ratio is greater than the tabulated value of t-ratio, reject the null hypothesis. If computed value of t-ratio is less than the tabulated value of t-ratio, do not reject the null hypothesis and accept the alternative null hypothesis.

8.3.2 Confidence Interval Approach

We have discussed about interval estimation in Unit 3 and Unit 5. Thus, here we bring out the essential points only.

(i) Remember that confidence interval (CI) is created individually for each parameter. There cannot be a single confidence interval for a group of parameters.

(ii) Confidence interval is build on the basis of the logic described above in the test of significance approach.

(iii) Suppose we have the null hypothesis $H_0: \beta_2 = 0$ and the alternative hypothesis is $H_A: \beta_2 \neq 0$. The estimator of β_2 is b_2. We know the standard error of b_2.

(iv) Here also we decide on the level of significance (α). We refer to the t-table and find out the t-ratio for desired level of significance.

(v) The degree of freedom is known to us, i.e., (n–k).

(vi) Since the above is case of two-tailed test, we take $\alpha/2$ on each side of the t probability curve. Therefore, we take the t-ratio corresponding to the probability $\alpha/2$ and the degrees of freedom applicable.

(vii) Remember that confidence interval is created with the help of the estimator and its standard error. We test whether the parameter lies within the confidence interval or not.

(viii) Construct the confidence interval as follows:

$$\left[b_2 - t_{\alpha/2}SE(b_2) \leq \beta_2 \leq b_2 + t_{\alpha/2}SE(b_2)\right]. \quad \ldots (8.6)$$

(ix) The probability of the parameter remaining in the confidence interval is $(1-\alpha)$. If we have taken the confidence interval as 5 per cent, then the probability that β_2 will remain in the confidence interval is 95 per cent.

$$P_r\left[b_2 - t_{\alpha/2}SE(b_2) \leq \beta_2 \leq b_2 + t_{\alpha/2}SE(b_2)\right] = (1-\alpha) \ldots (8.7)$$

(x) If the parameter (in this case, β_2) remains in the confidence interval, do not reject the null hypothesis.

(xi) If the parameter does not remain within the confidence interval, reject the null hypothesis, and accept the alternative null hypothesis.

Check Your Progress 2

1) Describe the steps you would follow in testing the hypothesis that $\beta_2 < 0$.

....

....

....

....

....

2) Create a confidence interval for the population parameter of the partial slope coefficient.

....

....

....

....

....

8.4 TEST OF OVERALL SIGNIFICANCE

The overall test of significance of a multiple regression model is carried out by applying *F*-test. We have discussed about the *F*-test in Unit 5 of this course in the context of two variable models. For testing of the overall significance of a multiple regression model we proceed as follows:

(i) Set the null hypothesis. The null hypothesis for testing the overall significance of a multiple regression model is given as follows:

$$H_0: \beta_2 = \beta_3 = \ldots \beta_k = 0 \quad \ldots (8.8)$$

(ii) Set the corresponding alternative hypothesis.

$$H_A: \beta_2 = \ldots = \beta_k \neq 0 \quad \ldots (8.9)$$

(iii) Decide on the level of significance. It has the same connotation as in the case of *t*-test described above.

(iv) For multiple regression model the *F*-statistic is given by

$$F = \frac{ESS/(k-1)}{RSS(n-k)} \quad \dots (8.10)$$

(v) Find out the degrees of freedom. The *F*-statistic mentioned in equation (8.10) follows *F* distribution with degrees of freedom (k–1, n–k).

(vi) Find out the computed value of *F* on the basis of equation (8.10). Compare it with the tabulated value of *F* (given at the end of the book). Read the tabulated *F* value for desired level of significance and applicable degrees of freedom.

(vii) If the computed value of *F* is greater than the tabulated value, then reject the null hypothesis.

(viii) If the computed value is less than the tabulated value, do not reject the null hypothesis.

8.5 TESTOF EQUALITY BETWEEN TWO PARAMETERS

We can compare between the parameters of a multiple regression model. Particularly, we can test whether two parameters are equal in a regression model. For this purpose we apply the same procedure as we have learnt in the course BECC 107.

Let us take the following regression model:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_i \quad \dots (8.11)$$

Recall that we do not know the variance of the parameters. Thus, for comparison of the parameters we apply the *t*-test. Secondly, we do not the parameters. Therefore, we take their OLS estimators for comparison purposes.

Our null hypothesis and alternative hypothesis are as follows:

$$H_0: \beta_3 = \beta_4 \quad \text{or} \quad (\beta_3 - \beta_4) = 0 \quad \dots (8.12)$$

$$H_1: \beta_3 \neq \beta_4 \quad \text{or} \quad (\beta_3 - \beta_4) \neq 0 \quad \dots (8.13)$$

For testing of the above hypothesis, the t -statistic is given as follows:

$$t = \frac{(b_3 - b_4) - (\beta_3 - \beta_4)}{SE(b_3 - b_4)} \quad \dots (8.14)$$

The above follows *t*-distribution with (n – k) degrees of freedom.

Since $\beta_3 = \beta_4$ under the null hypothesis, we can re-arrange equation (8.14) as follows:

$$t = \frac{b_3 - b_4}{\sqrt{V(b_3) + V(b_4) - 2\text{cov}(b_3, b_4)}} \quad \dots (8.15)$$

The computed value of t-statistic is obtained by equation (8.15). We compare the computed value of *t*-ratio with the tabulated value of *t*-ratio. We read the *t*-table for desired level of significance and applicable degrees of freedom.

If the computed value of *t*-ratio is greater than the tabulated value, then we reject the null hypothesis. If the computed value of *t*-ratio is less than the tabulated value, then we do not reject the null hypothesis and accept the alternative hypothesis.

We need to interpret our results. If we reject the null hypothesis we conclude that the partial slope coefficients β_3 and β_4 are statistically significantly different. If we do not reject the null hypothesis, we conclude that there is no statistically significant difference between the slope coefficients β_3 and β_4.

Check Your Progress 3

1) Mention the steps of carrying out a test of the overall significance a multiple regression model.

....

....

....

....

....

2) State how the equality between two parameters can be tested.

....

....

....

....

....

8.6 TEST OF LINEAR RESTRICTIONS ON PARAMETERS

Many times we come across situations where we have to test for linear restrictions on parameters. For example, let us consider the Cobb-Douglas production function.

$$Y_i = \beta_1 X_{2i}^{\beta_2} X_{3i}^{\beta_3} e^{u_i} \quad \text{… (8.16)}$$

where Y_i is output, X_{2i} is capital and X_{3i} is labour. The parameters are β_2 and β_3 . The stochastic error term is u_i. The subscript 'i' indicates the i^{th} observation. The Cobb-Douglas production function exhibits constant returns to scale if the parameters fulfil the following condition:

$$\beta_2 + \beta_3 = 1 \quad \text{… (8.17)}$$

As we have discussed in Unit 6, by taking natural log, the Cobb-Douglas production function can be expressed in linear form as

$$ln\, Y_i = ln\, \beta_1 + \beta_2 lnX_{2i} + \beta_3 lnX_{3i} + u_i \quad \ldots (8.18)$$

Suppose have collected data on a sample of firms; our sample size is *n*. The production function is Cobb-Douglas as given above. We want to test whether the production function exhibits constant returns to scale. For this purpose we need to apply the F-test. We can follow two approaches as discussed below.

8.6.1 The t-Test Approach

We will discuss two procedures for testing the hypothesis.

(a) For the In this case our null hypothesis and alternative hypothesis are as follows:

$$H_0: \beta_2 + \beta_3 = 1 \quad \ldots (8.19)$$

$$H_A: \beta_2 + \beta_3 \neq 1 \quad \ldots (8.20)$$

For testing of the above hypothesis, the *t* -statistic is given as follows:

$$t = \frac{(b_2+b_3)-(\beta_2+\beta_3)}{SE(b_2+b_3)} \quad \ldots (8.21)$$

The above follows *t*-distribution with (n – k) degrees of freedom.

We can re-arrange equation (8.21) as follows:

$$\mathrm{t} = \frac{b_2+b_3-1}{\sqrt{V(b_2)+V(b_3)+2\mathrm{cov}(b_2,b_3)}} \quad \ldots (8.22)$$

The computed value of t-statistic is obtained by equation (8.22). We compare the computed value of *t*-ratio with the tabulated value of *t*-ratio. We read the *t*-table for desired level of significance and applicable degrees of freedom.

If the computed value of *t*-ratio is greater than the tabulated value, then we reject the null hypothesis. If the computed value of *t*-ratio is less than the tabulated value, then we do not reject the null hypothesis and accept the alternative hypothesis.

We need to interpret our results. If we reject the null hypothesis we conclude that the firms do not exhibit constant returns to scale. If we do not reject the null hypothesis, we conclude that the firms exhibit constant returns to scale.

(b) Let us look again at the null hypothesis given at (8.19).

$$H_0: \beta_2 + \beta_3 = 1$$

If the above restriction holds, then we should have

$$\beta_2 = (1 - \beta_3)$$

Let us substitute the above relationship in the Cobb-Douglas production function

$$ln\, Y_i = ln\, \beta_1 + (1 - \beta_3) lnX_{2i} + \beta_3 lnX_{3i} + u_i \quad \ldots (8.23)$$

We can re-arrange terms in equation (8.23) to obtain

$$ln\, Y_i - lnX_{2i} = ln\, \beta_1 - \beta_3 lnX_{2i} + \beta_3 lnX_{3i} + u_i$$

Or,

$$ln(Y_i/X_{2i}) = \beta_0 + \beta_3 ln(X_{3i}/X_{2i}) + u_i \qquad \dots (8.24)$$

Note that the dependent variable in the above regression model is output-labour ratio and the explanatory variable is capital-labour ratio. We can estimate the regression model given at equation (8.24) and find the OLS estimator of β_3.

If $\beta_3 = 1$, then the Cobb-Douglas production will exhibit constant returns to scale.

Therefore, we set the null hypothesis and alternative hypothesis as

$H_0: \beta_3 = 1$ and $H_A: \beta_3 \neq 1$

We apply t-test for individual parameters as mentioned in sub-section 8.3.1. If the null hypothesis is rejected we conclude that the firms do not exhibit constant returns to scale.

8.6.2 Restricted Least Squares

The t-test approach mentioned above may not be suitable in all cases. There may be situations where we have more than two parameters to be tested. In such circumstances we apply the *F*-test. This approach is called the restricted least squares.

Let us consider the multiple regression model given at equation (8.11).

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_i$$

Suppose we have to test the hypothesis that X_3 and X_4 do not influence the dependent variable Y. In such a case, the parameters β_3 and β_4 should be zero.

Recall that if we increase the number of explanatory variables in a regression model, there is an increase R^2. Recall further that $R^2 = \frac{ESS}{TSS}$. Thus, if two of the explanatory variables in equation (8.11) are dropped (i.e., their coefficients are zero), there will be a decrease in the value R^2. If the variables that X_3 and X_4 are relevant, there will be a significant decline the value of R^2. On the other hand, if the variables X_3 and X_4 are not relevant for the regression model, then the decline in the value of R^2 will be insignificant. We use this property of the regression model to test hypotheses on a group of parameters. Therefore, while applying F-test in restricted least squares we estimated the regression model twice: (i) the unrestricted model, and (ii) the restricted model.

We proceed as follows:

(i) Suppose there are *k* explanatory variables in the regression model.

(ii) Out of these *k* explanatory variables, suppose the first *m* explanatory variables are not relevant.

(iii) Thus our null hypothesis will be as follows:

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_m = 0 \qquad \ldots (8.25)$$

(iv) The corresponding alternative hypothesis will be that the βs are not zero.

(v) Estimate the unrestricted regression model given at (8.11). Obtain the residual sum of squares (RRS) on the basis of the estimated regression equation. Denote it as RSS_{UR}.

(vi) Estimate the restricted regression model by excluding the explanatory variables for which the parameters are zero. Obtain the residual sum of squares (RRS) from this restricted model. Denote it as RSS_R.

(vii) Our F-statistic is

$$F = \frac{RSS_R - RSS_{UR}/m}{RS\ \ _{UR}/(n-k)} \qquad \ldots (8.26)$$

The F-statistic at (8.26) follows F-distribution with degrees of freedom (m, n–k).

(ix) Find out the computed value of F on the basis of equation (8.10). Compare it with the tabulated value of F (given at the end of the book). Read the tabulated F value for desired level of significance and applicable degrees of freedom.

(x) If the computed value of F is greater than the tabulated value, then reject the null hypothesis.

(xi) If the computed value is less than the tabulated value, do not reject the null hypothesis.

As mentioned earlier, the residual sum of squares (RSS) and the coefficient of determination (R^2) are related. Therefore, it is possible to carry out the F-test on the basis of R^2 also. If we have the coefficient of determination for the unrestricted model (R^2_{UR}) and the coefficient of determination for the restricted model (R^2_R), then we can test the joint hypothesis about the set of parameters.

The F-statistic will be

$$F = \frac{R^2_{UR} - R^2_R/m}{(1-R^2_{UR})/(n-k)} \qquad \ldots (8.27)$$

which follows F-distribution with degrees of freedom (m, n–k).

The conclusion to be drawn and interpretation of results will be the same as described in points (x) and (xi) above.

8.7 STRUCTURAL STABILITY OF A MODEL: CHOW TEST

Many times we come across situations where there is a change in the pattern of data. The dependent and independent variables may not remain the same throughout the sample. For example, saving behaviour of poor and rich households may be different. The production of an industry may be different after a policy change. In such situations it may not be appropriate to run a single regression for the entire dataset. There is a need to check for structural stability of the econometric model.

There are various procedures to bring in structural breaks in a regression model. We will discuss about the dummy variable cases in unit 9. In this Unit we discuss a very simple and specific case.

Suppose we have data on n observations. We suspect that the first n_1 observations are different from the remaining n_2 observations (we have $n_1 + n_2 = n$). In this case run the following three regression equations:

$Y_t = \lambda_1 + \lambda_2 X_t + u_t$ (number of observations: n_1) … (8.28)

$Y_t = r_1 + r_2 X_t + v_t$ (number of observations: n_2) … (8.29)

$Y_t = \alpha_1 + \alpha_2 X_t + w_t$ (number of observations: $n = n_1 + n_2$) … (8.30)

If both the sub-samples are the same, then we should have $\lambda_1 = r_1 = \alpha_1$

and $\lambda_2 = r_2 = \alpha_2$. If both the sub-samples are different then there will be a structural break in the sample. It implies the parameters of equations (8.28) and (8.29) are different. In order to test for the structural stability of the regression model we apply Chow test.

We process as follows:

(i) Run the regression model (8.28). Obtain residual sum of squares RSS_1.

(ii) Run regression model (8.29). Obtain residual sum of squares RSS_2.

(iii) Run regression model (8.30). Obtain residual sum of squares RSS_3.

(iv) In regression model (8.30) we are forcing the model to have the same parameters in both the sub-samples. Therefore, let us call the residual sum of squares obtained from this model RSS_R.

(v) Since regression models given at (8.28) and (8.29) are independent, let us call this the unrestricted model. Therefore, $RSS_{UR} = RSS_1 + RSS_2$

(vi) Suppose both the sub-samples are the same. In that case there should not be any difference between RSS_{UR} and RSS_R. Our null hypothesis in that case is H_0: There is not structural change (or, there is parameter stability).

(vii) Test the above by the following test statistic:

$$F = \frac{RSS_R - RSS_{UR})/K}{RSS_{UR}/n_1 + n_2 - 2k} \qquad \ldots (8.31)$$

It follows F-distribution with degrees of freedom $k, (n_1 + n_2 - 2k)$, where k is the number of explanatory variables in the regression model.

(viii) Check the F-distribution table given at the end of the book for desired level of significance and applicable degrees of freedom.

(ix) Draw the inference on the basis of computed value of the F-statistic obtained at step(vii).

(x) If the computed value of *F* is greater than the tabulated value, then reject the null hypothesis.

(xi) If the computed value is less than the tabulated value, do not reject the null hypothesis.

The Chow test helps us in testing for parameter stability. Note that there are three limitations of the Chow test.

(i) We assume that the error variance σ^2 is constant throughout the sample. There is no difference in the error variance between the sub-samples.

(ii) The point of structural break is not known to us. We assume that point of structural change.

(iii) We cannot apply Chow test if there are more than one structural break.

8.8 PREDICTION

In Unit 5 we explained how prediction is made on the basis of simple regression model. We extend the same procedure to multiple regression models. As in the case of simple regression models, there are two types of prediction in multiple regression models.

If we predict an individual value of the dependent variable corresponding to particular values of the explanatory variables, we obtain the 'individual prediction'. When we predict the expected value of *Y* corresponding to particular values of the explanatory variables, it is called 'mean prediction'. The expected of *Y* in both the cases (individual prediction and mean prediction) is the same. The difference between mean and individual predictions lies in their variances.

8.8.1 Mean Prediction

Let

$$X_0 = \begin{bmatrix} 1 \\ X_{02} \\ X_{03} \\ \vdots \\ X_{0k} \end{bmatrix} \qquad (8.32)$$

be the vector of values of the *X* variables for which we wish to predict $\hat{Y}_0$.

The estimated multiple regression equation, in scalar form, is

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \cdots \hat{\beta}_k X_{ki} + u_i \quad \text{... (8.33)}$$

which in matrix notation can be written compactly as

$$\hat{Y}_i = X_i' \hat{\beta} \quad \text{... (8.34)}$$

where

$$X_i' = [1 \;\; X_{2i} \; X_{3i} \ldots X_{ki}] \quad \text{... (8.35)}$$

and

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} \quad \text{... (8.36)}$$

Equation (8.34) is the mean predication of Y_i corresponding to given X_i'.

If X_i' is as given in (8.35), then (8.34) becomes

$$\left(\hat{Y}_i \middle| X_0'\right) = X_0' \hat{\beta} \quad \text{... (8.37)}$$

where the values of x_0 are fixed. You should note that (8.36) gives an unbiased prediction of $E\left(\hat{Y}_i \middle| X_0'\right)$, since $E\left(X_0' \hat{\beta}\right) = X_0' \hat{\beta}$.

<u>Variance of Mean Prediction</u>

The formula to estimate the variance of $\left(\hat{Y}_0 \middle| X_0'\right)$ is as follows:

$$\text{var}\left(\hat{Y}_0 \middle| X_0'\right) = \sigma^2 X_0' (X'X)^{-1} X_0 \quad \text{... (8.38)}$$

where σ^2 is the variance of u_i

X_0' are the *X* variables for which we wish to predict, and

since we do not know the error variance (σ^2), we replace it by its unbiased estimator $\hat{\sigma}^2$.

8.8.2 Individual Prediction

As mentioned earlier, expected value of individual prediction is the same as that of individual prediction, i.e., $\hat{Y}_i$. The variance of the individual prediction is

$$\text{var}(Y_0 | X_0) = \sigma^2 [1 + X'_0 (X'X)^{-1} X_0] \quad \text{... (8.39)}$$

where $\text{var}(Y_0|X_0)$ stands for $E\left[Y_0 - \hat{Y}_0 \mid X\right]^2$. In practice we replace σ^2 by its unbiased estimator $\hat{\sigma}^2$.

Check Your Progress 4

1) Consider a Cobb-Douglas production. Write down the steps of testing the hypothesis that it exhibits constant returns to scale.

....

....

....

....

2) Write down the steps of carrying out Chow test.

....

....

....

....

3) Point out why individual prediction has higher variance than mean prediction.

....

....

....

....

8.9 LET US SUM UP

This unit described the assumptions of classical multiple regression that fortifies normality of error term also tested by Jarque-Bera Test (J-Test for Normality). The testing of hypothesis about individual coefficients is distinguished from the overall significance test in the unit. The unit also describes the testing of equality of two regression coefficients. Later the structural stability is tested using Chow test. The multiple regression is also used for prediction of dependent variables for given values of independent variables. Both individual and joint hypothesis testing is described in the unit. Various tests such as likelihood ratio (LR), Wald (W) and Lagrange Multiplier Test (LM) are explained in the unit

8.10 ANSWERS/ HINTS TO CHECK YOUR PORGRESS EXERCISES

Check Your Progress 1

1) Refer to Sub-Section 8.2.1 and answer.
2) The Jarque-Bera test statistic is given at equation (8.2). Describe how thetest is carried out.

Check Your Progress 2

1) Refer to Sub-Section 8.3.1 and answer. Decide on the null and alternative hypotheses. Describe the staps you would follow.

2) Refer to Sub-Section 8.3.2 and answer.

Check Your Progress 3

1) It can be tested by F-test. See Section 8.4 for details.

2) Refer to Sub-Section 8.5 and answer.

Check Your Progress 4

1) We have explained in Sub-Section 8.6.1. Refer to it.

2) Refer to Sub-Section 8.7 and answer.

3) Refer to Sub-Section 8.8 and answer. It has the same logic as in the case of two variable models discussed in Section 5.7 of Unit 5.