

8

MULTIPLE REGRESSION ANALYSIS: THE PROBLEM OF INFERENCE

This chapter, a continuation of Chapter 5, extends the ideas of interval estimation and hypothesis testing developed there to models involving three or more variables. Although in many ways the concepts developed in Chapter 5 can be applied straightforwardly to the multiple regression model, a few additional features are unique to such models, and it is these features that will receive more attention in this chapter.

8.1 THE NORMALITY ASSUMPTION ONCE AGAIN

We know by now that if our sole objective is point estimation of the parameters of the regression models, the method of ordinary least squares (OLS), which does not make any assumption about the probability distribution of the disturbances u_i , will suffice. But if our objective is estimation as well as inference, then, as argued in Chapters 4 and 5, we need to assume that the u_i follow some probability distribution.

For reasons already clearly spelled out, we assumed that the u_i follow the normal distribution with zero mean and constant variance σ^2 . We continue to make the same assumption for multiple regression models. With the normality assumption and following the discussion of Chapters 4 and 7, we find that the OLS estimators of the partial regression coefficients, which are identical with the maximum likelihood (ML) estimators, are best linear unbiased estimators (BLUE).¹ Moreover, the estimators $\hat{\beta}_2$, $\hat{\beta}_3$, and $\hat{\beta}_1$ are

¹With the normality assumption, the OLS estimators $\hat{\beta}_2$, $\hat{\beta}_3$, and $\hat{\beta}_1$ are minimum-variance estimators in the entire class of unbiased estimators, whether linear or not. In short, they are BUE (best unbiased estimators). See C. R. Rao, *Linear Statistical Inference and Its Applications*, John Wiley & Sons, New York, 1965, p. 258.

themselves normally distributed with means equal to true β_2 , β_3 , and β_1 and the variances given in Chapter 7. Furthermore, $(n - 3)\hat{\sigma}^2/\sigma^2$ follows the χ^2 distribution with $n - 3$ df, and the three OLS estimators are distributed independently of $\hat{\sigma}^2$. The proofs follow the two-variable case discussed in Appendix 3. As a result and following Chapter 5, one can show that, upon replacing σ^2 by its unbiased estimator $\hat{\sigma}^2$ in the computation of the standard errors, each of the following variables

$$t = \frac{\hat{\beta}_1 - \beta_1}{\text{se}(\hat{\beta}_1)} \quad (8.1.1)$$

$$t = \frac{\hat{\beta}_2 - \beta_2}{\text{se}(\hat{\beta}_2)} \quad (8.1.2)$$

$$t = \frac{\hat{\beta}_3 - \beta_3}{\text{se}(\hat{\beta}_3)} \quad (8.1.3)$$

follows the t distribution with $n - 3$ df.

Note that the df are now $n - 3$ because in computing $\sum \hat{u}_i^2$ and hence $\hat{\sigma}^2$ we first need to estimate the three partial regression coefficients, which therefore put three restrictions on the residual sum of squares (RSS) (following this logic in the four-variable case there will be $n - 4$ df, and so on). Therefore, the t distribution can be used to establish confidence intervals as well as test statistical hypotheses about the true population partial regression coefficients. Similarly, the χ^2 distribution can be used to test hypotheses about the true σ^2 . To demonstrate the actual mechanics, we use the following illustrative example.

8.2 EXAMPLE 8.1: CHILD MORTALITY EXAMPLE REVISITED

In Chapter 7 we regressed child mortality (CM) on per capita GNP (PGNP) and the female literacy rate (FLR) for a sample of 64 countries. The regression results given in (7.6.2) are reproduced below with some additional information:

$$\begin{array}{rcccl} \widehat{CM}_i & = & 263.6416 & - & 0.0056 \text{ PGNP}_i & - & 2.2316 \text{ FLR}_i & & \\ \text{se} & = & (11.5932) & & (0.0019) & & (0.2099) & & \\ t & = & (22.7411) & & (-2.8187) & & (-10.6293) & & (8.2.1) \\ p \text{ value} & = & (0.0000)^* & & (0.0065) & & (0.0000)^* & & \\ & & & & & & R^2 = 0.7077 & & \bar{R}^2 = 0.6981 \end{array}$$

where * denotes extremely low value.

In Eq. (8.2.1) we have followed the format first introduced in Eq. (5.11.1), where the figures in the first set of parentheses are the estimated standard errors, those in the second set are the t values under the null hypothesis that the relevant population coefficient has a value of zero, and those in the third are the estimated p values. Also given are R^2 and adjusted R^2 values. We have already interpreted this regression in Example 7.1.

What about the statistical significance of the observed results? Consider, for example, the coefficient of PGNP of -0.0056 . Is this coefficient statistically significant, that is, statistically different from zero? Likewise, is the coefficient of FLR of -2.2316 statistically significant? Are both coefficients statistically significant? To answer this and related questions, let us first consider the kinds of hypothesis testing that one may encounter in the context of a multiple regression model.

8.3 HYPOTHESIS TESTING IN MULTIPLE REGRESSION: GENERAL COMMENTS

Once we go beyond the simple world of the two-variable linear regression model, hypothesis testing assumes several interesting forms, such as the following:

1. Testing hypotheses about an individual partial regression coefficient (Section 8.4)
2. Testing the overall significance of the estimated multiple regression model, that is, finding out if all the partial slope coefficients are simultaneously equal to zero (Section 8.5)
3. Testing that two or more coefficients are equal to one another (Section 8.6)
4. Testing that the partial regression coefficients satisfy certain restrictions (Section 8.7)
5. Testing the stability of the estimated regression model over time or in different cross-sectional units (Section 8.8)
6. Testing the functional form of regression models (Section 8.9)

Since testing of one or more of these types occurs so commonly in empirical analysis, we devote a section to each type.

8.4 HYPOTHESIS TESTING ABOUT INDIVIDUAL REGRESSION COEFFICIENTS

If we invoke the assumption that $u_i \sim N(0, \sigma^2)$, then, as noted in Section 8.1, we can use the t test to test a hypothesis about any *individual* partial regression coefficient. To illustrate the mechanics, consider the child mortality regression, (8.2.1). Let us postulate that

$$H_0: \beta_2 = 0 \quad \text{and} \quad H_1: \beta_2 \neq 0$$

The null hypothesis states that, with X_3 (female literacy rate) held constant, X_2 (PGNP) has no (linear) influence on Y (child mortality).² To test the null hypothesis, we use the t test given in (8.1.2). Following Chapter 5 (see Table 5.1), if the computed t value exceeds the critical t value at the chosen level of significance, we may reject the null hypothesis; otherwise, we may not reject it. For our illustrative example, using (8.1.2) and noting that $\beta_2 = 0$ under the null hypothesis, we obtain

$$t = \frac{-0.0056}{0.0020} = -2.8187 \quad (8.4.1)$$

as shown in Eq. (8.2.1).

Notice that we have 64 observations. Therefore, the degrees of freedom in this example are 61 (why?). If you refer to the t table given in **Appendix D**, we do not have data corresponding to 61 df. The closest we have are for 60 df. If we use these df, and assume α , the level of significance (i.e., the probability of committing a Type I error) of 5 percent, the critical t value is 2.0 for a two-tail test (look up $t_{\alpha/2}$ for 60 df) or 1.671 for a one-tail test (look up t_{α} for 60 df).

For our example, the alternative hypothesis is two-sided. Therefore, we use the two-tail t value. Since the computed t value of 2.8187 (in absolute terms) exceeds the critical t value of 2, we can reject the null hypothesis that PGNP has no effect on child mortality. To put it more positively, with the female literacy rate held constant, per capita GNP has a significant (negative) effect on child mortality, as one would expect a priori. Graphically, the situation is as shown in Figure 8.1.

In practice, one does not have to assume a particular value of α to conduct hypothesis testing. One can simply use the p value given in (8.2.2),

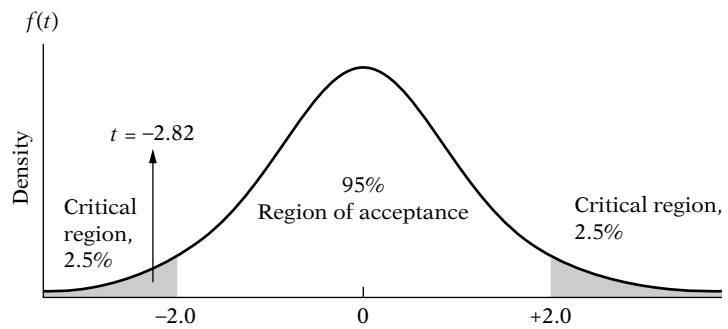


FIGURE 8.1 The 95% confidence interval for t (60 df).

²In most empirical investigations the null hypothesis is stated in this form, that is, taking the extreme position (a kind of straw man) that there is no relationship between the dependent variable and the explanatory variable under consideration. The idea here is to find out whether the relationship between the two is a trivial one to begin with.

which in the present case is 0.0065. The interpretation of this p value (i.e., the exact level of significance) is that if the null hypothesis were true, the probability of obtaining a t value of as much as 2.8187 or greater (in absolute terms) is only 0.0065 or 0.65 percent, which is indeed a small probability, much smaller than the artificially adopted value of $\alpha = 5\%$.

This example provides us an opportunity to decide whether we want to use a one-tail or a two-tail t test. Since a priori child mortality and per capita GNP are expected to be negatively related (why?), we should use the one-tail test. That is, our null and alternative hypothesis should be:

$$H_0: \beta_2 < 0 \quad \text{and} \quad H_1: \beta_2 \geq 0$$

As the reader knows by now, we can reject the null hypothesis on the basis of the one-tail t test in the present instance.

In Chapter 5 we saw the intimate connection between hypothesis testing and confidence interval estimation. For our example, the 95% confidence interval for β_2 is:

$$\hat{\beta}_2 - t_{\alpha/2} \text{se}(\hat{\beta}_2) \leq \beta_2 \leq \hat{\beta}_2 + t_{\alpha/2} \text{se}(\hat{\beta}_2)$$

which in our example becomes

$$-0.0056 - 2(0.0020) \leq \beta_2 \leq -0.0056 + 2(0.0020)$$

that is,

$$-0.0096 \leq \beta_2 \leq -0.0016 \quad (8.4.2)$$

that is, the interval, -0.0096 to -0.0016 includes the true β_2 coefficient with 95% confidence coefficient. Thus, if 100 samples of size 64 are selected and 100 confidence intervals like (8.4.2) are constructed, we expect 95 of them to contain the true population parameter β_2 . Since the interval (8.4.2) does not include the null-hypothesized value of zero, we can reject the null hypothesis that the true β_2 is zero with 95% confidence.

Thus, whether we use the t test of significance as in (8.4.1) or the confidence interval estimation as in (8.4.2), we reach the same conclusion. However, this should not be surprising in view of the close connection between confidence interval estimation and hypothesis testing.

Following the procedure just described, we can test hypotheses about the other parameters of our child mortality regression model. The necessary data are already provided in Eq. (8.2.1). For example, suppose we want to test the hypothesis that, with the influence of PGNP held constant, the female literacy rate has no effect whatsoever on child mortality. We can confidently reject this hypothesis, for under this null hypothesis the p value of obtaining an absolute t value of as much as 10.6 or greater is practically zero.

Before moving on, remember that the t -testing procedure is based on the assumption that the error term u_i follows the normal distribution. Although we cannot directly observe u_i , we can observe their proxy, the \hat{u}_i , that is, the

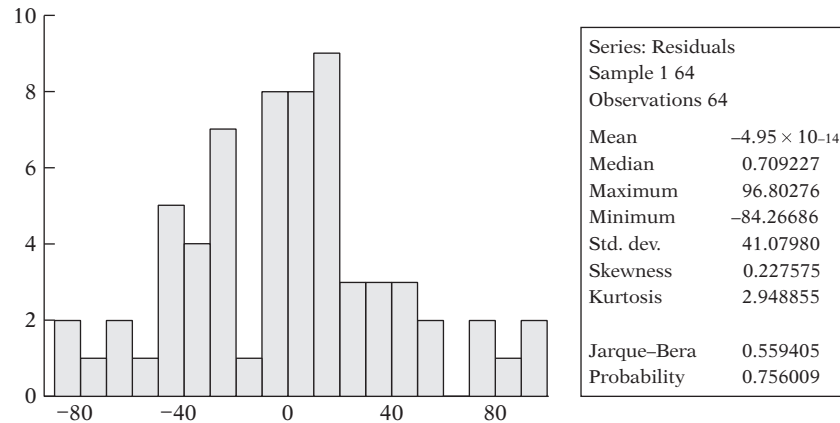


FIGURE 8.2 Histogram of residuals from regression (8.2.1).

residuals. For our mortality regression, the histogram of the residuals is as shown in Figure 8.2.

From the histogram it seems that the residuals are normally distributed. We can also compute the **Jarque-Bera (JB)** test of normality, as shown in Eq. (5.12.1). In our case the JB value is 0.5594 with a p value 0.76.³ Therefore, it seems that the error term in our example follows the normal distribution. Of course, keep in mind that the JB test is a large-sample test and our sample of 64 observations may not be necessarily large.

8.5 TESTING THE OVERALL SIGNIFICANCE OF THE SAMPLE REGRESSION

Throughout the previous section we were concerned with testing the significance of the estimated partial regression coefficients individually, that is, under the separate hypothesis that each true population partial regression coefficient was zero. But now consider the following hypothesis:

$$H_0: \beta_2 = \beta_3 = 0 \quad (8.5.1)$$

This null hypothesis is a joint hypothesis that β_2 and β_3 are jointly or simultaneously equal to zero. A test of such a hypothesis is called a test of the **overall significance** of the observed or estimated regression line, that is, whether Y is linearly related to both X_2 and X_3 .

Can the joint hypothesis in (8.5.1) be tested by testing the significance of $\hat{\beta}_2$ and $\hat{\beta}_3$ individually as in Section 8.4? The answer is no, and the reasoning is as follows.

³For our example, the skewness value is 0.2276 and the kurtosis value is 2.9488. Recall that for a normally distributed variable the skewness and kurtosis values are, respectively, 0 and 3.

In testing the individual significance of an observed partial regression coefficient in Section 8.4, we assumed implicitly that each test of significance was based on a different (i.e., independent) sample. Thus, in testing the significance of $\hat{\beta}_2$ under the hypothesis that $\beta_2 = 0$, it was assumed tacitly that the testing was based on a different sample from the one used in testing the significance of $\hat{\beta}_3$ under the null hypothesis that $\beta_3 = 0$. But to test the joint hypothesis of (8.5.1), if we use the same sample data, we shall be violating the assumption underlying the test procedure.⁴ The matter can be put differently: In (8.4.2) we established a 95% confidence interval for β_2 . But if we use the same sample data to establish a confidence interval for β_3 , say, with a confidence coefficient of 95%, we cannot assert that both β_2 and β_3 lie in their respective confidence intervals with a probability of $(1 - \alpha)(1 - \alpha) = (0.95)(0.95)$.

In other words, although the statements

$$\Pr [\hat{\beta}_2 - t_{\alpha/2} \text{se}(\hat{\beta}_2) \leq \beta_2 \leq \hat{\beta}_2 + t_{\alpha/2} \text{se}(\hat{\beta}_2)] = 1 - \alpha$$

$$\Pr [\hat{\beta}_3 - t_{\alpha/2} \text{se}(\hat{\beta}_3) \leq \beta_3 \leq \hat{\beta}_3 + t_{\alpha/2} \text{se}(\hat{\beta}_3)] = 1 - \alpha$$

are individually true, *it is not true* that the probability that the intervals

$$[\hat{\beta}_2 \pm t_{\alpha/2} \text{se}(\hat{\beta}_2), \hat{\beta}_3 \pm t_{\alpha/2} \text{se}(\hat{\beta}_3)]$$

simultaneously include β_2 and β_3 is $(1 - \alpha)^2$, because the intervals may not be independent when the same data are used to derive them. To state the matter differently,

... testing a series of single [individual] hypotheses is *not* equivalent to testing those same hypotheses jointly. The intuitive reason for this is that in a joint test of several hypotheses any single hypothesis is “affected” by the information in the other hypotheses.⁵

The upshot of the preceding argument is that for a given example (sample) only one confidence interval or only one test of significance can be obtained. How, then, does one test the simultaneous null hypothesis that $\beta_2 = \beta_3 = 0$? The answer follows.

The Analysis of Variance Approach to Testing the Overall Significance of an Observed Multiple Regression: The *F* Test

For reasons just explained, we cannot use the usual *t* test to test the joint hypothesis that the true partial slope coefficients are zero simultaneously. However, this joint hypothesis can be tested by the **analysis of variance** (ANOVA) technique first introduced in Section 5.9, which can be demonstrated as follows.

⁴In any given sample the cov ($\hat{\beta}_2, \hat{\beta}_3$) may not be zero; that is, $\hat{\beta}_2$ and $\hat{\beta}_3$ may be correlated. See (7.4.17).

⁵Thomas B. Fomby, R. Carter Hill, and Stanley R. Johnson, *Advanced Econometric Methods*, Springer-Verlag, New York, 1984, p. 37.

TABLE 8.1 ANOVA TABLE FOR THE THREE-VARIABLE REGRESSION

Source of variation	SS	df	MSS
Due to regression (ESS)	$\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}$	2	$\frac{\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}}{2}$
Due to residual (RSS)	$\sum \hat{u}_i^2$	$n - 3$	$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n - 3}$
Total	$\sum y_i^2$	$n - 1$	

Recall the identity

$$\sum y_i^2 = \hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i} + \sum \hat{u}_i^2 \quad (8.5.2)$$

$$\text{TSS} = \quad \text{ESS} \quad + \quad \text{RSS}$$

TSS has, as usual, $n - 1$ df and RSS has $n - 3$ df for reasons already discussed. ESS has 2 df since it is a function of $\hat{\beta}_2$ and $\hat{\beta}_3$. Therefore, following the ANOVA procedure discussed in Section 5.9, we can set up Table 8.1.

Now it can be shown⁶ that, under the assumption of normal distribution for u_i and the null hypothesis $\beta_2 = \beta_3 = 0$, the variable

$$F = \frac{(\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i})/2}{\sum \hat{u}_i^2 / (n - 3)} = \frac{\text{ESS/df}}{\text{RSS/df}} \quad (8.5.3)$$

is distributed as the F distribution with 2 and $n - 3$ df.

What use can be made of the preceding F ratio? It can be proved⁷ that under the assumption that the $u_i \sim N(0, \sigma^2)$,

$$E \frac{\sum \hat{u}_i^2}{n - 3} = E(\hat{\sigma}^2) = \sigma^2 \quad (8.5.4)$$

With the additional assumption that $\beta_2 = \beta_3 = 0$, it can be shown that

$$\frac{E(\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i})}{2} = \sigma^2 \quad (8.5.5)$$

Therefore, if the null hypothesis is true, both (8.5.4) and (8.5.5) give identical estimates of true σ^2 . This statement should not be surprising because if there is a trivial relationship between Y and X_2 and X_3 , the sole source of variation in Y is due to the random forces represented by u_i . If, however, the null hypothesis is false, that is, X_2 and X_3 definitely influence Y , the equality between (8.5.4) and (8.5.5) will not hold. In this case, the ESS will be

⁶See K. A. Brownlee, *Statistical Theory and Methodology in Science and Engineering*, John Wiley & Sons, New York, 1960, pp. 278–280.

⁷Ibid.

relatively larger than the RSS, taking due account of their respective df. Therefore, the F value of (8.5.3) provides a test of the null hypothesis that the true slope coefficients are simultaneously zero. If the F value computed from (8.5.3) exceeds the critical F value from the F table at the α percent level of significance, we reject H_0 ; otherwise we do not reject it. Alternatively, if the p value of the observed F is sufficiently low, we can reject H_0 .

Table 8.2 summarizes the F test. Turning to our illustrative example, we obtain the ANOVA table, as shown in Table 8.3.

Using (8.5.3), we obtain

$$F = \frac{128,681.2}{1742.88} = 73.8325 \quad (8.5.6)$$

The p value of obtaining an F value of as much as 73.8325 or greater is almost zero, leading to the rejection of the hypothesis that together PGNP and FLR have no effect on child mortality. If you were to use the conventional 5 percent level-of-significance value, the critical F value for 2 df in the numerator and 60 df in the denominator (the actual df, however, are 61) is about 3.15 or about 4.98 if you were to use the 1 percent level of significance. Obviously, the observed F of about 74 far exceeds any of these critical F values.

TABLE 8.2 A SUMMARY OF THE F STATISTIC

Null hypothesis H_0	Alternative hypothesis H_1	Critical region Reject H_0 if
$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 > \sigma_2^2$	$\frac{S_1^2}{S_2^2} > F_{\alpha, ndf, ddf}$
$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 \neq \sigma_2^2$	$\frac{S_1^2}{S_2^2} > F_{\alpha/2, ndf, ddf}$ or $< F_{(1-\alpha/2), ndf, ddf}$

Notes:

- σ_1^2 and σ_2^2 are the two population variances.
- S_1^2 and S_2^2 are the two sample variances.
- ndf and ddf denote, respectively, the numerator and denominator df.
- In computing the F ratio, put the larger S^2 value in the numerator.
- The critical F values are given in the last column. The first subscript of F is the level of significance and the second subscript is the numerator and denominator df.
- Note that $F_{(1-\alpha/2), ndf, ddf} = 1/F_{\alpha/2, ddf, ndf}$.

TABLE 8.3 ANOVA TABLE FOR THE CHILD MORTALITY EXAMPLE

Source of variation	SS	df	MSS
Due to regression	257,362.4	2	128,681.2
Due to residuals	106,315.6	61	1742.88
Total	363,678	63	

We can generalize the preceding F -testing procedure as follows.

Testing the Overall Significance of a Multiple Regression: The F Test

Decision Rule. Given the k -variable regression model:

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

To test the hypothesis

$$H_0: \beta_2 = \beta_3 = \cdots = \beta_k = 0$$

(i.e., all slope coefficients are simultaneously zero) versus

$$H_1: \text{Not all slope coefficients are simultaneously zero}$$

compute

$$F = \frac{\text{ESS}/\text{df}}{\text{RSS}/\text{df}} = \frac{\text{ESS}/(k-1)}{\text{RSS}/(n-k)} \quad (8.5.7)$$

If $F > F_\alpha(k-1, n-k)$, reject H_0 ; otherwise you do not reject it, where $F_\alpha(k-1, n-k)$ is the *critical F* value at the α level of significance and $(k-1)$ numerator df and $(n-k)$ denominator df. Alternatively, if the p value of F obtained from (8.5.7) is sufficiently low, one can reject H_0 .

Needless to say, in the three-variable case (Y and X_2, X_3) k is 3, in the four-variable case k is 4, and so on.

In passing, note that most regression packages routinely calculate the F value (given in the analysis of variance table) along with the usual regression output, such as the estimated coefficients, their standard errors, t values, etc. The null hypothesis for the t computation is usually assumed to be $\beta_i = 0$.

Individual versus Joint Testing of Hypotheses. In Section 8.4 we discussed the test of significance of a single regression coefficient and in Section 8.5 we have discussed the joint or overall test of significance of the estimated regression (i.e., all slope coefficients are simultaneously equal to zero). **We reiterate that these tests are different.** Thus, on the basis of the t test or confidence interval (of Section 8.4) it is possible to accept the hypothesis that a particular slope coefficient, β_k , is zero, and yet reject the joint hypothesis that all slope coefficients are zero.

The lesson to be learned is that the joint “message” of individual confidence intervals is no substitute for a joint confidence region [implied by the F test] in performing joint tests of hypotheses and making joint confidence statements.⁸

⁸Fomby et al., op. cit., p. 42.

An Important Relationship between R^2 and F

There is an intimate relationship between the coefficient of determination R^2 and the F test used in the analysis of variance. Assuming the normal distribution for the disturbances u_i and the null hypothesis that $\beta_2 = \beta_3 = 0$, we have seen that

$$F = \frac{\text{ESS}/2}{\text{RSS}/(n-3)} \quad (8.5.8)$$

is distributed as the F distribution with 2 and $n - 3$ df.

More generally, in the k -variable case (including intercept), if we assume that the disturbances are normally distributed and that the null hypothesis is

$$H_0: \beta_2 = \beta_3 = \dots = \beta_k = 0 \quad (8.5.9)$$

then it follows that

$$F = \frac{\text{ESS}/(k-1)}{\text{RSS}/(n-k)} \quad (8.5.7) = (8.5.10)$$

follows the F distribution with $k - 1$ and $n - k$ df. (*Note:* The total number of parameters to be estimated is k , of which one is the intercept term.)

Let us manipulate (8.5.10) as follows:

$$\begin{aligned} F &= \frac{n-k}{k-1} \frac{\text{ESS}}{\text{RSS}} \\ &= \frac{n-k}{k-1} \frac{\text{ESS}}{\text{TSS} - \text{ESS}} \\ &= \frac{n-k}{k-1} \frac{\text{ESS}/\text{TSS}}{1 - (\text{ESS}/\text{TSS})} \\ &= \frac{n-k}{k-1} \frac{R^2}{1 - R^2} \\ &= \frac{R^2/(k-1)}{(1 - R^2)/(n-k)} \end{aligned} \quad (8.5.11)$$

where use is made of the definition $R^2 = \text{ESS}/\text{TSS}$. Equation (8.5.11) shows how F and R^2 are related. These two vary directly. When $R^2 = 0$, F is zero ipso facto. The larger the R^2 , the greater the F value. In the limit, when $R^2 = 1$, F is infinite. *Thus the F test, which is a measure of the overall significance of the estimated regression, is also a test of significance of R^2 .* In other words, testing the null hypothesis (8.5.9) is equivalent to testing the null hypothesis that (the population) R^2 is zero.

TABLE 8.4 ANOVA TABLE IN TERMS OF R^2

Source of variation	SS	df	MSS*
Due to regression	$R^2(\sum y_i^2)$	2	$R^2(\sum y_i^2)/2$
Due to residuals	$(1 - R^2)(\sum y_i^2)$	$n - 3$	$(1 - R^2)(\sum y_i^2)/(n - 3)$
Total	$\sum y_i^2$	$n - 1$	

*Note that in computing the F value there is no need to multiply R^2 and $(1 - R^2)$ by $\sum y_i^2$ because it drops out, as shown in (8.5.12).

For the three-variable case (8.5.11) becomes

$$F = \frac{R^2/2}{(1 - R^2)/(n - 3)} \tag{8.5.12}$$

By virtue of the close connection between F and R^2 , the ANOVA Table 8.1 can be recast as Table 8.4.

For our illustrative example, using (8.5.12) we obtain:

$$F = \frac{0.7077/2}{(1 - 0.7077)/61} = 73.8726$$

which is about the same as obtained before, except for the rounding errors.

One advantage of the F test expressed in terms of R^2 is its ease of computation: All that one needs to know is the R^2 value. Therefore, the overall F test of significance given in (8.5.7) can be recast in terms of R^2 as shown in Table 8.4.

Testing the Overall Significance of a Multiple Regression in Terms of R^2

Decision Rule. Testing the overall significance of a regression in terms of R^2 : Alternative but equivalent test to (8.5.7).

Given the k -variable regression model:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + u_i$$

To test the hypothesis

$$H_0: \beta_2 = \beta_3 = \dots = \beta_k = 0$$

versus

H_1 : Not all slope coefficients are simultaneously zero

compute

$$F = \frac{R^2/(k - 1)}{(1 - R^2)/(n - k)} \tag{8.5.13}$$

If $F > F_{\alpha(k-1, n-k)}$, reject H_0 ; otherwise you may accept H_0 where $F_{\alpha(k-1, n-k)}$ is the critical F value at the α level of significance and $(k - 1)$ numerator df and $(n - k)$ denominator df. Alternatively, if the p value of F obtained from (8.5.13) is sufficiently low, reject H_0 .

Before moving on, return to Example 7.5 in Chapter 7. From regression (7.10.7) we observe that RGDP (relative per capita GDP) and RGDP squared explain only about 5.3 percent of the variation in GDPG (GDP growth rate) in a sample of 119 countries. This R^2 of 0.053 seems a “low” value. Is it really statistically different from zero? How do we find that out?

Recall our earlier discussion in “An Important Relationship between R^2 and F ” about the relationship between R^2 and the F value as given in (8.5.11) or (8.5.12) for the specific case of two regressors. As noted, if R^2 is zero, then F is zero ipso facto, which will be the case if the regressors have no impact whatsoever on the regressand. Therefore, if we insert $R^2 = 0.053$ into formula (8.5.12), we obtain

$$F = \frac{0.053/2}{(1 - 0.053)/116} = 3.2475 \quad (8.5.13)$$

Under the null hypothesis that $R^2 = 0$, the preceding F value follows the F distribution with 2 and 116 df in the numerator, respectively. (*Note:* There are 119 observations and two regressors.) From the F table we see that this F value is significant at about the 5 percent level; the p value is actually 0.0425. Therefore, we can reject the null hypothesis that the two regressors have no impact on the regressand, notwithstanding the fact that the R^2 is only 0.053.

This example brings out an important empirical observation that in cross-sectional data involving several observations, one generally obtains low R^2 because of the diversity of the cross-sectional units. Therefore, one should not be surprised or worried about finding low R^2 's in cross-sectional regressions. What is relevant is that the model is correctly specified, that the regressors have the correct (i.e., theoretically expected) signs, and that (hopefully) the regression coefficients are statistically significant. The reader should check that *individually* both the regressors in (7.10.7) are statistically significant at the 5 percent or better level (i.e., lower than 5 percent).

The “Incremental” or “Marginal” Contribution of an Explanatory Variable

In Chapter 7 we stated that generally we cannot allocate the R^2 value among the various regressors. In our child mortality example we found that the R^2 was 0.7077 but we cannot say what part of this value is due to the regressor PGNP and what part is due to female literacy rate (FLR) because of possible correlation between the two regressors in the sample at hand. We can shed more light on this using the analysis of covariance technique.

For our illustrative example we found that *individually* X_2 (PGNP) and X_3 (FLR) were statistically significant on the basis of (*separate*) t tests. We have also found that on the basis of the F test *collectively* both the regressors have a significant effect on the regressand Y (child mortality).

Now suppose we introduce PGNP and FLR *sequentially*; that is, we first regress child mortality on PGNP and assess its significance and then add FLR to the model to find out whether it contributes anything (of course, the order in which PGNP and FLR enter can be reversed). By contribution we mean whether the addition of the variable to the model increases ESS (and hence R^2) “significantly” in relation to the RSS. This contribution may appropriately be called the **incremental**, or **marginal**, contribution of an explanatory variable.

The topic of incremental contribution is an important one in practice. In most empirical investigations the researcher may not be completely sure whether it is worth adding an X variable to the model knowing that several other X variables are already present in the model. One does not wish to include variable(s) that contribute very little toward ESS. By the same token, one does not want to exclude variable(s) that substantially increase ESS. But how does one decide whether an X variable significantly reduces RSS? The analysis of variance technique can be easily extended to answer this question.

Suppose we first regress child mortality on PGNP and obtain the following regression:

$$\begin{aligned} \widehat{CM}_i &= 157.4244 - 0.0114 \text{ PGNP} && \text{(8.5.14)} \\ t &= (15.9894) \quad (-3.5156) && r^2 = 0.1662 \\ p \text{ value} &= (0.0000) \quad (0.0008) && \text{adj } r^2 = 0.1528 \end{aligned}$$

As these results show, PGNP has a significant effect on CM. The ANOVA table corresponding to the preceding regression is given in Table 8.5.

Assuming the disturbances u_i are normally distributed and the hypothesis that PGNP has no effect on CM, we obtain the F value of

$$F = \frac{60,449.5}{4890.7822} = 12.3598 \quad \text{(8.5.15)}$$

TABLE 8.5 ANOVA TABLE FOR REGRESSION (8.5.14)

Source of variation	SS	df	MSS
ESS (due to PGNP)	60,449.5	1	60,449.5
RSS	303,228.5	62	4890.7822
Total	363,678	63	

which follows the F distribution with 1 and 62 df. This F value is highly significant, as the computed p value is 0.0008. Thus, as before, we reject the hypothesis that PGNP has no effect on CM. Incidentally, note that $t^2 = (-3.5156)^2 = 12.3594$, which is approximately the same as the F value of (8.5.15), where the t value is obtained from (8.5.14). But this should not be surprising in view of the fact that the square of the t statistic with n df is equal to the F value with 1 df in the numerator and n df in the denominator, a relationship first established in Chapter 5. Note that in the present example, $n = 64$.

Having run the regression (8.5.14), let us suppose we decide to add FLR to the model and obtain the multiple regression (8.2.1). The questions we want to answer are:

1. What is the marginal, or incremental, contribution of FLR, knowing that PGNP is already in the model and that it is significantly related to CM?
2. Is the incremental contribution of FLR statistically significant?
3. What is the criterion for adding variables to the model?

The preceding questions can be answered by the ANOVA technique. To see this, let us construct Table 8.6. In this table X_2 refers to PGNP and X_3 refers to FLR.

To assess the *incremental* contribution of X_3 after allowing for the contribution of X_2 , we form

$$\begin{aligned}
 F &= \frac{Q_2/\text{df}}{Q_4/\text{df}} \\
 &= \frac{\text{ESS}_{\text{new}} - \text{ESS}_{\text{old}}/\text{number of new regressors}}{\text{RSS}_{\text{new}}/\text{df} (= n - \text{number of parameters in the new model})} \\
 &= \frac{Q_2/1}{Q_4/12} \text{ for our example}
 \end{aligned}$$

(8.5.16)

TABLE 8.6 ANOVA TABLE TO ASSESS INCREMENTAL CONTRIBUTION OF A VARIABLE(S)

Source of variation	SS	df	MSS
ESS due to X_2 alone	$Q_1 = \hat{\beta}_{12}^2 \sum x_2^2$	1	$\frac{Q_1}{1}$
ESS due to the addition of X_3	$Q_2 = Q_3 - Q_1$	1	$\frac{Q_2}{1}$
ESS due to both X_2, X_3	$Q_3 = \hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}$	2	$\frac{Q_3}{2}$
RSS	$Q_4 = Q_5 - Q_3$	$n - 3$	$\frac{Q_4}{n - 3}$
Total	$Q_5 = \sum y_i^2$	$n - 1$	

TABLE 8.7 ANOVA TABLE FOR THE ILLUSTRATIVE EXAMPLE: INCREMENTAL ANALYSIS

Source of variation	SS	df	MSS
ESS due to PGNP	60,449.5	1	60,449.5
ESS due to the addition of FLR	196,912.9	1	196,912.9
ESS due to PGNP and FLR	257,362.4	2	128,681.2
RSS	106,315.6	61	1742.8786
Total	363,678	63	

where $ESS_{\text{new}} = ESS$ under the new model (i.e., after adding the new regressors = Q_3), $ESS_{\text{old}} = ESS$ under the old model (= Q_1) and $RSS_{\text{new}} = RSS$ under the new model (i.e., after taking into account all the regressors = Q_4). For our illustrative example the results are as shown in Table 8.7.

Now applying (8.5.16), we obtain:

$$F = \frac{196,912.9}{1742.8786} = 112.9814 \quad (8.5.17)$$

Under the usual assumptions, this F value follows the F distribution with 1 and 62 df. The reader should check that this F value is highly significant, suggesting that addition of FLR to the model significantly increases ESS and hence the R^2 value. Therefore, FLR should be added to the model. Again, note that if you square the value of the FLR coefficient in the multiple regression (8.2.1), which is $(-10.6293)^2$, you will obtain the F value of (8.5.17), save for the rounding errors.

Incidentally, the F ratio of (8.5.16) can be recast by using the R^2 values only, as we did in (8.5.13). As exercise 8.2 shows, the F ratio of (8.5.16) is *equivalent* to the following F ratio:⁹

$$\begin{aligned}
 F &= \frac{(R^2_{\text{new}} - R^2_{\text{old}})/df}{(1 - R^2_{\text{new}})/df} \\
 &= \frac{(R^2_{\text{new}} - R^2_{\text{old}})/\text{number of new regressors}}{(1 - R^2_{\text{new}})/df (= n - \text{number of parameters in the new model})}
 \end{aligned}
 \quad (8.5.18)$$

This F ratio follows the F distribution with the appropriate numerator and denominator df, 1 and 61 in our illustrative example.

For our example, $R^2_{\text{new}} = 0.7077$ [from Eq. (8.2.1)] and $R^2_{\text{old}} = 0.1662$ [from Eq. (8.5.14)]. Therefore,

$$F = \frac{(0.7077 - 0.1662)/1}{(1 - 0.7077)/61} = 113.05 \quad (8.5.19)$$

⁹The following F test is a special case of the more general F test given in (8.7.9) or (8.7.10) in Sec. 8.7.

which is about the same as that obtained from (8.5.17), except for the rounding errors. This F is highly significant, reinforcing our earlier finding that the variable FLR belongs in the model.

A cautionary note: If you use the R^2 version of the F test given in (8.5.11), make sure that the dependent variable in the new and the old models is the same. If they are different, use the F test given in (8.5.16).

When to Add a New Variable. The F -test procedure just outlined provides a formal method of deciding whether a variable should be added to a regression model. Often researchers are faced with the task of choosing from several competing models **involving the same dependent variable** but with different explanatory variables. As a matter of ad hoc choice (because very often the theoretical foundation of the analysis is weak), these researchers frequently choose the model that gives the highest adjusted R^2 . Therefore, if the inclusion of a variable increases \bar{R}^2 , it is retained in the model although it does not reduce RSS significantly in the statistical sense. The question then becomes: When does the adjusted R^2 increase? It can be shown that \bar{R}^2 will increase if the t value of the coefficient of the newly added variable is larger than 1 in absolute value, where the t value is computed under the hypothesis that the population value of the said coefficient is zero [i.e., the t value computed from (5.3.2) under the hypothesis that the true β value is zero].¹⁰ The preceding criterion can also be stated differently: \bar{R}^2 will increase with the addition of an extra explanatory variable only if the $F (= t^2)$ value of that variable exceeds 1.

Applying either criterion, the FLR variable in our child mortality example with a t value of -10.6293 or an F value of 112.9814 should increase \bar{R}^2 , which indeed it does—when FLR is added to the model, \bar{R}^2 increases from 0.1528 to 0.6981 .

When to Add a Group of Variables. Can we develop a similar rule for deciding whether it is worth adding (or dropping) a group of variables from a model? The answer should be apparent from (8.5.18): *If adding (dropping) a group of variables to the model gives an F value greater (less) than 1, R^2 will increase (decrease).* Of course, from (8.5.18) one can easily find out whether the addition (subtraction) of a group of variables significantly increases (decreases) the explanatory power of a regression model.

8.6 TESTING THE EQUALITY OF TWO REGRESSION COEFFICIENTS

Suppose in the multiple regression

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_i \quad (8.6.1)$$

¹⁰For proof, see Dennis J. Aigner, *Basic Econometrics*, Prentice Hall, Englewood Cliffs, N.J., 1971, pp. 91–92.

we want to test the hypotheses

$$\begin{aligned} H_0: \beta_3 &= \beta_4 & \text{or} & & (\beta_3 - \beta_4) &= 0 \\ H_1: \beta_3 &\neq \beta_4 & \text{or} & & (\beta_3 - \beta_4) &\neq 0 \end{aligned} \quad (8.6.2)$$

that is, the two slope coefficients β_3 and β_4 are equal.

Such a null hypothesis is of practical importance. For example, let (8.6.1) represent the demand function for a commodity where Y = amount of a commodity demanded, X_2 = price of the commodity, X_3 = income of the consumer, and X_4 = wealth of the consumer. The null hypothesis in this case means that the income and wealth coefficients are the same. Or, if Y_i and the X 's are expressed in logarithmic form, the null hypothesis in (8.6.2) implies that the income and wealth elasticities of consumption are the same. (Why?)

How do we test such a null hypothesis? Under the classical assumptions, it can be shown that

$$t = \frac{(\hat{\beta}_3 - \hat{\beta}_4) - (\beta_3 - \beta_4)}{\text{se}(\hat{\beta}_3 - \hat{\beta}_4)} \quad (8.6.3)$$

follows the t distribution with $(n - 4)$ df because (8.6.1) is a four-variable model or, more generally, with $(n - k)$ df, where k is the total number of parameters estimated, including the constant term. The $\text{se}(\hat{\beta}_3 - \hat{\beta}_4)$ is obtained from the following well-known formula (see **Appendix A** for details):

$$\text{se}(\hat{\beta}_3 - \hat{\beta}_4) = \sqrt{\text{var}(\hat{\beta}_3) + \text{var}(\hat{\beta}_4) - 2 \text{cov}(\hat{\beta}_3, \hat{\beta}_4)} \quad (8.6.4)$$

If we substitute the null hypothesis and the expression for the $\text{se}(\hat{\beta}_3 - \hat{\beta}_4)$ into (8.6.3), our test statistic becomes

$$t = \frac{\hat{\beta}_3 - \hat{\beta}_4}{\sqrt{\text{var}(\hat{\beta}_3) + \text{var}(\hat{\beta}_4) - 2 \text{cov}(\hat{\beta}_3, \hat{\beta}_4)}} \quad (8.6.5)$$

Now the testing procedure involves the following steps:

1. Estimate $\hat{\beta}_3$ and $\hat{\beta}_4$. Any standard computer package can do that.
2. Most standard computer packages routinely compute the variances and covariances of the estimated parameters.¹¹ From these estimates the standard error in the denominator of (8.6.5) can be easily obtained.
3. Obtain the t ratio from (8.6.5). Note the null hypothesis in the present case is $(\beta_3 - \beta_4) = 0$.
4. If the t variable computed from (8.6.5) exceeds the critical t value at the designated level of significance for given df, then you can reject the null hypothesis; otherwise, you do not reject it. Alternatively, if the p value of the

¹¹The algebraic expression for the covariance formula is rather involved. **Appendix C** provides a compact expression for it, however, using matrix notation.

t statistic from (8.6.5) is reasonably low, one can reject the null hypothesis. Note that the lower the p value, the greater the evidence against the null hypothesis. Therefore, when we say that a p value is low or reasonably low, we mean that it is less than the significance level, such as 10, 5, or 1 percent. Some personal judgment is involved in this decision.

EXAMPLE 8.2

THE CUBIC COST FUNCTION REVISITED

Recall the cubic total cost function estimated in Section 7.10, which for convenience is reproduced below:

$$\begin{aligned} \hat{Y}_i &= 141.7667 + 63.4777X_i - 12.9615X_i^2 + 0.9396X_i^3 \\ \text{se} &= (6.3753) \quad (4.7786) \quad (0.9857) \quad (0.0591) \quad (7.10.6) \\ \text{cov}(\hat{\beta}_3, \hat{\beta}_4) &= -0.0576; \quad R^2 = 0.9983 \end{aligned}$$

where Y is total cost and X is output, and where the figures in parentheses are the estimated standard errors.

Suppose we want to test the hypothesis that the coefficients of the X^2 and X^3 terms in the cubic cost function are the same, that is, $\beta_3 = \beta_4$ or $(\beta_3 - \beta_4) = 0$. In the regression (7.10.6) we have all the necessary output to conduct the t test of (8.6.5). The actual mechanics are as follows:

$$\begin{aligned} t &= \frac{\hat{\beta}_3 - \hat{\beta}_4}{\sqrt{\text{var}(\hat{\beta}_3) + \text{var}(\hat{\beta}_4) - 2 \text{cov}(\hat{\beta}_3, \hat{\beta}_4)}} \\ &= \frac{-12.9615 - 0.9396}{\sqrt{(0.9867)^2 + (0.0591)^2 - 2(-0.0576)}} \quad (8.6.6) \\ &= \frac{-13.9011}{1.0442} = -13.3130 \end{aligned}$$

The reader can verify that for 6 df (why?) the observed t value exceeds the critical t value even at the 0.002 (or 0.2 percent) level of significance (two-tail test); the p value is extremely small, 0.000006. Hence we can reject the hypothesis that the coefficients of X^2 and X^3 in the cubic cost function are identical.

**8.7 RESTRICTED LEAST SQUARES: TESTING LINEAR
EQUALITY RESTRICTIONS**

There are occasions where economic theory may suggest that the coefficients in a regression model satisfy some linear equality restrictions. For instance, consider the Cobb–Douglas production function:

$$Y_i = \beta_1 X_{2i}^{\beta_2} X_{3i}^{\beta_3} e^{u_i} \quad (7.9.1) = (8.7.1)$$

where Y = output, X_2 = labor input, and X_3 = capital input. Written in log form, the equation becomes

$$\ln Y_i = \beta_0 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + u_i \quad (8.7.2)$$

where $\beta_0 = \ln \beta_1$.

Now if there are constant returns to scale (equiproportional change in output for an equiproportional change in the inputs), economic theory would suggest that

$$\beta_2 + \beta_3 = 1 \quad (8.7.3)$$

which is an example of a linear equality restriction.¹²

How does one find out if there are constant returns to scale, that is, if the restriction (8.7.3) is valid? There are two approaches.

The *t*-Test Approach

The simplest procedure is to estimate (8.7.2) in the usual manner without taking into account the restriction (8.7.3) explicitly. This is called the **unrestricted** or **unconstrained regression**. Having estimated β_2 and β_3 (say, by OLS method), a test of the hypothesis or restriction (8.7.3) can be conducted by the *t* test of (8.6.3), namely,

$$\begin{aligned} t &= \frac{(\hat{\beta}_2 + \hat{\beta}_3) - (\beta_2 + \beta_3)}{\text{se}(\hat{\beta}_2 + \hat{\beta}_3)} \\ &= \frac{(\hat{\beta}_2 + \hat{\beta}_3) - 1}{\sqrt{\text{var}(\hat{\beta}_2) + \text{var}(\hat{\beta}_3) + 2 \text{cov}(\hat{\beta}_2, \hat{\beta}_3)}} \end{aligned} \quad (8.7.4)$$

where $(\beta_2 + \beta_3) = 1$ under the null hypothesis and where the denominator is the standard error of $(\hat{\beta}_2 + \hat{\beta}_3)$. Then following Section 8.6, if the *t* value computed from (8.7.4) exceeds the critical *t* value at the chosen level of significance, we reject the hypothesis of constant returns to scale; otherwise we do not reject it.

The *F*-Test Approach: Restricted Least Squares

The preceding *t* test is a kind of postmortem examination because we try to find out whether the linear restriction is satisfied after estimating the “unrestricted” regression. A direct approach would be to incorporate the restriction (8.7.3) into the estimating procedure at the outset. In the present example, this procedure can be done easily. From (8.7.3) we see that

$$\beta_2 = 1 - \beta_3 \quad (8.7.5)$$

or

$$\beta_3 = 1 - \beta_2 \quad (8.7.6)$$

Therefore, using either of these equalities, we can eliminate one of the β coefficients in (8.7.2) and estimate the resulting equation. Thus, if we use (8.7.5), we can write the Cobb–Douglas production function as

$$\begin{aligned} \ln Y_i &= \beta_0 + (1 - \beta_3) \ln X_{2i} + \beta_3 \ln X_{3i} + u_i \\ &= \beta_0 + \ln X_{2i} + \beta_3 (\ln X_{3i} - \ln X_{2i}) + u_i \end{aligned}$$

¹²If we had $\beta_2 + \beta_3 < 1$, this relation would be an example of a linear inequality restriction. To handle such restrictions, one needs to use mathematical programming techniques.

or

$$(\ln Y_i - \ln X_{2i}) = \beta_0 + \beta_3(\ln X_{3i} - \ln X_{2i}) + u_i \quad (8.7.7)$$

or

$$\ln(Y_i/X_{2i}) = \beta_0 + \beta_3 \ln(X_{3i}/X_{2i}) + u_i \quad (8.7.8)$$

where (Y_i/X_{2i}) = output/labor ratio and (X_{3i}/X_{2i}) = capital labor ratio, quantities of great economic importance.

Notice how the original equation (8.7.2) is transformed. Once we estimate β_3 from (8.7.7) or (8.7.8), β_2 can be easily estimated from the relation (8.7.5). Needless to say, this procedure will guarantee that the sum of the estimated coefficients of the two inputs will equal 1. The procedure outlined in (8.7.7) or (8.7.8) is known as **restricted least squares (RLS)**. This procedure can be generalized to models containing any number of explanatory variables and more than one linear equality restriction. The generalization can be found in Theil.¹³ (See also general F testing below.)

How do we compare the unrestricted and restricted least-squares regressions? In other words, how do we know that, say, the restriction (8.7.3) is valid? This question can be answered by applying the F test as follows. Let

$$\sum \hat{u}_{UR}^2 = \text{RSS of the unrestricted regression (8.7.2)}$$

$$\sum \hat{u}_R^2 = \text{RSS of the restricted regression (8.7.7)}$$

m = number of linear restrictions (1 in the present example)

k = number of parameters in the unrestricted regression

n = number of observations

Then,

$$F = \frac{(\text{RSS}_R - \text{RSS}_{UR})/m}{\text{RSS}_{UR}/(n-k)} \quad (8.7.9)$$

$$= \frac{(\sum \hat{u}_R^2 - \sum \hat{u}_{UR}^2)/m}{\sum \hat{u}_{UR}^2/(n-k)}$$

follows the F distribution with m , $(n - k)$ df. (Note: UR and R stand for unrestricted and restricted, respectively.)

The F test above can also be expressed in terms of R^2 as follows:

$$F = \frac{(R_{UR}^2 - R_R^2)/m}{(1 - R_{UR}^2)/(n-k)} \quad (8.7.10)$$

where R_{UR}^2 and R_R^2 are, respectively, the R^2 values obtained from the unrestricted and restricted regressions, that is, from the regressions (8.7.2) and

¹³Henri Theil, *Principles of Econometrics*, John Wiley & Sons, New York, 1971, pp. 43–45.

(8.7.7). It should be noted that

$$R_{UR}^2 \geq R_R^2 \quad (8.7.11)$$

and

$$\sum \hat{u}_{UR}^2 \leq \sum \hat{u}_R^2 \quad (8.7.12)$$

In exercise 8.4 you are asked to justify these statements.

A Cautionary Note: In using (8.7.10) keep in mind that if the dependent variable in the restricted and unrestricted models is not the same, R_{UR}^2 and R_R^2 are not directly comparable. In that case, use the procedure described in Chapter 7 to render the two R^2 values comparable (see Example 8.3 below) or use the F test given in (8.7.9).

EXAMPLE 8.3

THE COBB–DOUGLAS PRODUCTION FUNCTION FOR THE MEXICAN ECONOMY, 1955–1974

By way of illustrating the preceding discussion consider the data given in Table 8.8. Attempting to fit the Cobb–Douglas production function to these data, yielded the following results:

$$\begin{aligned} \widehat{\ln \text{GDP}}_t &= -1.6524 + 0.3397 \ln \text{Labor}_t + 0.8460 \ln \text{Capital}_t & (8.7.13) \\ t &= (-2.7259) \quad (1.8295) \quad (9.0625) \\ p \text{ value} &= (0.0144) \quad (0.0849) \quad (0.0000) \\ & R^2 = 0.9951 \quad \text{RSS}_{UR} = 0.0136 \end{aligned}$$

where RSS_{UR} is the unrestricted RSS, as we have put no restrictions on estimating (8.7.13).

We have already seen in Chapter 7 how to interpret the coefficients of the Cobb–Douglas production function. As you can see, the output/labor elasticity is about 0.34 and the output/capital elasticity is about 0.85. If we add these coefficients, we obtain 1.19, suggesting that perhaps the Mexican economy during the stated time period was experiencing increasing returns to scale. Of course, we do not know if 1.19 is statistically different from 1.

To see if that is the case, let us impose the restriction of constant returns to scale, which gives the following regression:

$$\begin{aligned} \widehat{\ln (\text{GDP}/\text{Labor})}_t &= -0.4947 + 1.0153 \ln (\text{Capital}/\text{Labor})_t & (8.7.14) \\ t &= (-4.0612) \quad (28.1056) \\ p \text{ value} &= (0.0007) \quad (0.0000) \\ & R_R^2 = 0.9777 \quad \text{RSS}_R = 0.0166 \end{aligned}$$

where RSS_R is the restricted RSS, for we have imposed the restriction that there are constant returns to scale.

(Continued)

EXAMPLE 9.3 (Continued)

Since the dependent variable in the preceding two regressions is different, we have to use the F test given in (8.7.9). We have the necessary data to obtain the F value.

$$\begin{aligned} F &= \frac{(RSS_R - RSS_{UR})/m}{RSS_{UR}/(n-k)} \\ &= \frac{(0.0166 - 0.0136)/1}{(0.0136)/(20 - 3)} \\ &= 3.75 \end{aligned}$$

Note in the present case $m = 1$, as we have imposed only one restriction and $(n - k)$ is 17, since we have 20 observations and three parameters in the unrestricted regression.

This F value follows the F distribution with 1 df in the numerator and 17 df in the denominator. The reader can easily check that this F value is not significant at the 5% level. (See **Appendix D**, Table D.3.)

The conclusion then is that the Mexican economy was probably characterized by constant returns to scale over the sample period and therefore there may be no harm in using the restricted regression given in (8.7.14). As this regression shows, if capital/labor ratio increased by 1 percent, on average, labor productivity went up by about 1 percent.

TABLE 8.8
REAL GDP, EMPLOYMENT, AND REAL FIXED CAPITAL—MEXICO

Year	GDP*	Employment [†]	Fixed capital [‡]
1955	114043	8310	182113
1956	120410	8529	193749
1957	129187	8738	205192
1958	134705	8952	215130
1959	139960	9171	225021
1960	150511	9569	237026
1961	157897	9527	248897
1962	165286	9662	260661
1963	178491	10334	275466
1964	199457	10981	295378
1965	212323	11746	315715
1966	226977	11521	337642
1967	241194	11540	363599
1968	260881	12066	391847
1969	277498	12297	422382
1970	296530	12955	455049
1971	306712	13338	484677
1972	329030	13738	520553
1973	354057	15924	561531
1974	374977	14154	609825

*Millions of 1960 pesos;

[†]Thousands of people;

[‡]Millions of 1960 pesos.

Source: Victor J. Elias, *Sources of Growth: A Study of Seven Latin American Economies*, International Center for Economic Growth, ICS Press, San Francisco, 1992. Data from Tables E5, E12, and E14.

General F Testing¹⁴

The F test given in (8.7.10) or its equivalent (8.7.9) provides a general method of testing hypotheses about one or more parameters of the k -variable regression model:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i \quad (8.7.15)$$

The F test of (8.5.16) or the t test of (8.6.3) is but a specific application of (8.7.10). Thus, hypotheses such as

$$H_0: \beta_2 = \beta_3 \quad (8.7.16)$$

$$H_0: \beta_3 + \beta_4 + \beta_5 = 3 \quad (8.7.17)$$

which involve some linear restrictions on the parameters of the k -variable model, or hypotheses such as

$$H_0: \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0 \quad (8.7.18)$$

which imply that some regressors are absent from the model, can all be tested by the F test of (8.7.10).

From the discussion in Sections 8.5 and 8.7, the reader will have noticed that the general strategy of F testing is this: There is a larger model, the *unconstrained model* (8.7.15), and then there is a smaller model, the *constrained* or *restricted model*, which is obtained from the larger model by deleting some variables from it, e.g., (8.7.18), or by putting some linear restrictions on one or more coefficients of the larger model, e.g., (8.7.16) or (8.7.17).

We then fit the unconstrained and constrained models to the data and obtain the respective coefficients of determination, namely, R_{UR}^2 and R_R^2 . We note the df in the unconstrained model ($= n - k$) and also note the df in the constrained model ($= m$), m being the number of linear restriction [e.g., 1 in (8.7.16) or (8.7.18)] or the number of regressors omitted from the model [e.g., $m = 4$ if (8.7.18) holds, since four regressors are assumed to be absent from the model]. We then compute the F ratio as indicated in (8.7.9) or (8.7.10) and use this *Decision Rule*: *If the computed F exceeds $F_\alpha(m, n - k)$, where $F_\alpha(m, n - k)$ is the critical F at the α level of significance, we reject the null hypothesis; otherwise we do not reject it.*

¹⁴If one is using the maximum likelihood approach to estimation, then a test similar to the one discussed shortly is the **likelihood ratio test**, which is slightly involved and is therefore discussed in the appendix to the chapter. For further discussion, see Theil, op. cit., pp. 179–184.

Let us illustrate:

EXAMPLE 8.4

THE DEMAND FOR CHICKEN IN THE UNITED STATES, 1960–1982

In exercise 7.19, among other things, you were asked to consider the following demand function for chicken:

$$\ln Y_t = \beta_1 + \beta_2 \ln X_{2t} + \beta_3 \ln X_{3t} + \beta_4 \ln X_{4t} + \beta_5 \ln X_{5t} + u_t \quad (8.7.19)$$

where Y = per capita consumption of chicken, lb, X_2 = real disposable per capita income, \$, X_3 = real retail price of chicken per lb, ¢, X_4 = real retail price of pork per lb, ¢, and X_5 = real retail price of beef per lb, ¢.

In this model β_2 , β_3 , β_4 , and β_5 are, respectively, the income, own-price, cross-price (pork), and cross-price (beef) elasticities. (Why?) According to economic theory,

$$\begin{aligned} \beta_2 &> 0 \\ \beta_3 &< 0 \\ \beta_4 &> 0, & \text{if chicken and pork are competing products} \\ &< 0, & \text{if chicken and pork are complementary products} \\ &= 0, & \text{if chicken and pork are unrelated products} \\ \beta_5 &> 0, & \text{if chicken and beef are competing products} \\ &< 0, & \text{if they are complementary products} \\ &= 0, & \text{if they are unrelated products} \end{aligned} \quad (8.7.20)$$

Suppose someone maintains that chicken and pork and beef are unrelated products in the sense that chicken consumption is not affected by the prices of pork and beef. In short,

$$H_0: \beta_4 = \beta_5 = 0 \quad (8.7.21)$$

Therefore, the constrained regression becomes

$$\ln Y_t = \beta_1 + \beta_2 \ln X_{2t} + \beta_3 \ln X_{3t} + u_t \quad (8.7.22)$$

Equation (8.7.19) is of course the unconstrained regression.

Using the data given in exercise 7.19, we obtain the following:

Unconstrained regression

$$\begin{aligned} \widehat{\ln Y}_t = & 2.1898 + 0.3425 \ln X_{2t} - 0.5046 \ln X_{3t} + 0.1485 \ln X_{4t} + 0.0911 \ln X_{5t} \\ & (0.1557) \quad (0.0833) \quad (0.1109) \quad (0.0997) \quad (0.1007) \\ & R_{UR}^2 = 0.9823 \quad (8.7.23) \end{aligned}$$

Constrained regression

$$\begin{aligned} \widehat{\ln Y}_t = & 2.0328 + 0.4515 \ln X_{2t} - 0.3772 \ln X_{3t} \\ & (0.1162) \quad (0.0247) \quad (0.0635) \quad (8.7.24) \\ & R_R^2 = 0.9801 \end{aligned}$$

(Continued)

EXAMPLE 8.4 (Continued)

where the figures in parentheses are the estimated standard errors. *Note:* The R^2 values of (8.7.23) and (8.7.24) are comparable since the dependent variable in the two models is the same.

Now the F ratio to test the hypothesis (8.7.21) is

$$F = \frac{(R_{UR}^2 - R_R^2) / m}{(1 - R_{UR}^2) / (n - k)} \quad (8.7.10)$$

The value of m in the present case is 2, since there are two restrictions involved: $\beta_4 = 0$ and $\beta_5 = 0$. The denominator df , $(n - k)$, is 18, since $n = 23$ and $k = 5$ (5 β coefficients).

Therefore, the F ratio is

$$\begin{aligned} F &= \frac{(0.9823 - 0.9801) / 2}{(1 - 0.9823) / 18} \\ &= 1.1224 \end{aligned} \quad (8.7.25)$$

which has the F distribution with 2 and 18 df .

At 5 percent, clearly this F value is not statistically significant [$F_{0.5}(2, 18) = 3.55$]. The p value is 0.3472. Therefore, there is no reason to reject the null hypothesis—the demand for chicken does not depend on pork and beef prices. In short, we can accept the constrained regression (8.7.24) as representing the demand function for chicken.

Notice that the demand function satisfies a priori economic expectations in that the own-price elasticity is negative and that the income elasticity is positive. However, the estimated price elasticity, in absolute value, is statistically less than unity, implying that the demand for chicken is price inelastic. (Why?) Also, the income elasticity, although positive, is also statistically less than unity, suggesting that chicken is not a luxury item; by convention, an item is said to be a luxury item if its income elasticity is greater than one.

8.8 TESTING FOR STRUCTURAL OR PARAMETER STABILITY OF REGRESSION MODELS: THE CHOW TEST

When we use a regression model involving time series data, it may happen that there is a **structural change** in the relationship between the regressand Y and the regressors. By structural change, we mean that the values of the parameters of the model do not remain the same through the entire time period. Sometime the structural change may be due to external forces (e.g., the oil embargoes imposed by the OPEC oil cartel in 1973 and 1979 or the Gulf War of 1990–1991), or due to policy changes (such as the switch from a fixed exchange-rate system to a flexible exchange-rate system around 1973) or action taken by Congress (e.g., the tax changes initiated by President Reagan in his two terms in office or changes in the minimum wage rate) or to a variety of other causes.

How do we find out that a structural change has in fact occurred? To be specific, consider the data given in Table 8.9. This table gives data on disposable personal income and personal savings, in billions of dollars, for the United States for the period 1970–1995. Suppose we want to estimate a

TABLE 8.9 SAVINGS AND PERSONAL DISPOSABLE INCOME (BILLIONS OF DOLLARS), UNITED STATES, 1970–1995

Observation	Savings	Income	Observation	Savings	Income
1970	61.0	727.1	1983	167.0	2522.4
1971	68.6	790.2	1984	235.7	2810.0
1972	63.6	855.3	1985	206.2	3002.0
1973	89.6	965.0	1986	196.5	3187.6
1974	97.6	1054.2	1987	168.4	3363.1
1975	104.4	1159.2	1988	189.1	3640.8
1976	96.4	1273.0	1989	187.8	3894.5
1977	92.5	1401.4	1990	208.7	4166.8
1978	112.6	1580.1	1991	246.4	4343.7
1979	130.1	1769.5	1992	272.6	4613.7
1980	161.8	1973.3	1993	214.4	4790.2
1981	199.1	2200.2	1994	189.4	5021.7
1982	205.5	2347.3	1995	249.3	5320.8

Source: Economic Report of the President, 1997, Table B-28, p. 332.

simple savings function that relates savings (Y) to disposable personal income DPI (X). Since we have the data, we can obtain an OLS regression of Y on X . But if we do that, we are maintaining that the relationship between savings and DPI has not changed much over the span of 26 years. That may be a tall assumption. For example, it is well known that in 1982 the United States suffered its worst peacetime recession. The civilian unemployment rate that year reached 9.7 percent, the highest since 1948. An event such as this might disturb the relationship between savings and DPI. To see if this happened, let us divide our sample data into two time periods: 1970–1981 and 1982–1995, the pre- and post-1982 recession periods.

Now we have three possible regressions:

$$\text{Time period 1970–1981: } Y_t = \lambda_1 + \lambda_2 X_t + u_{1t} \quad n_1 = 12 \quad (8.8.1)$$

$$\text{Time period 1982–1995: } Y_t = \gamma_1 + \gamma_2 X_t + u_{2t} \quad n_2 = 14 \quad (8.8.2)$$

$$\text{Time period 1970–1995: } Y_t = \alpha_1 + \alpha_2 X_t + u_t \quad n = (n_1 + n_2) = 26 \quad (8.8.3)$$

Regression (8.8.3) assumes that there is no difference between the two time periods and therefore estimates the relationship between savings and DPI for the entire time period consisting of 26 observations. In other words, this regression assumes that the intercept as well as the slope coefficient remains the same over the entire period; that is, there is no structural change. If this is in fact the situation, then $\alpha_1 = \lambda_1 = \gamma_1$ and $\alpha_2 = \lambda_2 = \gamma_2$.

Regressions (8.8.1) and (8.8.2) assume that the regressions in the two time periods are different; that is, the intercept and the slope coefficients are different, as indicated by the subscripted parameters. In the preceding regressions, the u 's represent the error terms and the n 's represent the number of observations.

For the data given in Table 8.9, the empirical counterparts of the preceding three regressions are as follows:

$$\begin{aligned}\hat{Y}_t &= 1.0161 + 0.0803 X_t \\ t &= (0.0873) \quad (9.6015) && \text{(8.8.1a)} \\ R^2 &= 0.9021 \quad \text{RSS}_1 = 1785.032 \quad \text{df} = 10\end{aligned}$$

$$\begin{aligned}\hat{Y}_t &= 153.4947 + 0.0148X_t \\ t &= (4.6922) \quad (1.7707) && \text{(8.8.2a)} \\ R^2 &= 0.2971 \quad \text{RSS}_2 = 10,005.22 \quad \text{df} = 12\end{aligned}$$

$$\begin{aligned}\hat{Y}_t &= 62.4226 + 0.0376 X_t + \dots \\ t &= (4.8917) \quad (8.8937) + \dots && \text{(8.8.3a)} \\ R^2 &= 0.7672 \quad \text{RSS}_3 = 23,248.30 \quad \text{df} = 24\end{aligned}$$

In the preceding regressions, RSS denotes the residual sum of squares, and the figures in parentheses are the estimated t values.

A look at the estimated regressions suggests that the relationship between savings and DPI is not the same in the two subperiods. The slope in the preceding savings-income regressions represents the **marginal propensity to save (MPS)**, that is, the (mean) change in savings as a result of a dollar's increase in disposable personal income. In the period 1970–1981 the MPS was about 0.08, whereas in the period 1982–1995 it was about 0.02. Whether this change was due to the economic policies pursued by President Reagan is hard to say. This further suggests that the **pooled regression** (8.8.3a)—that is, the one that pools all the 26 observations and runs a common regression, disregarding possible differences in the two subperiods may not be appropriate. Of course, the preceding statements need to be supported by appropriate statistical test(s). Incidentally, the scattergrams and the estimated regression lines are as shown in Figure 8.3.

Now the possible differences, that is, structural changes, may be caused by differences in the intercept or the slope coefficient or both. How do we find that out? A visual feeling about this can be obtained as shown in Figure 8.2. But it would be useful to have a formal test.

This is where the **Chow test** comes in handy.¹⁵ This test assumes that:

1. $u_{1t} \sim N(0, \sigma^2)$ and $u_{2t} \sim N(0, \sigma^2)$. That is, the error terms in the subperiod regressions are normally distributed with the same (homoscedastic) variance σ^2 .

¹⁵Gregory C. Chow, "Tests of Equality Between Sets of Coefficients in Two Linear Regressions," *Econometrica*, vol. 28, no. 3, 1960, pp. 591–605.

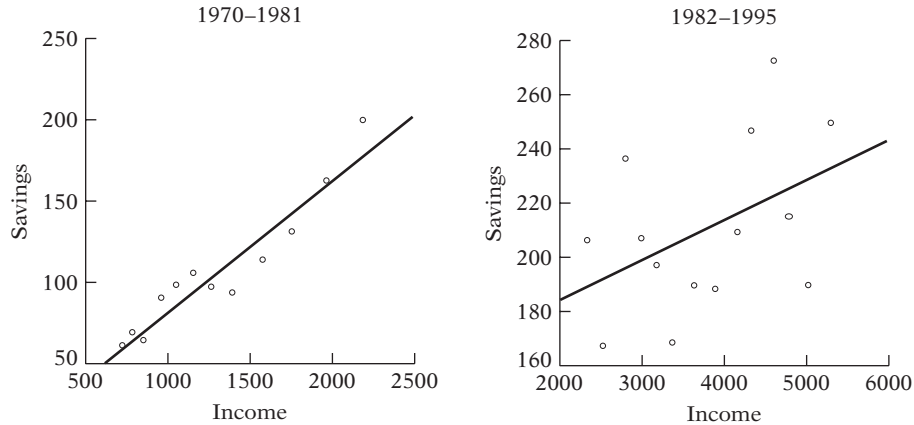


FIGURE 8.3

2. The two error terms u_{1t} and u_{2t} are independently distributed.

The mechanics of the Chow test are as follows:

1. Estimate regression (8.8.3), which is appropriate if there is no parameter instability, and obtain RSS_3 with $df = (n_1 + n_2 - k)$, where k is the number of parameters estimated, 2 in the present case. For our example $RSS_3 = 23,248.30$. We call RSS_3 the **restricted residual sum of squares (RSS_R)** because it is obtained by imposing the restrictions that $\lambda_1 = \gamma_1$ and $\lambda_2 = \gamma_2$, that is, the subperiod regressions are not different.

2. Estimate (8.8.1) and obtain its residual sum of squares, RSS_1 , with $df = (n_1 - k)$. In our example, $RSS_1 = 1785.032$ and $df = 10$.

3. Estimate (8.8.2) and obtain its residual sum of squares, RSS_2 , with $df = (n_2 - k)$. In our example, $RSS_2 = 10,005.22$ with $df = 12$.

4. Since the two sets of samples are deemed independent, we can add RSS_1 and RSS_2 to obtain what may be called the **unrestricted residual sum of squares (RSS_{UR})**, that is, obtain:

$$RSS_{UR} = RSS_1 + RSS_2 \quad \text{with } df = (n_1 + n_2 - 2k)$$

In the present case,

$$RSS_{UR} = (1785.032 + 10,005.22) = 11,790.252$$

5. Now the idea behind the Chow test is that if in fact there is no structural change [i.e., regressions (8.8.1) and (8.8.2) are essentially the same], then the RSS_R and RSS_{UR} should not be statistically different. Therefore, if we form the following ratio:

$$F = \frac{(RSS_R - RSS_{UR})/k}{(RSS_{UR})/(n_1 + n_2 - 2k)} \sim F_{[k, (n_1 + n_2 - 2k)]} \quad (8.8.4)$$

then Chow has shown that under the null hypothesis the regressions (8.8.1) and (8.8.2) are (statistically) the same (i.e., no structural change or break) and the F ratio given above follows the F distribution with k and $(n_1 + n_2 - 2k)$ df in the numerator and denominator, respectively.

6. Therefore, we do not reject the null hypothesis of *parameter stability* (i.e., no structural change) if the computed F value in an application does not exceed the critical F value obtained from the F table at the chosen level of significance (or the p value). In this case we may be justified in using the pooled (restricted?) regression (8.8.3). Contrarily, if the computed F value exceeds the critical F value, we reject the hypothesis of parameter stability and conclude that the regressions (8.8.1) and (8.8.2) are different, in which case the pooled regression (8.8.3) is of dubious value, to say the least.

Returning to our example, we find that

$$\begin{aligned} F &= \frac{(23,248.30 - 11,790.252)/2}{(11,790.252)/22} \\ &= 10.69 \end{aligned} \quad (8.8.5)$$

From the F tables, we find that for 2 and 22 df the 1 percent critical F value is 5.72. Therefore, the probability of obtaining an F value of as much as or greater than 10.69 is much smaller than 1 percent; actually the p value is only 0.00057.

The Chow test therefore seems to support our earlier hunch that the savings–income relation has undergone a structural change in the United States over the period 1970–1995, assuming that the assumptions underlying the test are fulfilled. We will have more to say about this shortly.

Incidentally, note that the Chow test can be easily generalized to handle cases of more than one structural break. For example, if we believe that the savings–income relation changed after President Clinton took office in January 1992, we could divide our sample into three periods: 1970–1981, 1982–1991, 1992–1995, and carry out the Chow test. Of course, we will have four RSS terms, one for each subperiod and one for the pooled data. But the logic of the test remains the same. Data through 2001 are now available to extend the last period to 2001.

There are some caveats about the Chow test that must be kept in mind:

1. The assumptions underlying the test must be fulfilled. For example, one should find out if the error variances in the regressions (8.8.1) and (8.8.2) are the same. We will discuss this point shortly.

2. The Chow test will tell us only if the two regressions (8.8.1) and (8.8.2) are different, without telling us whether the difference is on account of the intercepts, or the slopes, or both. But in Chapter 9, on dummy variables, we will see how we can answer this question.

3. The Chow test assumes that we know the point(s) of structural break. In our example, we assumed it to be in 1982. However, if it is not possible to

determine when the structural change actually took place, we may have to use other methods.¹⁶

Before we leave the Chow test and our savings–income regression, let us examine one of the assumptions underlying the Chow test, namely, that the error variances in the two periods are the same. Since we cannot observe the true error variances, we can obtain their estimates from the RSS given in the regressions (8.8.1a) and (8.8.2a), namely,

$$\hat{\sigma}_1^2 = \frac{\text{RSS}_1}{n_1 - 2} = \frac{1785.032}{10} = 178.5032 \quad (8.8.6)$$

$$\hat{\sigma}_2^2 = \frac{\text{RSS}_2}{n_2 - 2} = \frac{10,005.22}{14 - 2} = 833.7683 \quad (8.8.7)$$

Notice that, since there are two parameters estimated in each equation, we deduct 2 from the number of observations to obtain the df. Given the assumptions underlying the Chow test, $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ are unbiased estimators of the true variances in the two subperiods. As a result, it can be shown that $\sigma_1^2 = \sigma_2^2$ —that is, the variances in the two subpopulations are the same (as assumed by the Chow test)—then it can be shown that

$$\frac{(\hat{\sigma}_1^2/\sigma_1^2)}{(\hat{\sigma}_2^2/\sigma_2^2)} \sim F_{(n_1-k), (n_2-k)} \quad (8.8.8)$$

follows the F distribution with $(n_1 - k)$ and $(n_2 - k)$ df in the numerator and the denominator, respectively, in our example $k = 2$, since there are only two parameters in each subregression.

Of course, $\sigma_1^2 = \sigma_2^2$, the preceding F test reduces to computing

$$F = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} \quad (8.8.9)$$

Note: By convention we put the larger of the two estimated variances in the numerator. (See **Appendix A** for the details of the F and other probability distributions.)

Computing this F in an application and comparing it with the critical F value with the appropriate df, one can decide to reject or not reject the null hypothesis that the variances in the two subpopulations are the same. If the null hypothesis is not rejected, then one can use the Chow test.

Returning to our savings–income regression, we obtain the following result:

$$F = \frac{833.7683}{178.5032} = 4.6701 \quad (8.8.10)$$

¹⁶For a detailed discussion, see William H. Greene, *Econometric Analysis*, 4th ed., Prentice Hall, Englewood Cliffs, N.J., 2000, pp. 293–297.

Under the null hypothesis of equality of variances in the two subpopulations, this F value follows the F distribution with 12 and 10 df, in the numerator and denominator, respectively. (*Note:* We have put the larger of the two estimated variances in the numerator.) From the F tables in **Appendix D**, we see that the 5 and 1 percent critical F values for 12 and 10 df are 2.91 and 4.71, respectively. The computed F value is significant at the 5 percent level and is almost significant at the 1 percent level. Thus, our conclusion would be that the two subpopulation variances are not the same and, therefore, strictly speaking we should not use the Chow test.

Our purpose here has been to demonstrate the mechanics of the Chow test, which is used popularly in applied work. If the error variances in the two subpopulations are heteroscedastic, the Chow test can be modified. But the procedure is beyond the scope of this book.¹⁷

Another point we made earlier was that the Chow test is sensitive to the choice of the time at which the regression parameters might have changed. In our example, we assumed that the change probably took place in the recession year of 1982. If we had assumed it to be 1981, when Ronald Reagan began his presidency, we might find that the computed F value is different. As a matter of fact, in exercise 8.34 the reader is asked to check this out.

If we do not want to choose the point at which the break in the underlying relationship might have occurred, we could choose alternative methods, such as the **recursive residual test**. We will take this topic up in Chapter 13, the chapter on model specification analysis.

8.9 PREDICTION WITH MULTIPLE REGRESSION

In Section 5.10 we showed how the estimated two-variable regression model can be used for (1) *mean prediction*, that is, predicting the point on the population regression function (PRF), as well as for (2) *individual prediction*, that is, predicting an individual value of Y given the value of the regressor $X = X_0$, where X_0 is the specified numerical value of X .

The estimated multiple regression too can be used for similar purposes, and the procedure for doing that is a straightforward extension of the two-variable case, except the formulas for estimating the variances and standard errors of the forecast value [comparable to (5.10.2) and (5.10.6) of the two-variable model] are rather involved and are better handled by the matrix methods discussed in **Appendix C**. Of course, most standard regression packages can do this routinely, so there is no need to look up the matrix formulation. It is given in **Appendix C** for the benefit of the mathematically inclined students. This appendix also gives a fully worked out example.

¹⁷For a discussion of the Chow test under heteroscedasticity, see William H. Greene, *Econometric Analysis*, 4th ed., Prentice Hall, Englewood Cliffs, N.J., 2000, pp. 292–293, and Adrian C. Darnell, *A Dictionary of Econometrics*, Edward Elgar, U.K., 1994, p. 51.

***8.10 THE TROIKA OF HYPOTHESIS TESTS: THE LIKELIHOOD RATIO (LR), WALD (W), AND LAGRANGE MULTIPLIER (LM) TESTS¹⁸**

In this and the previous chapters we have, by and large, used the t , F , and chi-square tests to test a variety of hypotheses in the context of linear (in-parameter) regression models. But once we go beyond the somewhat comfortable world of linear regression models, we need method(s) to test hypotheses that can handle regression models, linear or not.

The well-known trinity of **likelihood, Wald, and Lagrange multiplier tests** can accomplish this purpose. The interesting thing to note is that *asymptotically* (i.e., in large samples) all three tests are equivalent in that the test statistic associated with each of these tests follows the chi-square distribution.

Although we will discuss the **likelihood ratio test** in the appendix to this chapter, in general we will not use these tests in this textbook for the pragmatic reason that in small, or finite, samples, which is unfortunately what most researchers deal with, the F test that we have used so far will suffice. As Davidson and MacKinnon note:

For linear regression models, with or without normal errors, there is of course no need to look at LM, W and LR at all, since no information is gained from doing so over and above what is already contained in F .¹⁹

***8.11 TESTING THE FUNCTIONAL FORM OF REGRESSION: CHOOSING BETWEEN LINEAR AND LOG-LINEAR REGRESSION MODELS**

The choice between a linear regression model (the regressand is a linear function of the regressors) or a log-linear regression model (the log of the regressand is a function of the logs of the regressors) is a perennial question in empirical analysis. We can use a test proposed by MacKinnon, White, and Davidson, which for brevity we call the **MWD test** to choose between the two models.²⁰

To illustrate this test, assume the following

H_0 : *Linear Model*: Y is a linear function of regressors, the X 's.

H_1 : *Log-Linear Model*: $\ln Y$ is a linear function of logs of regressors, the logs of X 's.

where, as usual, H_0 and H_1 denote the null and alternative hypotheses.

*Optional.

¹⁸For an accessible discussion, see A. Buse, "The Likelihood Ratio, Wald and Lagrange Multiplier Tests: An Expository Note," *American Statistician*, vol. 36, 1982, pp. 153–157.

¹⁹Russell Davidson and James G. MacKinnon, *Estimation and Inference in Econometrics*, Oxford University Press, New York, 1993, p. 456.

²⁰J. MacKinnon, H. White, and R. Davidson, "Tests for Model Specification in the Presence of Alternative Hypothesis; Some Further Results." *Journal of Econometrics*, vol. 21, 1983, pp. 53–70. A similar test is proposed in A. K. Bera and C. M. Jarque, "Model Specification Tests: A Simultaneous Approach," *Journal of Econometrics*, vol. 20, 1982, pp. 59–82.

The MWD test involves the following steps²¹:

Step I: Estimate the linear model and obtain the estimated Y values. Call them \hat{Y}_f (i.e., \hat{Y}).

Step II: Estimate the log-linear model and obtain the estimated $\ln Y$ values; call them $\ln \hat{Y}_f$ (i.e., $\widehat{\ln Y}$).

Step III: Obtain $Z_1 = (\ln \hat{Y}_f - \ln Y)$.

Step IV: Regress Z_1 on X 's and Z_1 obtained in Step III. Reject H_0 if the coefficient of Z_1 is statistically significant by the usual t test.

Step V: Obtain $Z_2 = (\text{antilog of } \ln \hat{Y}_f - Y)$.

Step VI: Regress Z_2 on the logs of X 's and Z_2 . Reject H_1 if the coefficient of Z_2 is statistically significant by the usual t test.

Although the MWD test seems involved, the logic of the test is quite simple. If the linear model is in fact the correct model, the constructed variable Z_1 should not be statistically significant in Step IV, for in that case the estimated Y values from the linear model and those estimated from the log-linear model (after taking their antilog values for comparative purposes) should not be different. The same comment applies to the alternative hypothesis H_1 .

EXAMPLE 8.5

THE DEMAND FOR ROSES

Refer to exercise 7.16 where we have presented data on the demand for roses in the Detroit metropolitan area for the period 1971–II to 1975–II. For illustrative purposes, we will consider the demand for roses as a function only of the prices of roses and carnations, leaving out the income variable for the time being. Now we consider the following models:

$$\text{Linear model: } Y_t = \alpha_1 + \alpha_2 X_{2t} + \alpha_3 X_{3t} + u_t \quad (8.11.1)$$

$$\text{Log-linear model: } \ln Y_t = \beta_1 + \beta_2 \ln X_{2t} + \beta_3 \ln X_{3t} + u_t \quad (8.11.2)$$

where Y is the quantity of roses in dozens, X_2 is the average wholesale price of roses (\$/dozen), and X_3 is the average wholesale price of carnations (\$/dozen). A priori, α_2 and β_2 are expected to be negative (why?), and α_3 and β_3 are expected to be positive (why?). As we know, the slope coefficients in the log-linear model are elasticity coefficients.

The regression results are as follows:

$$\begin{aligned} \hat{Y}_t &= 9734.2176 - 3782.1956X_{2t} + 2815.2515X_{3t} \\ t &= (3.3705) \quad (-6.6069) \quad (2.9712) \end{aligned} \quad (8.11.3)$$

$$F = 21.84 \quad R^2 = 0.77096$$

$$\begin{aligned} \widehat{\ln Y}_t &= 9.2278 - 1.7607 \ln X_{2t} + 1.3398 \ln X_{3t} \\ t &= (16.2349) \quad (-5.9044) \quad (2.5407) \end{aligned} \quad (8.11.4)$$

$$F = 17.50 \quad R^2 = 0.7292$$

(Continued)

²¹This discussion is based on William H. Greene, *ET. The Econometrics Toolkit Version 3*, Econometric Software, Bellport, New York, 1992, pp. 245–246.

EXAMPLE 8.5 (Continued)

As these results show, both the linear and the log-linear models seem to fit the data reasonably well: The parameters have the expected signs and the t and R^2 values are statistically significant.

To decide between these models on the basis of the **MWD test**, we first test the hypothesis that the true model is linear. Then, following Step IV of the test, we obtain the following regression:

$$\begin{aligned} \hat{Y}_t &= 9727.5685 - 3783.0623X_{2t} + 2817.7157X_{3t} + 85.2319Z_{1t} \\ t &= (3.2178) \quad (-6.3337) \quad (2.8366) \quad (0.0207) \quad \mathbf{(8.11.5)} \\ & \quad \quad \quad F = 13.44 \quad R^2 = 0.7707 \end{aligned}$$

Since the coefficient of Z_1 is not statistically significant (the p value of the estimated t is 0.98), we do not reject the hypothesis that the true model is linear.

Suppose we switch gears and assume that the true model is log-linear. Following step VI of the MWD test, we obtain the following regression results:

$$\begin{aligned} \widehat{\ln Y}_t &= 9.1486 - 1.9699 \ln X_t + 1.5891 \ln X_{2t} - 0.0013Z_{2t} \\ t &= (17.0825) \quad (-6.4189) \quad (3.0728) \quad (-1.6612) \quad \mathbf{(8.11.6)} \\ & \quad \quad \quad F = 14.17 \quad R^2 = 0.7798 \end{aligned}$$

The coefficient of Z_2 is statistically significant at about the 12 percent level (p value is 0.1225). Therefore, we can reject the hypothesis that the true model is log-linear at this level of significance. Of course, if one sticks to the conventional 1 or 5 percent significance levels, then one cannot reject the hypothesis that the true model is log-linear. As this example shows, it is quite possible that in a given situation we cannot reject either of the specifications.

8.12 SUMMARY AND CONCLUSIONS

1. This chapter extended and refined the ideas of interval estimation and hypothesis testing first introduced in Chapter 5 in the context of the two-variable linear regression model.

2. In a multiple regression, testing the *individual significance* of a partial regression coefficient (using the t test) and testing the *overall significance* of the regression (i.e., H_0 : all partial slope coefficients are zero or $R^2 = 0$) are not the same thing.

3. In particular, the finding that one or more partial regression coefficients are statistically insignificant on the basis of the *individual t test* does not mean that all partial regression coefficients are also (collectively) statistically insignificant. The latter hypothesis can be tested only by the F test.

4. The **F test** is versatile in that it can test a variety of hypotheses, such as whether (1) an individual regression coefficient is statistically significant, (2) all partial slope coefficients are zero, (3) two or more coefficients are statistically equal, (4) the coefficients satisfy some linear restrictions, and (5) there is structural stability of the regression model.

5. As in the two-variable case, the multiple regression model can be used for the purpose of mean and or individual prediction.