5

TWO-VARIABLE REGRESSION: INTERVAL ESTIMATION AND HYPOTHESIS TESTING

> Beware of testing too many hypotheses; the more you torture the data, the more likely they are to confess, but confession obtained under duress may not be admissible in the court of scientific opinion.[1]

As pointed out in Chapter 4, estimation and hypothesis testing constitute the two major branches of classical statistics. The theory of estimation consists of two parts: point estimation and interval estimation. We have discussed point estimation thoroughly in the previous two chapters where we introduced the OLS and ML methods of point estimation. In this chapter we first consider interval estimation and then take up the topic of hypothesis testing, a topic intimately related to interval estimation.

5.1 STATISTICAL PREREQUISITES

Before we demonstrate the actual mechanics of establishing confidence intervals and testing statistical hypotheses, it is assumed that the reader is familiar with the fundamental concepts of probability and statistics. Although not a substitute for a basic course in statistics, **Appendix A** provides the essentials of statistics with which the reader should be totally familiar. Key concepts such as **probability, probability distributions, Type I and Type II errors, level of significance, power of a statistical test,** and **confidence interval** are crucial for understanding the material covered in this and the following chapters.

[1]Stephen M. Stigler, "Testing Hypothesis or Fitting Models? Another Look at Mass Extinctions," in Matthew H. Nitecki and Antoni Hoffman, eds., *Neutral Models in Biology,* Oxford University Press, Oxford, 1987, p. 148.

5.2 INTERVAL ESTIMATION: SOME BASIC IDEAS

To fix the ideas, consider the hypothetical consumption-income example of Chapter 3. Equation (3.6.2) shows that the estimated marginal propensity to consume (MPC) β_2 is 0.5091, which is a single (point) estimate of the unknown population MPC β_2. How reliable is this estimate? As noted in Chapter 3, because of sampling fluctuations, a single estimate is likely to differ from the true value, although in repeated sampling its mean value is expected to be equal to the true value. [*Note:* $E(\hat{\beta}_2) = \beta_2$.] Now in statistics the reliability of a point estimator is measured by its standard error. Therefore, instead of relying on the point estimate alone, we may construct an interval around the point estimator, say within two or three standard errors on either side of the point estimator, such that this interval has, say, 95 percent probability of including the true parameter value. This is roughly the idea behind **interval estimation.**

To be more specific, assume that we want to find out how "close" is, say, $\hat{\beta}_2$ to β_2. For this purpose we try to find out two positive numbers δ and α, the latter lying between 0 and 1, such that the probability that the **random interval** $(\hat{\beta}_2 - \delta, \hat{\beta}_2 + \delta)$ contains the true β_2 is $1 - \alpha$. Symbolically,

$$\Pr\,(\hat{\beta}_2 - \delta \leq \beta_2 \leq \hat{\beta}_2 + \delta) = 1 - \alpha \qquad \textbf{(5.2.1)}$$

Such an interval, if it exists, is known as a **confidence interval;** $1 - \alpha$ is known as the **confidence coefficient;** and α $(0 < \alpha < 1)$ is known as the **level of significance.**[2] The endpoints of the confidence interval are known as the **confidence limits** (also known as *critical* values), $\hat{\beta}_2 - \delta$ being the **lower confidence** *limit* and $\hat{\beta}_2 + \delta$ the **upper confidence** *limit*. In passing, note that in practice α and $1 - \alpha$ are often expressed in percentage forms as 100α and $100(1 - \alpha)$ percent.

Equation (5.2.1) shows that an **interval estimator,** in contrast to a point estimator, is an interval constructed in such a manner that it has a specified probability $1 - \alpha$ of including within its limits the true value of the parameter. For example, if $\alpha = 0.05$, or 5 percent, (5.2.1) would read: The probability that the (random) interval shown there includes the true β_2 is 0.95, or 95 percent. The interval estimator thus gives a range of values within which the true β_2 may lie.

It is very important to know the following aspects of interval estimation:

1. Equation (5.2.1) does not say that the probability of β_2 lying between the given limits is $1 - \alpha$. Since β_2, although an unknown, is assumed to be some fixed number, either it lies in the interval or it does not. What (5.2.1)

[2]Also known as the **probability of committing a Type I error.** A Type I error consists in rejecting a true hypothesis, whereas a Type II error consists in accepting a false hypothesis. (This topic is discussed more fully in **App. A**.) The symbol α is also known as the **size of the (statistical) test.**

states is that, for the method described in this chapter, the probability of constructing an interval that contains β_2 is $1 - \alpha$.

2. The interval (5.2.1) is a **random interval;** that is, it will vary from one sample to the next because it is based on $\hat{\beta}_2$, which is random. (Why?)

3. Since the confidence interval is random, the probability statements attached to it should be understood in the long-run sense, that is, repeated sampling. More specifically, (5.2.1) means: If in repeated sampling confidence intervals like it are constructed a great many times on the $1 - \alpha$ probability basis, then, in the long run, on the average, such intervals will enclose in $1 - \alpha$ of the cases the true value of the parameter.

4. As noted in 2, the interval (5.2.1) is random so long as $\hat{\beta}_2$ is not known. But once we have a specific sample and once we obtain a specific numerical value of $\hat{\beta}_2$, the interval (5.2.1) is no longer random; it is fixed. In this case, we **cannot** make the probabilistic statement (5.2.1); that is, we cannot say that the probability is $1 - \alpha$ that a given *fixed* interval includes the true β_2. In this situation β_2 is either in the fixed interval or outside it. Therefore, the probability is either 1 or 0. Thus, for our hypothetical consumption-income example, if the 95% confidence interval were obtained as ($0.4268 \leq \beta_2 \leq 0.5914$), as we do shortly in (5.3.9), we **cannot** say the probability is 95% that this interval includes the true β_2. That probability is either 1 or 0.

How are the confidence intervals constructed? From the preceding discussion one may expect that if the **sampling or probability distributions** of the estimators are known, one can make confidence interval statements such as (5.2.1). In Chapter 4 we saw that under the assumption of normality of the disturbances u_i the OLS estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ are themselves normally distributed and that the OLS estimator $\hat{\sigma}^2$ is related to the χ^2 (chi-square) distribution. It would then seem that the task of constructing confidence intervals is a simple one. And it is!

5.3 CONFIDENCE INTERVALS FOR REGRESSION COEFFICIENTS β_1 AND β_2

Confidence Interval for β_2

It was shown in Chapter 4, Section 4.3, that, with the normality assumption for u_i, the OLS estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ are themselves normally distributed with means and variances given therein. Therefore, for example, the variable

$$Z = \frac{\hat{\beta}_2 - \beta_2}{\text{se}\,(\hat{\beta}_2)} = \frac{(\hat{\beta}_2 - \beta_2)\sqrt{\sum x_i^2}}{\sigma} \tag{5.3.1}$$

as noted in (4.3.6), is a standardized normal variable. It therefore seems that we can use the normal distribution to make probabilistic statements about β_2 provided the true population variance σ^2 is known. If σ^2 is known, an important property of a normally distributed variable with mean μ and variance σ^2 is that the area under the normal curve between $\mu \pm \sigma$ is about 68 percent, that between the limits $\mu \pm 2\sigma$ is about 95 percent, and that between $\mu \pm 3\sigma$ is about 99.7 percent.

But σ^2 is rarely known, and in practice it is determined by the unbiased estimator $\hat{\sigma}^2$. If we replace σ by $\hat{\sigma}$, (5.3.1) may be written as

$$\begin{aligned} t &= \frac{\hat{\beta}_2 - \beta_2}{\text{se}\,(\hat{\beta}_2)} = \frac{\text{estimator} - \text{parameter}}{\text{estimated standard error of estimator}} \\ &= \frac{(\hat{\beta}_2 - \beta_2)\sqrt{\sum x_i^2}}{\hat{\sigma}} \end{aligned} \tag{5.3.2}$$

where the se $(\hat{\beta}_2)$ now refers to the estimated standard error. It can be shown (see Appendix 5A, Section 5A.2) that the t variable thus defined follows the t distribution with $n - 2$ df. [Note the difference between (5.3.1) and (5.3.2).] Therefore, instead of using the normal distribution, we can use the t distribution to establish a confidence interval for β_2 as follows:

$$\Pr\,(-t_{\alpha/2} \leq t \leq t_{\alpha/2}) = 1 - \alpha \tag{5.3.3}$$

where the t value in the middle of this double inequality is the t value given by (5.3.2) and where $t_{\alpha/2}$ is the value of the t variable obtained from the t distribution for $\alpha/2$ level of significance and $n - 2$ df; it is often called the **critical** t value at $\alpha/2$ level of significance. Substitution of (5.3.2) into (5.3.3) yields

$$\Pr\left[-t_{\alpha/2} \leq \frac{\hat{\beta}_2 - \beta_2}{\text{se}\,(\hat{\beta}_2)} \leq t_{\alpha/2}\right] = 1 - \alpha \tag{5.3.4}$$

Rearranging (5.3.4), we obtain

$$\Pr\,[\hat{\beta}_2 - t_{\alpha/2}\ \text{se}\,(\hat{\beta}_2) \leq \beta_2 \leq \hat{\beta}_2 + t_{\alpha/2}\ \text{se}\,(\hat{\beta}_2)] = 1 - \alpha \tag{5.3.5}$$[3]

[3]Some authors prefer to write (5.3.5) with the df explicitly indicated. Thus, they would write

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

But for simplicity we will stick to our notation; the context clarifies the appropriate df involved.

Equation (5.3.5) provides a $100(1-\alpha)$ percent **confidence interval** for β_2, which can be written more compactly as

$$100(1-\alpha)\% \text{ confidence interval for } \beta_2:$$

$$\hat{\beta}_2 \pm t_{\alpha/2}\, \text{se}\,(\hat{\beta}_2) \tag{5.3.6}$$

Arguing analogously, and using (4.3.1) and (4.3.2), we can then write:

$$\Pr\,[\hat{\beta}_1 - t_{\alpha/2}\, \text{se}\,(\hat{\beta}_1) \leq \beta_1 \leq \hat{\beta}_1 + t_{\alpha/2}\, \text{se}\,(\hat{\beta}_1)] = 1-\alpha \tag{5.3.7}$$

or, more compactly,

$$100(1-\alpha)\% \text{ confidence interval for } \beta_1:$$

$$\hat{\beta}_1 \pm t_{\alpha/2}\, \text{se}\,(\hat{\beta}_1) \tag{5.3.8}$$

Notice an important feature of the confidence intervals given in (5.3.6) and (5.3.8): In both cases *the width of the confidence interval is proportional to the standard error of the estimator.* That is, the larger the standard error, the larger is the width of the confidence interval. Put differently, the larger the standard error of the estimator, the greater is the uncertainty of estimating the true value of the unknown parameter. Thus, the standard error of an estimator is often described as a measure of the **precision** of the estimator, i.e., how precisely the estimator measures the true population value.

Returning to our illustrative consumption–income example, in Chapter 3 (Section 3.6) we found that $\hat{\beta}_2 = 0.5091$, $\text{se}\,(\hat{\beta}_2) = 0.0357$, and $\text{df} = 8$. If we assume $\alpha = 5\%$, that is, 95% confidence coefficient, then the t table shows that for 8 df the **critical** $t_{\alpha/2} = t_{0.025} = 2.306$. Substituting these values in (5.3.5), the reader should verify that the 95% confidence interval for β_2 is as follows:

$$0.4268 \leq \beta_2 \leq 0.5914 \tag{5.3.9}$$

Or, using (5.3.6), it is

$$0.5091 \pm 2.306(0.0357)$$

that is,

$$0.5091 \pm 0.0823 \tag{5.3.10}$$

The interpretation of this confidence interval is: Given the confidence coefficient of 95%, in the long run, in 95 out of 100 cases intervals like

(0.4268, 0.5914) will contain the true β_2. But, as warned earlier, we cannot say that the probability is 95 percent that the specific interval (0.4268 to 0.5914) contains the true β_2 because this interval is now fixed and no longer random; therefore, β_2 either lies in it or does not: The probability that the specified fixed interval includes the true β_2 is therefore 1 or 0.

Confidence Interval for β_1

Following (5.3.7), the reader can easily verify that the 95% confidence interval for β_1 of our consumption–income example is

$$9.6643 \leq \beta_1 \leq 39.2448 \tag{5.3.11}$$

Or, using (5.3.8), we find it is

$$24.4545 \pm 2.306(6.4138)$$

that is,

$$24.4545 \pm 14.7902 \tag{5.3.12}$$

Again you should be careful in interpreting this confidence interval. In the long run, in 95 out of 100 cases intervals like (5.3.11) will contain the true β_1; the probability that this particular fixed interval includes the true β_1 is either 1 or 0.

Confidence Interval for β_1 and β_2 Simultaneously

There are occasions when one needs to construct a *joint confidence interval* for β_1 and β_2 such that with a confidence coefficient $(1 - \alpha)$, say, 95%, that interval includes β_1 and β_2 simultaneously. Since this topic is involved, the interested reader may want to consult appropriate references.[4] We will touch on this topic briefly in Chapters 8 and 10.

5.4 CONFIDENCE INTERVAL FOR σ^2

As pointed out in Chapter 4, Section 4.3, under the normality assumption, the variable

$$\chi^2 = (n-2)\frac{\hat{\sigma}^2}{\sigma^2} \tag{5.4.1}$$

[4]For an accessible discussion, see John Neter, William Wasserman, and Michael H. Kutner, *Applied Linear Regression Models,* Richard D. Irwin, Homewood, Ill., 1983, Chap. 5.

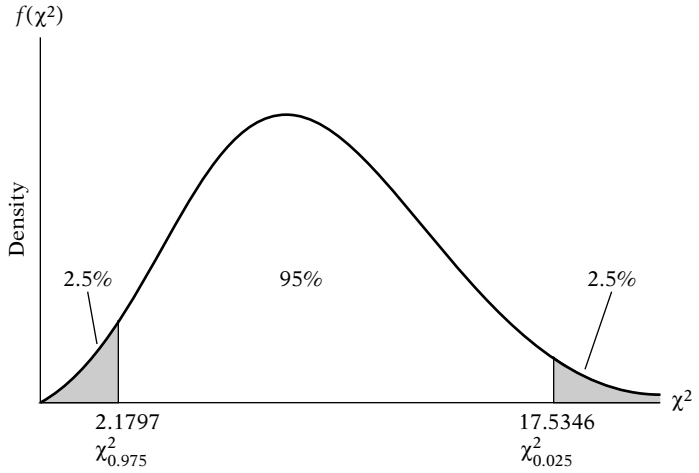


FIGURE 5.1 The 95% confidence interval for χ^2 (8 df).

follows the χ^2 distribution with $n - 2$ df.[5] Therefore, we can use the χ^2 distribution to establish a confidence interval for σ^2

$$\Pr\left(\chi^2_{1-\alpha/2} \leq \chi^2 \leq \chi^2_{\alpha/2}\right) = 1 - \alpha \qquad \textbf{(5.4.2)}$$

where the χ^2 value in the middle of this double inequality is as given by (5.4.1) and where $\chi^2_{1-\alpha/2}$ and $\chi^2_{\alpha/2}$ are two values of χ^2 (the **critical** χ^2 values) obtained from the chi-square table for $n - 2$ df in such a manner that they cut off $100(\alpha/2)$ percent tail areas of the χ^2 distribution, as shown in Figure 5.1.

Substituting χ^2 from (5.4.1) into (5.4.2) and rearranging the terms, we obtain

$$\Pr\left[(n-2)\frac{\hat{\sigma}^2}{\chi^2_{\alpha/2}} \leq \sigma^2 \leq (n-2)\frac{\hat{\sigma}^2}{\chi^2_{1-\alpha/2}}\right] = 1 - \alpha \qquad \textbf{(5.4.3)}$$

which gives the $100(1 - \alpha)\%$ confidence interval for σ^2.

To illustrate, consider this example. From Chapter 3, Section 3.6, we obtain $\hat{\sigma}^2 = 42.1591$ and df $= 8$. If α is chosen at 5 percent, the chi-square table for 8 df gives the following critical values: $\chi^2_{0.025} = 17.5346$, and $\chi^2_{0.975} = 2.1797$. These values show that the probability of a chi-square value exceeding 17.5346 is 2.5 percent and that of 2.1797 is 97.5 percent. Therefore, the interval between these two values is the 95% confidence interval for χ^2, as shown diagrammatically in Figure 5.1. (Note the skewed characteristic of the chi-square distribution.)

[5]For proof, see Robert V. Hogg and Allen T. Craig, *Introduction to Mathematical Statistics,* 2d ed., Macmillan, New York, 1965, p. 144.

Substituting the data of our example into (5.4.3), the reader should verify that the 95% confidence interval for σ^2 is as follows:

$$19.2347 \leq \sigma^2 \leq 154.7336 \qquad \textbf{(5.4.4)}$$

The interpretation of this interval is: If we establish 95% confidence limits on σ^2 and if we maintain a priori that these limits will include true σ^2, we shall be right in the long run 95 percent of the time.

5.5 HYPOTHESIS TESTING: GENERAL COMMENTS

Having discussed the problem of point and interval estimation, we shall now consider the topic of hypothesis testing. In this section we discuss briefly some general aspects of this topic; **Appendix A** gives some additional details.

The problem of statistical hypothesis testing may be stated simply as follows: *Is a given observation or finding compatible with some stated hypothesis or not?* The word "compatible," as used here, means "sufficiently" close to the hypothesized value so that we do not reject the stated hypothesis. Thus, if some theory or prior experience leads us to believe that the true slope coefficient β_2 of the consumption–income example is unity, is the observed $\hat{\beta}_2 = 0.5091$ obtained from the sample of Table 3.2 consistent with the stated hypothesis? If it is, we do not reject the hypothesis; otherwise, we may reject it.

In the language of statistics, the stated hypothesis is known as the **null hypothesis** and is denoted by the symbol H_0. The null hypothesis is usually tested against an **alternative hypothesis** (also known as **maintained hypothesis**) denoted by H_1, which may state, for example, that true β_2 is different from unity. The alternative hypothesis may be **simple** or **composite.**[6] For example, $H_1: \beta_2 = 1.5$ is a simple hypothesis, but $H_1: \beta_2 \neq 1.5$ is a composite hypothesis.

The theory of hypothesis testing is concerned with developing rules or procedures for deciding whether to reject or not reject the null hypothesis. There are two *mutually complementary* approaches for devising such rules, namely, **confidence interval** and **test of significance.** Both these approaches predicate that the variable (statistic or estimator) under consideration has some probability distribution and that hypothesis testing involves making statements or assertions about the value(s) of the parameter(s) of such distribution. For example, we know that with the normality assumption $\hat{\beta}_2$ is normally distributed with mean equal to β_2 and variance given by (4.3.5). If we hypothesize that $\beta_2 = 1$, we are making an assertion about one

[6] A statistical hypothesis is called a **simple hypothesis** if it specifies the precise value(s) of the parameter(s) of a probability density function; otherwise, it is called a **composite hypothesis.** For example, in the normal pdf $(1/\sigma\sqrt{2\pi})\exp\{-\frac{1}{2}[(X-\mu)/\sigma]^2\}$, if we assert that $H_1: \mu = 15$ and $\sigma = 2$, it is a simple hypothesis; but if $H_1: \mu = 15$ and $\sigma > 15$, it is a composite hypothesis, because the standard deviation does not have a specific value.

of the parameters of the normal distribution, namely, the mean. Most of the statistical hypotheses encountered in this text will be of this type—making assertions about one or more values of the parameters of some assumed probability distribution such as the normal, F, t, or χ^2. How this is accomplished is discussed in the following two sections.

5.6 HYPOTHESIS TESTING: THE CONFIDENCE-INTERVAL APPROACH

Two-Sided or Two-Tail Test

To illustrate the confidence-interval approach, once again we revert to the consumption–income example. As we know, the estimated marginal propensity to consume (MPC), $\hat{\beta}_2$, is 0.5091. Suppose we postulate that

$$H_0: \beta_2 = 0.3$$
$$H_1: \beta_2 \neq 0.3$$

that is, the true MPC is 0.3 under the null hypothesis but it is less than or greater than 0.3 under the alternative hypothesis. The null hypothesis is a simple hypothesis, whereas the alternative hypothesis is composite; actually it is what is known as a **two-sided hypothesis.** Very often such a two-sided alternative hypothesis reflects the fact that we do not have a strong a priori or theoretical expectation about the direction in which the alternative hypothesis should move from the null hypothesis.

Is the observed $\hat{\beta}_2$ compatible with H_0? To answer this question, let us refer to the confidence interval (5.3.9). We know that in the long run intervals like (0.4268, 0.5914) will contain the true β_2 with 95 percent probability. Consequently, in the long run (i.e., repeated sampling) such intervals provide a range or limits within which the true β_2 may lie with a confidence coefficient of, say, 95%. Thus, the confidence interval provides a set of plausible null hypotheses. Therefore, if β_2 under H_0 falls within the $100(1 - \alpha)\%$ confidence interval, we do not reject the null hypothesis; if it lies outside the interval, we may reject it.[7] This range is illustrated schematically in Figure 5.2.

> **Decision Rule:** Construct a $100(1 - \alpha)\%$ confidence interval for β_2. If the β_2 under H_0 falls within this confidence interval, do not reject H_0, but if it falls outside this interval, reject H_0.

Following this rule, for our hypothetical example, $H_0: \beta_2 = 0.3$ clearly lies outside the 95% confidence interval given in (5.3.9). Therefore, we can reject

[7] Always bear in mind that there is a 100α percent chance that the confidence interval does not contain β_2 under H_0 even though the hypothesis is correct. In short, there is a 100α percent chance of committing a **Type I error.** Thus, if $\alpha = 0.05$, there is a 5 percent chance that we could reject the null hypothesis even though it is true.

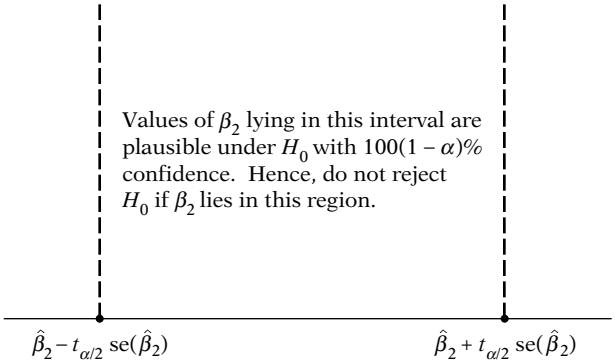


FIGURE 5.2 A $100(1-\alpha)\%$ confidence interval for β_2.

the hypothesis that the true MPC is 0.3, with 95% confidence. If the null hypothesis were true, the probability of our obtaining a value of MPC of as much as 0.5091 by sheer chance or fluke is at the most about 5 percent, a small probability.

In statistics, when we reject the null hypothesis, we say that our finding is **statistically significant.** On the other hand, when we do not reject the null hypothesis, we say that our finding is **not statistically significant.**

Some authors use a phrase such as "highly statistically significant." By this they usually mean that when they reject the null hypothesis, the probability of committing a Type I error (i.e., α) is a small number, usually 1 percent. But as our discussion of the **p value** in Section 5.8 will show, it is better to leave it to the researcher to decide whether a statistical finding is "significant," "moderately significant," or "highly significant."

One-Sided or One-Tail Test

Sometimes we have a strong a priori or theoretical expectation (or expectations based on some previous empirical work) that the alternative hypothesis is one-sided or unidirectional rather than two-sided, as just discussed. Thus, for our consumption–income example, one could postulate that

$$H_0: \beta_2 \leq 0.3 \qquad \text{and} \qquad H_1: \beta_2 > 0.3$$

Perhaps economic theory or prior empirical work suggests that the marginal propensity to consume is greater than 0.3. Although the procedure to test this hypothesis can be easily derived from (5.3.5), the actual mechanics are better explained in terms of the test-of-significance approach discussed next.[8]

[8]If you want to use the confidence interval approach, construct a $(100-\alpha)\%$ *one-sided* or *one-tail* confidence interval for β_2. Why?

5.7 HYPOTHESIS TESTING: THE TEST-OF-SIGNIFICANCE APPROACH

Testing the Significance of Regression Coefficients: The *t* Test

An *alternative but complementary approach* to the confidence-interval method of testing statistical hypotheses is the **test-of-significance approach** developed along independent lines by R. A. Fisher and jointly by Neyman and Pearson.[9] **Broadly speaking, a test of significance is a procedure by which sample results are used to verify the truth or falsity of a null hypothesis.** The key idea behind tests of significance is that of a **test statistic** (estimator) and the sampling distribution of such a statistic under the null hypothesis. The decision to accept or reject H_0 is made on the basis of the value of the test statistic obtained from the data at hand.

As an illustration, recall that under the normality assumption the variable

$$t = \frac{\hat{\beta}_2 - \beta_2}{\text{se}\,(\hat{\beta}_2)} = \frac{(\hat{\beta}_2 - \beta_2)\sqrt{\sum x_i^2}}{\hat{\sigma}} \tag{5.3.2}$$

follows the t distribution with $n - 2$ df. If the value of true β_2 is specified under the null hypothesis, the t value of (5.3.2) can readily be computed from the available sample, and therefore it can serve as a test statistic. And since this test statistic follows the t distribution, confidence-interval statements such as the following can be made:

$$\Pr\left[-t_{\alpha/2} \leq \frac{\hat{\beta}_2 - \beta_2^*}{\text{se}\,(\hat{\beta}_2)} \leq t_{\alpha/2}\right] = 1 - \alpha \tag{5.7.1}$$

where β_2^* is the value of β_2 under H_0 and where $-t_{\alpha/2}$ and $t_{\alpha/2}$ are the values of t (the **critical** t values) obtained from the t table for $(\alpha/2)$ level of significance and $n - 2$ df [cf. (5.3.4)]. The t table is given in **Appendix D**.

Rearranging (5.7.1), we obtain

$$\Pr\left[\beta_2^* - t_{\alpha/2}\,\text{se}\,(\hat{\beta}_2) \leq \hat{\beta}_2 \leq \beta_2^* + t_{\alpha/2}\,\text{se}\,(\hat{\beta}_2)\right] = 1 - \alpha \tag{5.7.2}$$

which gives the interval in which $\hat{\beta}_2$ will fall with $1 - \alpha$ probability, given $\beta_2 = \beta_2^*$. In the language of hypothesis testing, the $100(1 - \alpha)\%$ confidence interval established in (5.7.2) is known as the **region of acceptance** (of

[9] Details may be found in E. L. Lehman, *Testing Statistical Hypotheses,* John Wiley & Sons, New York, 1959.

the null hypothesis) and the *region(s)* outside the confidence interval is (are) called the **region(s) of rejection** (of H_0) or the **critical region(s).** As noted previously, the confidence limits, the endpoints of the confidence interval, are also called **critical values.**

The intimate connection between the confidence-interval and test-of-significance approaches to hypothesis testing can now be seen by comparing (5.3.5) with (5.7.2). In the confidence-interval procedure we try to establish a range or an interval that has a certain probability of including the true but unknown β_2, whereas in the test-of-significance approach we hypothesize some value for β_2 and try to see whether the computed $\hat{\beta}_2$ lies within reasonable (confidence) limits around the hypothesized value.

Once again let us revert to our consumption–income example. We know that $\hat{\beta}_2 = 0.5091$, se $(\hat{\beta}_2) = 0.0357$, and df $= 8$. If we assume $\alpha = 5$ percent, $t_{\alpha/2} = 2.306$. If we let H_0: $\beta_2 = \beta_2^* = 0.3$ and H_1: $\beta_2 \neq 0.3$, (5.7.2) becomes

$$\Pr\,(0.2177 \leq \hat{\beta}_2 \leq 0.3823) = 0.95 \qquad \textbf{(5.7.3)}^{10}$$

as shown diagrammatically in Figure 5.3. Since the observed $\hat{\beta}_2$ lies in the critical region, we reject the null hypothesis that true $\beta_2 = 0.3$.

In practice, there is no need to estimate (5.7.2) explicitly. One can compute the t value in the middle of the double inequality given by (5.7.1) and see whether it lies between the critical t values or outside them. For our example,

$$t = \frac{0.5091 - 0.3}{0.0357} = 5.86 \qquad \textbf{(5.7.4)}$$

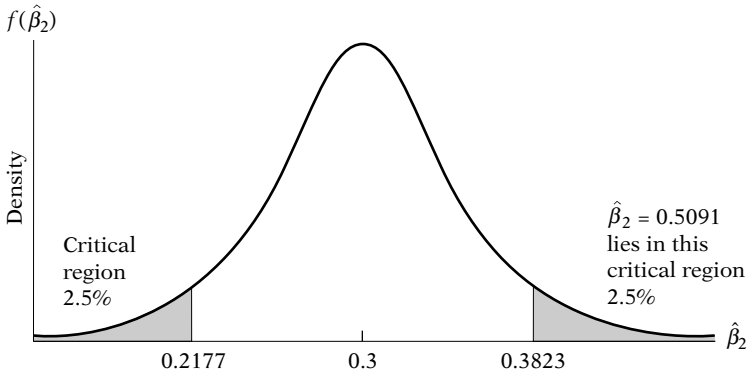


FIGURE 5.3 The 95% confidence interval for $\hat{\beta}_2$ under the hypothesis that $\beta_2 = 0.3$.

[10] In Sec. 5.2, point 4, it was stated that we *cannot* say that the probability is 95 percent that the fixed interval (0.4268, 0.5914) includes the true β_2. But we can make the probabilistic statement given in (5.7.3) because $\hat{\beta}_2$, being an estimator, is a random variable.

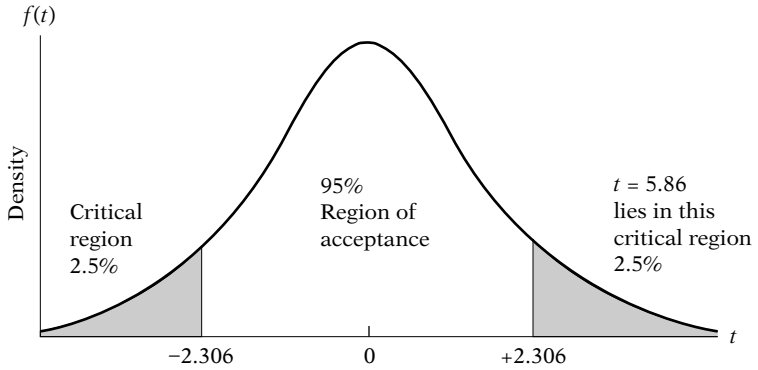


FIGURE 5.4 The 95% confidence interval for t(8 df).

which clearly lies in the critical region of Figure 5.4. The conclusion remains the same; namely, we reject H_0.

Notice that if the estimated β_2 ($= \hat{\beta}_2$) is equal to the hypothesized β_2, the t value in (5.7.4) will be zero. However, as the estimated β_2 value departs from the hypothesized β_2 value, $|t|$ (that is, the absolute t value; *note:* t can be positive as well as negative) will be increasingly large. *Therefore, a "large" $|t|$ value will be evidence against the null hypothesis.* Of course, we can always use the t table to determine whether a particular t value is large or small; the answer, as we know, depends on the degrees of freedom as well as on the probability of Type I error that we are willing to accept. If you take a look at the t table given in **Appendix D**, you will observe that for any given value of df the probability of obtaining an increasingly large $|t|$ value becomes progressively smaller. Thus, for 20 df the probability of obtaining a $|t|$ value of 1.725 or greater is 0.10 or 10 percent, but for the same df the probability of obtaining a $|t|$ value of 3.552 or greater is only 0.002 or 0.2 percent.

Since we use the t distribution, the preceding testing procedure is called appropriately the ***t* test. In the language of significance tests, a statistic is said to be statistically significant if the value of the test statistic lies in the critical region. In this case the null hypothesis is rejected. By the same token, a test is said to be statistically insignificant if the value of the test statistic lies in the acceptance region.** In this situation, the null hypothesis is not rejected. In our example, the t test is significant and hence we reject the null hypothesis.

Before concluding our discussion of hypothesis testing, note that the testing procedure just outlined is known as a **two-sided, or two-tail,** test-of-significance procedure in that we consider the two extreme tails of the relevant probability distribution, the rejection regions, and reject the null hypothesis if it lies in either tail. But this happens because our H_1 was a

two-sided composite hypothesis; $\beta_2 \neq 0.3$ means β_2 is either greater than or less than 0.3. But suppose prior experience suggests to us that the MPC is expected to be greater than 0.3. In this case we have: $H_0: \beta_2 \leq 0.3$ and $H_1: \beta_2 > 0.3$. Although H_1 is still a composite hypothesis, it is now one-sided. To test this hypothesis, we use the **one-tail test** (the right tail), as shown in Figure 5.5. (See also the discussion in Section 5.6.)

The test procedure is the same as before except that the upper confidence limit or critical value now corresponds to $t_\alpha = t_{.05}$, that is, the 5 percent level. As Figure 5.5 shows, we need not consider the lower tail of the t distribution in this case. Whether one uses a two- or one-tail test of significance will depend upon how the alternative hypothesis is formulated, which, in turn, may depend upon some a priori considerations or prior empirical experience. (But more on this in Section 5.8.)

We can summarize the t test of significance approach to hypothesis testing as shown in Table 5.1.

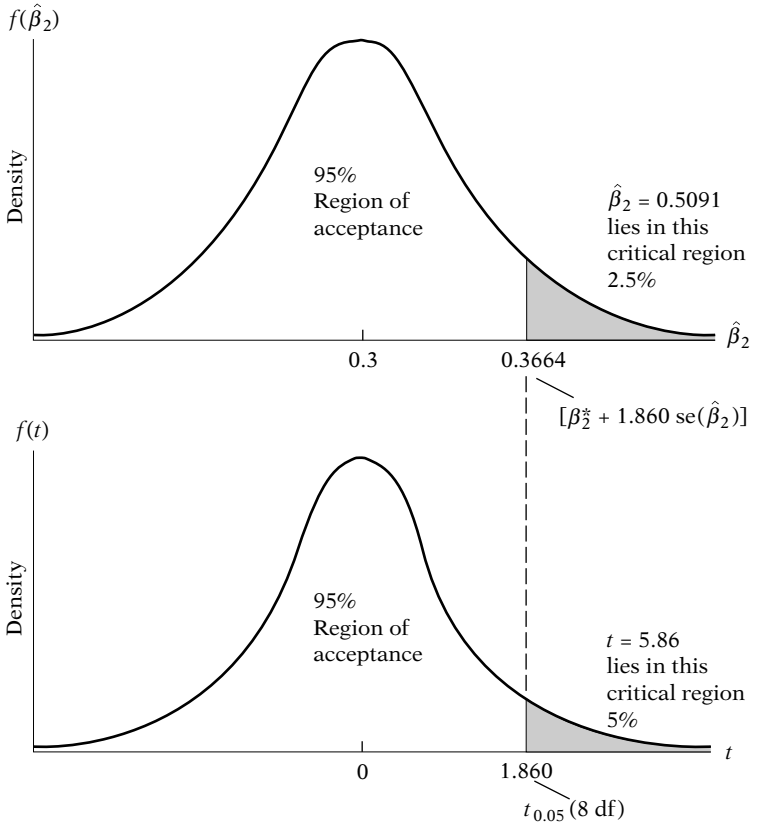


FIGURE 5.5 One-tail test of significance.

TABLE 5.1 THE *t* TEST OF SIGNIFICANCE: DECISION RULES

Type of hypothesis	H_0: the null hypothesis	H_1: the alternative hypothesis	Decision rule: reject H_0 if
Two-tail	$\beta_2 = \beta_2^*$	$\beta_2 \neq \beta_2^*$	$\lvert t \rvert > t_{\alpha/2,\text{df}}$
Right-tail	$\beta_2 \leq \beta_2^*$	$\beta_2 > \beta_2^*$	$t > t_{\alpha,\text{df}}$
Left-tail	$\beta_2 \geq \beta_2^*$	$\beta_2 < \beta_2^*$	$t < -t_{\alpha,\text{df}}$

Notes: β_2^* is the hypothesized numerical value of β_2.
$|t|$ means the absolute value of t.
t_α or $t_{\alpha/2}$ means the critical t value at the α or $\alpha/2$ level of significance.
df: degrees of freedom, $(n-2)$ for the two-variable model, $(n-3)$ for the three-variable model, and so on.
The same procedure holds to test hypotheses about β_1.

Testing the Significance of σ^2: The χ^2 Test

As another illustration of the test-of-significance methodology, consider the following variable:

$$\chi^2 = (n-2)\frac{\hat{\sigma}^2}{\sigma^2} \tag{5.4.1}$$

which, as noted previously, follows the χ^2 distribution with $n-2$ df. For the hypothetical example, $\hat{\sigma}^2 = 42.1591$ and df $= 8$. If we postulate that $H_0: \sigma^2 = 85$ vs. $H_1: \sigma^2 \neq 85$, Eq. (5.4.1) provides the test statistic for H_0. Substituting the appropriate values in (5.4.1), it can be found that under H_0, $\chi^2 = 3.97$. If we assume $\alpha = 5\%$, the critical χ^2 values are 2.1797 and 17.5346. Since the computed χ^2 lies between these limits, the data support the null hypothesis and we do not reject it. (See Figure 5.1.) This test procedure is called the **chi-square test of significance.** The χ^2 test of significance approach to hypothesis testing is summarized in Table 5.2.

TABLE 5.2 A SUMMARY OF THE χ^2 TEST

H_0: the null hypothesis	H_1: the alternative hypothesis	Critical region: reject H_0 if
$\sigma^2 = \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$\frac{\text{df}(\hat{\sigma}^2)}{\sigma_0^2} > \chi^2_{\alpha,\text{df}}$
$\sigma^2 = \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$\frac{\text{df}(\hat{\sigma}^2)}{\sigma_0^2} < \chi^2_{(1-\alpha),\text{df}}$
$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$\frac{\text{df}(\hat{\sigma}^2)}{\sigma_0^2} > \chi^2_{\alpha/2,\text{df}}$ or $< \chi^2_{(1-\alpha/2),\text{df}}$

Note: σ_0^2 is the value of σ^2 under the null hypothesis. The first subscript on χ^2 in the last column is the level of significance, and the second subscript is the degrees of freedom. These are critical chi-square values. Note that df is $(n-2)$ for the two-variable regression model, $(n-3)$ for the three-variable regression model, and so on.

5.8 HYPOTHESIS TESTING: SOME PRACTICAL ASPECTS

The Meaning of "Accepting" or "Rejecting" a Hypothesis

If on the basis of a test of significance, say, the t test, we decide to "accept" the null hypothesis, all we are saying is that on the basis of the sample evidence we have no reason to reject it; we are not saying that the null hypothesis is true beyond any doubt. Why? To answer this, let us revert to our consumption–income example and assume that H_0: β_2 (MPC) $= 0.50$. Now the estimated value of the MPC is $\hat{\beta}_2 = 0.5091$ with a se $(\hat{\beta}_2) = 0.0357$. Then on the basis of the t test we find that $t = (0.5091 - 0.50)/0.0357 = 0.25$, which is insignificant, say, at $\alpha = 5\%$. Therefore, we say "accept" H_0. But now let us assume H_0: $\beta_2 = 0.48$. Applying the t test, we obtain $t = (0.5091 - 0.48)/0.0357 = 0.82$, which too is statistically insignificant. So now we say "accept" this H_0. Which of these two null hypotheses is the "truth"? We do not know. Therefore, in "accepting" a null hypothesis we should always be aware that another null hypothesis may be equally compatible with the data. It is therefore preferable to say that we *may* accept the null hypothesis rather than we (do) accept it. Better still,

> . . . just as a court pronounces a verdict as "not guilty" rather than "innocent," so the conclusion of a statistical test is "do not reject" rather than "accept."[11]

The "Zero" Null Hypothesis and the "2-*t*" Rule of Thumb

A null hypothesis that is commonly tested in empirical work is H_0: $\beta_2 = 0$, that is, the slope coefficient is zero. This "zero" null hypothesis is a kind of straw man, the objective being to find out whether Y is related at all to X, the explanatory variable. If there is no relationship between Y and X to begin with, then testing a hypothesis such as $\beta_2 = 0.3$ or any other value is meaningless.

This null hypothesis can be easily tested by the confidence interval or the t-test approach discussed in the preceding sections. But very often such formal testing can be shortcut by adopting the "2-t" rule of significance, which may be stated as

> **"2-*t*" Rule of Thumb.** If the number of degrees of freedom is 20 or more and if α, the level of significance, is set at 0.05, then the null hypothesis $\beta_2 = 0$ can be rejected if the t value $[= \hat{\beta}_2/\text{se}\,(\hat{\beta}_2)]$ computed from (5.3.2) exceeds 2 in absolute value.

The rationale for this rule is not too difficult to grasp. From (5.7.1) we know that we will reject H_0: $\beta_2 = 0$ if

$$t = \hat{\beta}_2/\text{se}\,(\hat{\beta}_2) > t_{\alpha/2} \qquad \text{when } \hat{\beta}_2 > 0$$

[11] Jan Kmenta, *Elements of Econometrics,* Macmillan, New York, 1971, p. 114.

or

$$t = \hat{\beta}_2/\text{se}\,(\hat{\beta}_2) < -t_{\alpha/2} \qquad \text{when } \hat{\beta}_2 < 0$$

or when

$$|t| = \left|\frac{\hat{\beta}_2}{\text{se}\,(\hat{\beta}_2)}\right| > t_{\alpha/2} \tag{5.8.1}$$

for the appropriate degrees of freedom.

Now if we examine the t table given in **Appendix D**, we see that for df of about 20 or more a computed t value in excess of 2 (in absolute terms), say, 2.1, is statistically significant at the 5 percent level, implying rejection of the null hypothesis. Therefore, if we find that for 20 or more df the computed t value is, say, 2.5 or 3, we do not even have to refer to the t table to assess the significance of the estimated slope coefficient. Of course, one can always refer to the t table to obtain the precise level of significance, and one should always do so when the df are fewer than, say, 20.

In passing, note that if we are testing the one-sided hypothesis $\beta_2 = 0$ versus $\beta_2 > 0$ or $\beta_2 < 0$, then we should reject the null hypothesis if

$$|t| = \left|\frac{\hat{\beta}_2}{\text{se}\,(\hat{\beta}_2)}\right| > t_{\alpha} \tag{5.8.2}$$

If we fix α at 0.05, then from the t table we observe that for 20 or more df a t value in excess of 1.73 is statistically significant at the 5 percent level of significance (one-tail). Hence, whenever a t value exceeds, say, 1.8 (in absolute terms) and the df are 20 or more, one need not consult the t table for the statistical significance of the observed coefficient. Of course, if we choose α at 0.01 or any other level, we will have to decide on the appropriate t value as the benchmark value. But by now the reader should be able to do that.

Forming the Null and Alternative Hypotheses[12]

Given the null and the alternative hypotheses, testing them for statistical significance should no longer be a mystery. But how does one formulate these hypotheses? There are no hard-and-fast rules. Very often the phenomenon under study will suggest the nature of the null and alternative hypotheses. For example, consider the capital market line (CML) of portfolio theory, which postulates that $E_i = \beta_1 + \beta_2\sigma_i$, where E = expected return on portfolio and σ = the standard deviation of return, a measure of risk. Since return and risk are expected to be positively related—the higher the risk, the

[12]For an interesting discussion about formulating hypotheses, see J. Bradford De Long and Kevin Lang, "Are All Economic Hypotheses False?" *Journal of Political Economy,* vol. 100, no. 6, 1992, pp. 1257–1272.

higher the return—the natural alternative hypothesis to the null hypothesis that $\beta_2 = 0$ would be $\beta_2 > 0$. That is, one would not choose to consider values of β_2 less than zero.

But consider the case of the demand for money. As we shall show later, one of the important determinants of the demand for money is income. Prior studies of the money demand functions have shown that the income elasticity of demand for money (the percent change in the demand for money for a 1 percent change in income) has typically ranged between 0.7 and 1.3. Therefore, in a new study of demand for money, if one postulates that the income-elasticity coefficient β_2 is 1, the alternative hypothesis could be that $\beta_2 \neq 1$, a two-sided alternative hypothesis.

Thus, theoretical expectations or prior empirical work or both can be relied upon to formulate hypotheses. But no matter how the hypotheses are formed, *it is extremely important that the researcher establish these hypotheses before carrying out the empirical investigation*. Otherwise, he or she will be guilty of circular reasoning or self-fulfilling prophesies. That is, if one were to formulate hypotheses after examining the empirical results, there may be the temptation to form hypotheses that justify one's results. Such a practice should be avoided at all costs, at least for the sake of scientific objectivity. Keep in mind the Stigler quotation given at the beginning of this chapter!

Choosing α, the Level of Significance

It should be clear from the discussion so far that whether we reject or do not reject the null hypothesis depends critically on α, the level of significance or the *probability of committing a* ***Type I*** *error*—the probability of rejecting the true hypothesis. In **Appendix A** we discuss fully the nature of a Type I error, its relationship to a ***Type II*** *error* (the probability of accepting the false hypothesis) and why classical statistics generally concentrates on a Type I error. But even then, why is α commonly fixed at the 1, 5, or at the most 10 percent levels? As a matter of fact, there is nothing sacrosanct about these values; any other values will do just as well.

In an introductory book like this it is not possible to discuss in depth why one chooses the 1, 5, or 10 percent levels of significance, for that will take us into the field of statistical decision making, a discipline unto itself. A brief summary, however, can be offered. As we discuss in **Appendix A,** for a given sample size, if we try to reduce a *Type I error*, a *Type II error* increases, and vice versa. That is, given the sample size, if we try to reduce the probability of rejecting the true hypothesis, we at the same time increase the probability of accepting the false hypothesis. So there is a tradeoff involved between these two types of errors, given the sample size. Now the only way we can decide about the tradeoff is to find out the relative costs of the two types of errors. Then,

> If the error of rejecting the null hypothesis which is in fact true (Error Type I) is costly relative to the error of not rejecting the null hypothesis which is in fact

false (Error Type II), it will be rational to set the probability of the first kind of error low. If, on the other hand, the cost of making Error Type I is low relative to the cost of making Error Type II, it will pay to make the probability of the first kind of error high (thus making the probability of the second type of error low).[13]

Of course, the rub is that we rarely know the costs of making the two types of errors. Thus, applied econometricians generally follow the practice of setting the value of α at a 1 or a 5 or at most a 10 percent level and choose a test statistic that would make the probability of committing a Type II error as small as possible. Since one minus the probability of committing a Type II error is known as the **power of the test,** this procedure amounts to maximizing the power of the test. (See **Appendix A** for a discussion of the power of a test.)

But all this problem with choosing the appropriate value of α can be avoided if we use what is known as the ***p value*** of the test statistic, which is discussed next.

The Exact Level of Significance: The *p* Value

As just noted, the Achilles heel of the classical approach to hypothesis testing is its arbitrariness in selecting α. Once a test statistic (e.g., the t statistic) is obtained in a given example, why not simply go to the appropriate statistical table and find out the actual probability of obtaining a value of the test statistic as much as or greater than that obtained in the example? This probability is called the ***p* value** (i.e., **probability value**), also known as the **observed or exact level of significance** or the **exact probability of committing a Type I error.** More technically, the p value is defined as **the lowest significance level at which a null hypothesis can be rejected.**

To illustrate, let us return to our consumption–income example. Given the null hypothesis that the true MPC is 0.3, we obtained a t value of 5.86 in (5.7.4). What is the p value of obtaining a t value of as much as or greater than 5.86? Looking up the t table given in **Appendix D**, we observe that for 8 df the probability of obtaining such a t value must be much smaller than 0.001 (one-tail) or 0.002 (two-tail). By using the computer, it can be shown that the probability of obtaining a t value of 5.86 or greater (for 8 df) is about 0.000189.[14] This is the p value of the observed t statistic. This observed, or exact, level of significance of the t statistic is much smaller than the conventionally, and arbitrarily, fixed level of significance, such as 1, 5, or 10 percent. As a matter of fact, if we were to use the p value just computed,

[13]Jan Kmenta, *Elements of Econometrics*, Macmillan, New York, 1971, pp. 126–127.

[14]One can obtain the p value using electronic statistical tables to several decimal places. Unfortunately, the conventional statistical tables, for lack of space, cannot be that refined. Most statistical packages now routinely print out the p values.

and reject the null hypothesis that the true MPC is 0.3, the probability of our committing a Type I error is only about 0.02 percent, that is, only about 2 in 10,000!

As we noted earlier, if the data do not support the null hypothesis, $|t|$ obtained under the null hypothesis will be "large" and therefore the p value of obtaining such a $|t|$ value will be "small." In other words, for a given sample size, as $|t|$ increases, the p value decreases, and one can therefore reject the null hypothesis with increasing confidence.

What is the relationship of the p value to the level of significance α? If we make the habit of fixing α equal to the p value of a test statistic (e.g., the t statistic), then there is no conflict between the two values. To put it differently, **it is better to give up fixing α arbitrarily at some level and simply choose the p value of the test statistic.** It is preferable to leave it to the reader to decide whether to reject the null hypothesis at the given p value. If in an application the p value of a test statistic happens to be, say, 0.145, or 14.5 percent, and if the reader wants to reject the null hypothesis at this (exact) level of significance, so be it. Nothing is wrong with taking a chance of being wrong 14.5 percent of the time if you reject the true null hypothesis. Similarly, as in our consumption–income example, there is nothing wrong if the researcher wants to choose a p value of about 0.02 percent and not take a chance of being wrong more than 2 out of 10,000 times. After all, some investigators may be risk-lovers and some risk-averters!

In the rest of this text, we will generally quote the p value of a given test statistic. Some readers may want to fix α at some level and reject the null hypothesis if the p value is less than α. That is their choice.

Statistical Significance versus Practical Significance

Let us revert to our consumption–income example and now hypothesize that the true MPC is 0.61 (H_0: $\beta_2 = 0.61$). On the basis of our sample result of $\hat{\beta}_2 = 0.5091$, we obtained the interval (0.4268, 0.5914) with 95 percent confidence. Since this interval does not include 0.61, we can, with 95 percent confidence, say that our estimate is statistically significant, that is, significantly different from 0.61.

But what is the practical or substantive significance of our finding? That is, what difference does it make if we take the MPC to be 0.61 rather than 0.5091? Is the 0.1009 difference between the two MPCs that important practically?

The answer to this question depends on what we really do with these estimates. For example, from macroeconomics we know that the income multiplier is $1/(1 - \text{MPC})$. Thus, if MPC is 0.5091, the multiplier is 2.04, but it is 2.56 if MPC is equal to 0.61. That is, if the government were to increase its expenditure by \$1 to lift the economy out of a recession, income will eventually increase by \$2.04 if the MPC is 0.5091 but by \$2.56 if the MPC is 0.61. And that difference could very well be crucial to resuscitating the economy.

The point of all this discussion is that one should not confuse statistical significance with practical, or economic, significance. As Goldberger notes:

> When a null, say, $\beta_j = 1$, is specified, the likely intent is that β_j is *close* to 1, so close that for all practical purposes it may be treated *as if it were* 1. But whether 1.1 is "practically the same as" 1.0 is a matter of economics, not of statistics. One cannot resolve the matter by relying on a hypothesis test, because the test statistic $[t =](b_j - 1)/\hat{\sigma}_{bj}$ measures the estimated coefficient in standard error units, which are not meaningful units in which to measure the economic parameter $\beta_j - 1$. It may be a good idea to reserve the term "significance" for the statistical concept, adopting "substantial" for the economic concept.[15]

The point made by Goldberger is important. As sample size becomes very large, issues of statistical significance become much less important but issues of economic significance become critical. Indeed, since with very large samples almost any null hypothesis will be rejected, there may be studies in which the magnitude of the point estimates may be the only issue.

The Choice between Confidence-Interval and Test-of-Significance Approaches to Hypothesis Testing

In most applied economic analyses, the null hypothesis is set up as a straw man and the objective of the empirical work is to knock it down, that is, reject the null hypothesis. Thus, in our consumption–income example, the null hypothesis that the MPC $\beta_2 = 0$ is patently absurd, but we often use it to dramatize the empirical results. Apparently editors of reputed journals do not find it exciting to publish an empirical piece that does not reject the null hypothesis. Somehow the finding that the MPC is statistically different from zero is more newsworthy than the finding that it is equal to, say, 0.7!

Thus, J. Bradford De Long and Kevin Lang argue that it is better for economists

> . . . to concentrate on the magnitudes of coefficients and to report confidence levels and not significance tests. If all or almost all null hypotheses are false, there is little point in concentrating on whether or not an estimate is indistinguishable from its predicted value under the null. Instead, we wish to cast light on what models are good approximations, which requires that we know ranges of parameter values that are excluded by empirical estimates.[16]

In short, these authors prefer the confidence-interval approach to the test-of-significance approach. The reader may want to keep this advice in mind.[17]

---

[15]Arthur S. Goldberger, *A Course in Econometrics*, Harvard University Press, Cambridge, Massachusetts, 1991, p. 240. Note b_j is the OLS estimator of β_j and $\hat{\sigma}_{bj}$ is its standard error. For a corroborating view, see D. N. McCloskey, "The Loss Function Has Been Mislaid: The Rhetoric of Significance Tests," *American Economic Review*, vol. 75, 1985, pp. 201–205. See also D. N. McCloskey and S. T. Ziliak, "The Standard Error of Regression," *Journal of Economic Literature*, vol. 37, 1996, pp. 97–114.

[16]See their article cited in footnote 12, p. 1271.

[17]For a somewhat different perspective, see Carter Hill, William Griffiths, and George Judge, *Undergraduate Econometrics,* Wiley & Sons, New York, 2001, p. 108.

5.9 REGRESSION ANALYSIS AND ANALYSIS OF VARIANCE

In this section we study regression analysis from the point of view of the analysis of variance and introduce the reader to an illuminating and complementary way of looking at the statistical inference problem.

In Chapter 3, Section 3.5, we developed the following identity:

$$\sum y_i^2 = \sum \hat{y}_i^2 + \sum \hat{u}_i^2 = \hat{\beta}_2^2 \sum x_i^2 + \sum \hat{u}_i^2 \qquad \textbf{(3.5.2)}$$

that is, TSS = ESS + RSS, which decomposed the total sum of squares (TSS) into two components: explained sum of squares (ESS) and residual sum of squares (RSS). A study of these components of TSS is known as the **analysis of variance** (ANOVA) from the regression viewpoint.

Associated with any sum of squares is its df, the number of independent observations on which it is based. TSS has $n - 1$ df because we lose 1 df in computing the sample mean $\bar{Y}$. RSS has $n - 2$ df. (Why?) (*Note:* This is true only for the two-variable regression model with the intercept β_1 present.) ESS has 1 df (again true of the two-variable case only), which follows from the fact that $\text{ESS} = \hat{\beta}_2^2 \sum x_i^2$ is a function of $\hat{\beta}_2$ only, since $\sum x_i^2$ is known.

Let us arrange the various sums of squares and their associated df in Table 5.3, which is the standard form of the AOV table, sometimes called the **ANOVA table.** Given the entries of Table 5.3, we now consider the following variable:

$$\begin{aligned} F &= \frac{\text{MSS of ESS}}{\text{MSS of RSS}} \\ &= \frac{\hat{\beta}_2^2 \sum x_i^2}{\sum \hat{u}_i^2 / (n-2)} \\ &= \frac{\hat{\beta}_2^2 \sum x_i^2}{\hat{\sigma}^2} \end{aligned} \qquad \textbf{(5.9.1)}$$

If we assume that the disturbances u_i are normally distributed, which we do under the CNLRM, and if the null hypothesis (H_0) is that $\beta_2 = 0$, then it can be shown that the F variable of (5.9.1) follows the F distribution with

TABLE 5.3 ANOVA TABLE FOR THE TWO-VARIABLE REGRESSION MODEL

Source of variation	SS*	df	MSS†
Due to regression (ESS)	$\sum \hat{y}_i^2 = \hat{\beta}_2^2 \sum x_i^2$	1	$\hat{\beta}_2^2 \sum x_i^2$
Due to residuals (RSS)	$\sum \hat{u}_i^2$	$n - 2$	$\dfrac{\sum \hat{u}_i^2}{n-2} = \hat{\sigma}^2$
TSS	$\sum y_i^2$	$n - 1$	

*SS means sum of squares.
†Mean sum of squares, which is obtained by dividing SS by their df.

1 df in the numerator and $(n - 2)$ df in the denominator. (See Appendix 5A, Section 5A.3, for the proof. The general properties of the F distribution are discussed in **Appendix A**.)

What use can be made of the preceding F ratio? It can be shown[18] that

$$E\left(\hat{\beta}_2^2 \sum x_i^2\right) = \sigma^2 + \beta_2^2 \sum x_i^2 \tag{5.9.2}$$

and

$$E\frac{\sum \hat{u}_i^2}{n-2} = E(\hat{\sigma}^2) = \sigma^2 \tag{5.9.3}$$

(Note that β_2 and σ^2 appearing on the right sides of these equations are the true parameters.) Therefore, if β_2 is in fact zero, Eqs. (5.9.2) and (5.9.3) both provide us with identical estimates of true σ^2. In this situation, the explanatory variable X has no linear influence on Y whatsoever and the entire variation in Y is explained by the random disturbances u_i. If, on the other hand, β_2 is not zero, (5.9.2) and (5.9.3) will be different and part of the variation in Y will be ascribable to X. Therefore, the F ratio of (5.9.1) provides a test of the null hypothesis H_0: $\beta_2 = 0$. Since all the quantities entering into this equation can be obtained from the available sample, this F ratio provides a test statistic to test the null hypothesis that true β_2 is zero. All that needs to be done is to compute the F ratio and compare it with the critical F value obtained from the F tables at the chosen level of significance, or obtain the ***p* value** of the computed F statistic.

To illustrate, let us continue with our consumption–income example. The ANOVA table for this example is as shown in Table 5.4. The computed F value is seen to be 202.87. The p value of this F statistic corresponding to 1 and 8 df cannot be obtained from the F table given in **Appendix D**, but by using electronic statistical tables it can be shown that the p value is 0.0000001, an extremely small probability indeed. If you decide to choose the level-of-significance approach to hypothesis testing and fix α at 0.01, or a 1 percent level, you can see that the computed F of 202.87 is obviously significant at this level. Therefore, if we reject the null hypothesis that $\beta_2 = 0$, the probability of committing a Type I error is very small. For all practical

TABLE 5.4 ANOVA TABLE FOR THE CONSUMPTION–INCOME EXAMPLE

Source of variation	SS	df	MSS	
Due to regression (ESS)	8552.73	1	8552.73	$F = \frac{8552.73}{42.159}$
Due to residuals (RSS)	337.27	8	42.159	$= 202.87$
TSS	8890.00	9		

[18]For proof, see K. A. Brownlee, *Statistical Theory and Methodology in Science and Engineering*, John Wiley & Sons, New York, 1960, pp. 278–280.

purposes, our sample could not have come from a population with zero β_2 value and we can conclude with great confidence that X, income, does affect Y, consumption expenditure.

Refer to Theorem 5.7 of Appendix 5A.1, which states that the square of the t value with k df is an F value with 1 df in the numerator and k df in the denominator. For our consumption–income example, if we assume H_0: $\beta_2 = 0$, then from (5.3.2) it can be easily verified that the estimated t value is 14.26. This t value has 8 df. Under the same null hypothesis, the F value was 202.87 with 1 and 8 df. Hence $(14.24)^2 = F$ value, except for the rounding errors.

Thus, the t and the F tests provide us with two alternative but complementary ways of testing the null hypothesis that $\beta_2 = 0$. If this is the case, why not just rely on the t test and not worry about the F test and the accompanying analysis of variance? For the two-variable model there really is no need to resort to the F test. But when we consider the topic of multiple regression we will see that the F test has several interesting applications that make it a very useful and powerful method of testing statistical hypotheses.

5.10 APPLICATION OF REGRESSION ANALYSIS: THE PROBLEM OF PREDICTION

On the basis of the sample data of Table 3.2 we obtained the following sample regression:

$$\hat{Y}_i = 24.4545 + 0.5091X_i \qquad \textbf{(3.6.2)}$$

where $\hat{Y}_t$ is the estimator of true $E(Y_i)$ corresponding to given X. What use can be made of this **historical regression?** One use is to "predict" or "forecast" the future consumption expenditure Y corresponding to some given level of income X. Now there are two kinds of predictions: (1) prediction of the conditional mean value of Y corresponding to a chosen X, say, X_0, that is the point on the population regression line itself (see Figure 2.2), and (2) prediction of an individual Y value corresponding to X_0. We shall call these two predictions the **mean prediction** and **individual prediction.**

Mean Prediction[19]

To fix the ideas, assume that $X_0 = 100$ and we want to predict $E(Y \mid X_0 = 100)$. Now it can be shown that the historical regression (3.6.2) provides the point estimate of this mean prediction as follows:

$$\begin{aligned} \hat{Y}_0 &= \hat{\beta}_1 + \hat{\beta}_2 X_0 \\ &= 24.4545 + 0.5091(100) \\ &= 75.3645 \end{aligned} \qquad \textbf{(5.10.1)}$$

[19]For the proofs of the various statements made, see App. 5A, Sec. 5A.4.

where $\hat{Y}_0$ = estimator of $E(Y \mid X_0)$. It can be proved that this point predictor is a best linear unbiased estimator (BLUE).

Since $\hat{Y}_0$ is an estimator, it is likely to be different from its true value. The difference between the two values will give some idea about the prediction or forecast error. To assess this error, we need to find out the sampling distribution of $\hat{Y}_0$. It is shown in Appendix 5A, Section 5A.4, that $\hat{Y}_0$ in Eq. (5.10.1) is normally distributed with mean $(\beta_1 + \beta_2 X_0)$ and the variance is given by the following formula:

$$\operatorname{var}(\hat{Y}_0) = \sigma^2 \left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2} \right] \tag{5.10.2}$$

By replacing the unknown σ^2 by its unbiased estimator $\hat{\sigma}^2$, we see that the variable

$$t = \frac{\hat{Y}_0 - (\beta_1 + \beta_2 X_0)}{\operatorname{se}(\hat{Y}_0)} \tag{5.10.3}$$

follows the t distribution with $n - 2$ df. The t distribution can therefore be used to derive confidence intervals for the true $E(Y_0 \mid X_0)$ and test hypotheses about it in the usual manner, namely,

$$\Pr[\hat{\beta}_1 + \hat{\beta}_2 X_0 - t_{\alpha/2} \operatorname{se}(\hat{Y}_0) \leq \beta_1 + \beta_2 X_0 \leq \hat{\beta}_1 + \hat{\beta}_2 X_0 + t_{\alpha/2} \operatorname{se}(\hat{Y}_0)] = 1 - \alpha \tag{5.10.4}$$

where $\operatorname{se}(\hat{Y}_0)$ is obtained from (5.10.2).

For our data (see Table 3.3),

$$\begin{aligned} \operatorname{var}(\hat{Y}_0) &= 42.159 \left[\frac{1}{10} + \frac{(100 - 170)^2}{33{,}000} \right] \\ &= 10.4759 \end{aligned}$$

and

$$\operatorname{se}(\hat{Y}_0) = 3.2366$$

Therefore, the 95% confidence interval for true $E(Y \mid X_0) = \beta_1 + \beta_2 X_0$ is given by

$$75.3645 - 2.306(3.2366) \leq E(Y_0 \mid X = 100) \leq 75.3645 + 2.306(3.2366)$$

that is,

$$67.9010 \leq E(Y \mid X = 100) \leq 82.8381 \tag{5.10.5}$$

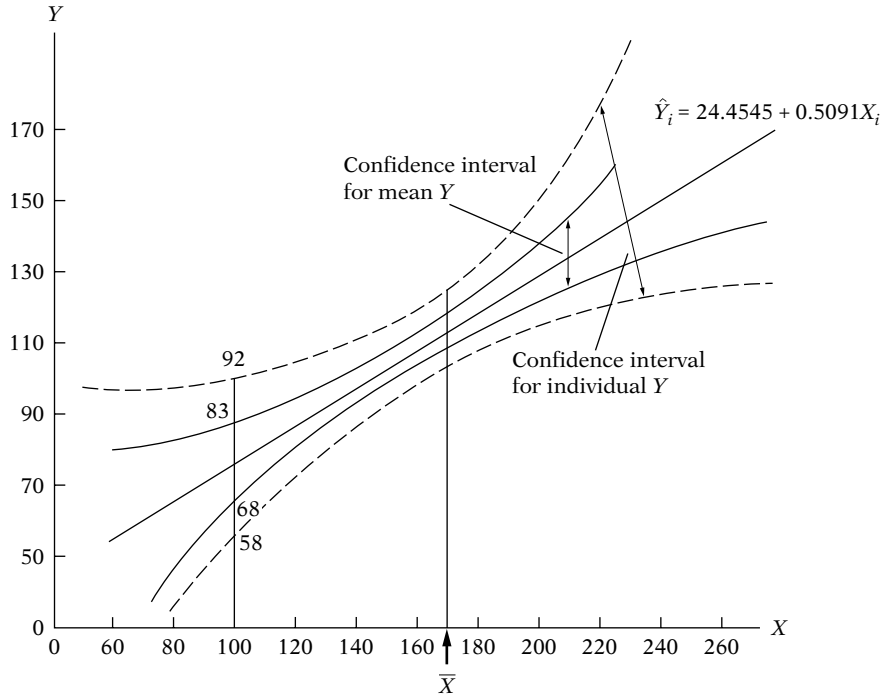


FIGURE 5.6 Confidence intervals (bands) for mean Y and individual Y values.

Thus, given $X_0 = 100$, in repeated sampling, 95 out of 100 intervals like (5.10.5) will include the true mean value; the single best estimate of the true mean value is of course the point estimate 75.3645.

If we obtain 95% confidence intervals like (5.10.5) for each of the X values given in Table 3.2, we obtain what is known as the **confidence interval,** or **confidence band,** for the population regression function, which is shown in Figure 5.6.

Individual Prediction

If our interest lies in predicting an individual Y value, Y_0, corresponding to a given X value, say, X_0, then, as shown in Appendix 5, Section 5A.3, a best linear unbiased estimator of Y_0 is also given by (5.10.1), but its variance is as follows:

$$\text{var}\,(Y_0 - \hat{Y}_0) = E[Y_0 - \hat{Y}_0]^2 = \sigma^2 \left[1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2}\right] \qquad \textbf{(5.10.6)}$$

It can be shown further that Y_0 also follows the normal distribution with mean and variance given by (5.10.1) and (5.10.6), respectively. Substituting $\hat{\sigma}^2$

for the unknown σ^2, it follows that

$$t = \frac{Y_0 - \hat{Y}_0}{\text{se}\,(Y_0 - \hat{Y}_0)}$$

also follows the t distribution. Therefore, the t distribution can be used to draw inferences about the true Y_0. Continuing with our consumption–income example, we see that the point prediction of Y_0 is 75.3645, the same as that of $\hat{Y}_0$, and its variance is 52.6349 (the reader should verify this calculation). Therefore, the 95% confidence interval for Y_0 corresponding to $X_0 = 100$ is seen to be

$$(58.6345 \leq Y_0 \mid X_0 = 100 \leq 92.0945) \qquad \textbf{(5.10.7)}$$

Comparing this interval with (5.10.5), we see that the confidence interval for individual Y_0 is wider than that for the mean value of Y_0. (Why?) Computing confidence intervals like (5.10.7) conditional upon the X values given in Table 3.2, we obtain the 95% confidence band for the individual Y values corresponding to these X values. This confidence band along with the confidence band for $\hat{Y}_0$ associated with the same X's is shown in Figure 5.6.

Notice an important feature of the confidence bands shown in Figure 5.6. The width of these bands is smallest when $X_0 = \bar{X}$. (Why?) However, the width widens sharply as X_0 moves away from $\bar{X}$. (Why?) This change would suggest that the predictive ability of the *historical* sample regression line falls markedly as X_0 departs progressively from $\bar{X}$. **Therefore, one should exercise great caution in "extrapolating" the historical regression line to predict $E(Y \mid X_0)$ or Y_0 associated with a given X_0 that is far removed from the sample mean $\bar{X}$.**

5.11 REPORTING THE RESULTS OF REGRESSION ANALYSIS

There are various ways of reporting the results of regression analysis, but in this text we shall use the following format, employing the consumption–income example of Chapter 3 as an illustration:

$$\begin{aligned} \hat{Y}_i &= 24.4545 \quad + \quad 0.5091 X_i \\ \text{se} &= (6.4138) \qquad (0.0357) \qquad r^2 = 0.9621 \\ t &= (3.8128) \qquad (14.2605) \qquad \text{df} = 8 \\ p &= (0.002571) \qquad (0.000000289) \qquad F_{1,8} = 202.87 \end{aligned} \qquad \textbf{(5.11.1)}$$

In Eq. (5.11.1) the figures in the first set of parentheses are the estimated standard errors of the regression coefficients, the figures in the second set are estimated t values computed from (5.3.2) under the null hypothesis that

the true population value of each regression coefficient individually is zero (e.g., $3.8128 = 24.4545 \div 6.4138$), and the figures in the third set are the estimated p values. Thus, for 8 df the probability of obtaining a t value of 3.8128 or greater is 0.0026 and the probability of obtaining a t value of 14.2605 or larger is about 0.0000003.

By presenting the p values of the estimated t coefficients, we can see at once the exact level of significance of each estimated t value. Thus, under the null hypothesis that the true population intercept value is zero, the exact probability (i.e., the p value) of obtaining a t value of 3.8128 or greater is only about 0.0026. Therefore, if we reject this null hypothesis, the probability of our committing a Type I error is about 26 in 10,000, a very small probability indeed. For all practical purposes we can say that the true population intercept is different from zero. Likewise, the p value of the estimated slope coefficient is zero for all practical purposes. If the true MPC were in fact zero, our chances of obtaining an MPC of 0.5091 would be practically zero. Hence we can reject the null hypothesis that the true MPC is zero.

Earlier we showed the intimate connection between the F and t statistics, namely, $F_{1,k} = t_k^2$. Under the null hypothesis that the true $\beta_2 = 0$, (5.11.1) shows that the F value is 202.87 (for 1 numerator and 8 denominator df) and the t value is about 14.24 (8 df); as expected, the former value is the square of the latter value, except for the roundoff errors. The ANOVA table for this problem has already been discussed.

5.12 EVALUATING THE RESULTS OF REGRESSION ANALYSIS

In Figure I.4 of the Introduction we sketched the anatomy of econometric modeling. Now that we have presented the results of regression analysis of our consumption–income example in (5.11.1), we would like to question the adequacy of the fitted model. How "good" is the fitted model? We need some criteria with which to answer this question.

First, are the signs of the estimated coefficients in accordance with theoretical or prior expectations? A priori, β_2, the marginal propensity to consume (MPC) in the consumption function, should be positive. In the present example it is. Second, if theory says that the relationship should be not only positive but also statistically significant, is this the case in the present application? As we discussed in Section 5.11, the MPC is not only positive but also statistically significantly different from zero; the p value of the estimated t value is extremely small. The same comments apply about the intercept coefficient. Third, how well does the regression model explain variation in the consumption expenditure? One can use r^2 to answer this question. In the present example r^2 is about 0.96, which is a very high value considering that r^2 can be at most 1.

Thus, the model we have chosen for explaining consumption expenditure behavior seems quite good. But before we sign off, we would like to find out

whether our model satisfies the assumptions of CNLRM. We will not look at the various assumptions now because the model is patently so simple. But there is one assumption that we would like to check, namely, the normality of the disturbance term, u_i. Recall that the t and F tests used before require that the error term follow the normal distribution. Otherwise, the testing procedure will not be valid in small, or finite, samples.

Normality Tests

Although several tests of normality are discussed in the literature, we will consider just three: (1) histogram of residuals; (2) normal probability plot (NPP), a graphical device; and (3) the **Jarque–Bera** test.

Histogram of Residuals. A histogram of residuals is a simple graphic device that is used to learn something about the shape of the PDF of a random variable. On the horizontal axis, we divide the values of the variable of interest (e.g., OLS residuals) into suitable intervals, and in each class interval we erect rectangles equal in height to the number of observations (i.e., frequency) in that class interval. If you mentally superimpose the bell-shaped normal distribution curve on the histogram, you will get some idea as to whether normal (PDF) approximation may be appropriate. A concrete example is given in Section 5.13 (see Figure 5.8). It is always a good practice to plot the histogram of the residuals as a rough and ready method of testing for the normality assumption.

Normal Probability Plot. A comparatively simple graphical device to study the shape of the probability density function (PDF) of a random variable is the **normal probability plot (NPP)** which makes use of *normal probability paper*, a specially designed graph paper. On the horizontal, or x, axis, we plot values of the variable of interest (say, OLS residuals, $\hat{u}_i$), and on the vertical, or y, axis, we show the expected value of this variable if it were normally distributed. Therefore, if the variable is in fact from the normal population, the NPP will be approximately a straight line. The NPP of the residuals from our consumption–income regression is shown in Figure 5.7, which is obtained from the MINITAB software package, version 13. As noted earlier, if the fitted line in the NPP is approximately a straight line, one can conclude that the variable of interest is normally distributed. In Figure 5.7, we see that residuals from our illustrative example are approximately normally distributed, because a straight line seems to fit the data reasonably well.

MINITAB also produces the **Anderson–Darling normality test,** known as the A^2 **statistic.** The underlying null hypothesis is that the variable under consideration is normally distributed. As Figure 5.7 shows, for our example, the computed A^2 statistic is 0.394. The *p value* of obtaining such a value of A^2 is 0.305, which is reasonably high. Therefore, we do not reject the

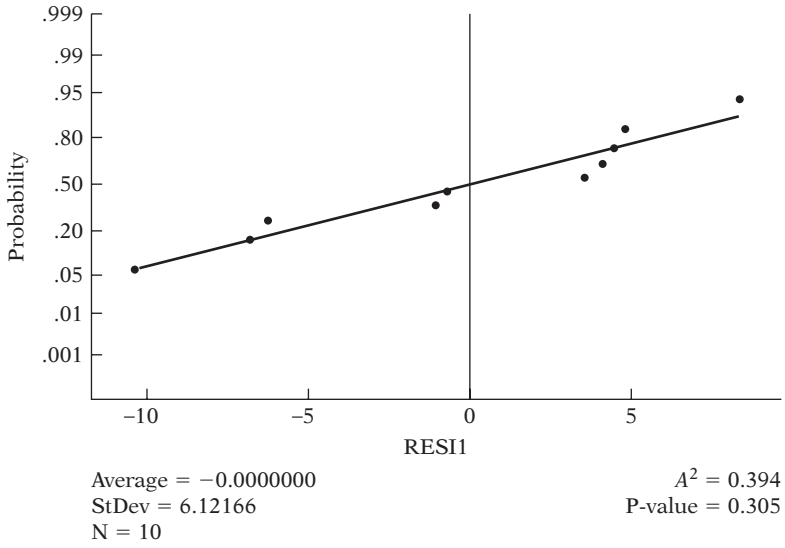


FIGURE 5.7 Residuals from consumption–income regression.

hypothesis that the residuals from our consumption–income example are normally distributed. Incidentally, Figure 5.7 shows the parameters of the (normal) distribution, the mean is approximately 0 and the standard deviation is about 6.12.

Jarque–Bera (JB) Test of Normality.[20] The JB test of normality is an *asymptotic,* or large-sample, test. It is also based on the OLS residuals. This test first computes the **skewness** and **kurtosis** (discussed in **Appendix A**) measures of the OLS residuals and uses the following test statistic:

$$\text{JB} = n\left[\frac{S^2}{6} + \frac{(K-3)^2}{24}\right] \tag{5.12.1}$$

where n = sample size, S = skewness coefficient, and K = kurtosis coefficient. For a normally distributed variable, $S = 0$ and $K = 3$. Therefore, the JB test of normality is a test of the joint hypothesis that S and K are 0 and 3, respectively. In that case the value of the JB statistic is expected to be 0.

Under the null hypothesis that the residuals are normally distributed, Jarque and Bera showed that *asymptotically (i.e., in large samples) the JB statistic given in (5.12.1) follows the chi-square distribution with 2 df*. If the computed p value of the JB statistic in an application is sufficiently low, which will happen if the value of the statistic is very different from 0, one can reject the hypothesis that the residuals are normally distributed. But if

[20]See C. M. Jarque and A. K. Bera, "A Test for Normality of Observations and Regression Residuals," *International Statistical Review*, vol. 55, 1987, pp. 163–172.

the p value is reasonably high, which will happen if the value of the statistic is close to zero, we do not reject the normality assumption.

The sample size in our consumption–income example is rather small. Hence, strictly speaking one should not use the JB statistic. If we mechanically apply the JB formula to our example, the JB statistic turns out to be 0.7769. The p value of obtaining such a value from the chi-square distribution with 2 df is about 0.68, which is quite high. In other words, we may not reject the normality assumption for our example. Of course, bear in mind the warning about the sample size.

Other Tests of Model Adequacy

Remember that the CNLRM makes many more assumptions than the normality of the error term. As we examine econometric theory further, we will consider several tests of model adequacy (see Chapter 13). Until then, keep in mind that our regression modeling is based on several simplifying assumptions that may not hold in each and every case.

A CONCLUDING EXAMPLE

Let us return to Example 3.2 about food expenditure in India. Using the data given in (3.7.2) and adopting the format of (5.11.1), we obtain the following expenditure equation:

$$\begin{aligned} \widehat{\text{FoodExp}}_i &= 94.2087 + 0.4368\ \text{TotalExp}_i \\ \text{se} &= (50.8563)\quad (0.0783) \\ t &= (1.8524)\quad (5.5770) \\ p &= (0.0695)\quad (0.0000)^* \\ r^2 &= 0.3698; \quad \text{df} = 53 \\ F_{1,53} &= 31.1034 \quad (p \text{ value} = 0.0000)^* \end{aligned} \tag{5.12.2}$$

where * denotes extremely small.

First, let us interpret this regression. As expected, there is a positive relationship between expenditure on food and total expenditure. If total expenditure went up by a rupee, on average, expenditure on food increased by about 44 paise. If total expenditure were zero, the average expenditure on food would be about 94 rupees. Of course, this mechanical interpretation of the intercept may not make much economic sense. The r^2 value of about 0.37 means that 37 percent of the variation in food expenditure is explained by total expenditure, a proxy for income.

Suppose we want to test the null hypothesis that there is no relationship between food expenditure and total expenditure, that is, the true slope coefficient $\beta_2 = 0$. The estimated value of β_2 is 0.4368. If the null hypothesis were true, what is the probability of obtaining a value of 0.4368? Under the null hypothesis, we observe from (5.12.2) that the t value is 5.5770 and the p value of obtaining such a t value is practically zero. In other words, we can reject the null hypothesis resoundingly. But suppose the null hypothesis were that $\beta_2 = 0.5$. Now what? Using the t test we obtain:

$$t = \frac{0.4368 - 0.5}{0.0783} = -0.8071$$

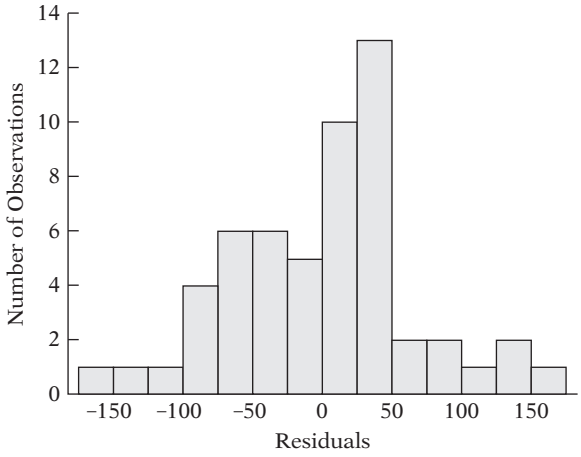
The probability of obtaining a $|t|$ of 0.8071 is greater than 20 percent. Hence we do not reject the hypothesis that the true β_2 is 0.5.

Notice that, under the null hypothesis, the true slope coefficient is zero, the F value is 31.1034, as shown in (5.12.2). Under the same null hypothesis, we obtained a t value of 5.5770. If we square this value, we obtain 31.1029, which is about the same as the F value, again showing the close relationship between the t and the F statistic. (*Note:* The numerator df for the F statistic must be 1, which is the case here.)

Using the estimated residuals from the regression, what can we say about the probability distribution of the error term? The information is given in Figure 5.8. As the figure shows, the residuals from the food expenditure regression seem to be symmetrically distributed. Application of the Jarque–Bera test shows that the JB statistic is about 0.2576, and the probability of obtaining such a statistic under the normality assumption is about

(*Continued*)

A CONCLUDING EXAMPLE (*Continued*)



Series: Residuals	
Sample 1 55	
Observations 55	
Mean	-1.19×10^{-14}
Median	7.747849
Maximum	171.5859
Minimum	−153.7664
Std. dev.	66.23382
Skewness	0.119816
Kurtosis	3.234473
Jarque–Bera	0.257585
Probability	0.879156

FIGURE 5.8 Residuals from the food expenditure regression.

88 percent. Therefore, we do not reject the hypothesis that the error terms are normally distributed. But keep in mind that the sample size of 55 observations may not be large enough.

We leave it to the reader to establish confidence intervals for the two regression coefficients as well as to obtain the normal probability plot and do mean and individual predictions.

5.13 SUMMARY AND CONCLUSIONS

1. Estimation and hypothesis testing constitute the two main branches of classical statistics. Having discussed the problem of estimation in Chapters 3 and 4, we have taken up the problem of hypothesis testing in this chapter.

2. Hypothesis testing answers this question: Is a given finding compatible with a stated hypothesis or not?

3. There are two mutually complementary approaches to answering the preceding question: **confidence interval** and **test of significance.**

4. Underlying the confidence-interval approach is the concept of **interval estimation.** An interval estimator is an interval or range constructed in such a manner that it has a specified probability of including within its limits the true value of the unknown parameter. The interval thus constructed is known as a **confidence interval,** which is often stated in percent form, such as 90 or 95%. The confidence interval provides a set of plausible hypotheses about the value of the unknown parameter. If the null-hypothesized value lies in the confidence interval, the hypothesis is not rejected, whereas if it lies outside this interval, the null hypothesis can be rejected.

5. In the **significance test** procedure, one develops a **test statistic** and examines its sampling distribution under the null hypothesis. The test statistic usually follows a well-defined probability distribution such as the normal, t, F, or chi-square. Once a test statistic (e.g., the t statistic) is computed

from the data at hand, its p value can be easily obtained. The p value gives the exact probability of obtaining the estimated test statistic under the null hypothesis. If this p value is small, one can reject the null hypothesis, but if it is large one may not reject it. What constitutes a small or large p value is up to the investigator. In choosing the p value the investigator has to bear in mind the probabilities of committing **Type I** and **Type II errors.**

6. In practice, one should be careful in fixing α, the probability of committing a **Type I error,** at arbitrary values such as 1, 5, or 10 percent. It is better to quote the ***p* value** of the test statistic. Also, the statistical significance of an estimate should not be confused with its practical significance.

7. Of course, hypothesis testing presumes that the model chosen for empirical analysis is adequate in the sense that it does not violate one or more assumptions underlying the classical normal linear regression model. Therefore, tests of model adequacy should precede tests of hypothesis. This chapter introduced one such test, the **normality test,** to find out whether the error term follows the normal distribution. Since in small, or finite, samples, the t, F, and chi-square tests require the normality assumption, it is important that this assumption be checked formally.

8. If the model is deemed practically adequate, it may be used for forecasting purposes. But in forecasting the future values of the regressand, one should not go too far out of the sample range of the regressor values. Otherwise, forecasting errors can increase dramatically.

EXERCISES

Questions

5.1. State with reason whether the following statements are true, false, or uncertain. Be precise.

- **a.** The t test of significance discussed in this chapter requires that the sampling distributions of estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ follow the normal distribution.
- **b.** Even though the disturbance term in the CLRM is not normally distributed, the OLS estimators are still unbiased.
- **c.** If there is no intercept in the regression model, the estimated u_i $(= \hat{u}_i)$ will not sum to zero.
- **d.** The p value and the size of a test statistic mean the same thing.
- **e.** In a regression model that contains the intercept, the sum of the residuals is always zero.
- **f.** If a null hypothesis is not rejected, it is true.
- **g.** The higher the value of σ^2, the larger is the variance of $\hat{\beta}_2$ given in (3.3.1).
- **h.** The conditional and unconditional means of a random variable are the same things.
- **i.** In the two-variable PRF, if the slope coefficient β_2 is zero, the intercept β_1 is estimated by the sample mean $\bar{Y}$.
- **j.** The conditional variance, var $(Y_i \mid X_i) = \sigma^2$, and the unconditional variance of Y, var $(Y) = \sigma_Y^2$, will be the same if X had no influence on Y.

5.2. Set up the ANOVA table in the manner of Table 5.4 for the regression model given in (3.7.2) and test the hypothesis that there is no relationship between food expenditure and total expenditure in India.

5.3. From the data given in Table 2.6 on earnings and education, we obtained the following regression [see Eq. (3.7.3)]:

$$\widehat{\text{Meanwage}}_i = 0.7437 + 0.6416\ \text{Education}_i$$
$$\text{se} = (0.8355) \quad (\quad)$$
$$t = (\quad) \quad (9.6536) \qquad r^2 = 0.8944 \qquad n = 13$$

a. Fill in the missing numbers.
b. How do you interpret the coefficient 0.6416?
c. Would you reject the hypothesis that education has no effect whatsoever on wages? Which test do you use? And why? What is the p value of your test statistic?
d. Set up the ANOVA table for this example and test the hypothesis that the slope coefficient is zero. Which test do you use and why?
e. Suppose in the regression given above the r^2 value was not given to you. Could you have obtained it from the other information given in the regression?

5.4. Let ρ^2 represent the true population coefficient of correlation. Suppose you want to test the hypothesis that $\rho^2 = 0$. Verbally explain how you would test this hypothesis. *Hint:* Use Eq. (3.5.11). See also exercise 5.7.

5.5. What is known as the **characteristic line** of modern investment analysis is simply the regression line obtained from the following model:

$$r_{it} = \alpha_i + \beta_i r_{mt} + u_t$$

where r_{it} = the rate of return on the ith security in time t
r_{mt} = the rate of return on the market portfolio in time t
u_t = stochastic disturbance term

In this model β_i is known as the **beta coefficient** of the ith security, a measure of market (or systematic) risk of a security.[*]

On the basis of 240 monthly rates of return for the period 1956–1976, Fogler and Ganapathy obtained the following characteristic line for IBM stock in relation to the market portfolio index developed at the University of Chicago[†]:

$$\hat{r}_{it} = 0.7264 + 1.0598 r_{mt} \qquad r^2 = 0.4710$$
$$\text{se} = (0.3001)\,(0.0728) \qquad \text{df} = 238$$
$$F_{1,238} = 211.896$$

a. A security whose beta coefficient is greater than one is said to be a volatile or aggressive security. Was IBM a volatile security in the time period under study?

[*]See Haim Levy and Marshall Sarnat, *Portfolio and Investment Selection: Theory and Practice*, Prentice-Hall International, Englewood Cliffs, N.J., 1984, Chap. 12.

[†]H. Russell Fogler and Sundaram Ganapathy, *Financial Econometrics*, Prentice Hall, Englewood Cliffs, N.J., 1982, p. 13.

b. Is the intercept coefficient significantly different from zero? If it is, what is its practical meaning?

5.6. Equation (5.3.5) can also be written as

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

That is, the weak inequality ($\leq$) can be replaced by the strong inequality ($<$). Why?

5.7. R. A. Fisher has derived the sampling distribution of the correlation coefficient defined in (3.5.13). If it is assumed that the variables X and Y are jointly normally distributed, that is, if they come from a bivariate normal distribution (see Appendix 4A, exercise 4.1), then under the assumption that the population correlation coefficient ρ is zero, it can be shown that $t = r\sqrt{n-2}/\sqrt{1-r^2}$ follows Student's t distribution with $n-2$ df.* Show that this t value is identical with the t value given in (5.3.2) under the null hypothesis that $\beta_2 = 0$. Hence establish that under the same null hypothesis $F = t^2$. (See Section 5.9.)

Problems

5.8. Consider the following regression output†:

$$\begin{aligned}\hat{Y}_i &= 0.2033 + 0.6560X_t \\ \text{se} &= (0.0976)\ \ (0.1961) \\ r^2 &= 0.397 \qquad \text{RSS} = 0.0544 \qquad \text{ESS} = 0.0358\end{aligned}$$

where Y = labor force participation rate (LFPR) of women in 1972 and X = LFPR of women in 1968. The regression results were obtained from a sample of 19 cities in the United States.

a. How do you interpret this regression?

b. Test the hypothesis: $H_0: \beta_2 = 1$ against $H_1: \beta_2 > 1$. Which test do you use? And why? What are the underlying assumptions of the test(s) you use?

c. Suppose that the LFPR in 1968 was 0.58 (or 58 percent). On the basis of the regression results given above, what is the mean LFPR in 1972? Establish a 95% confidence interval for the mean prediction.

d. How would you test the hypothesis that the error term in the population regression is normally distribute? Show the necessary calculations.

5.9. Table 5.5 gives data on average public teacher pay (annual salary in dollars) and spending on public schools per pupil (dollars) in 1985 for 50 states and the District of Columbia.

*If ρ is in fact zero, Fisher has shown that r follows the same t distribution provided either X or Y is normally distributed. But if ρ is not equal to zero, both variables must be normally distributed. See R. L. Anderson and T. A. Bancroft, *Statistical Theory in Research*, McGraw-Hill, New York, 1952, pp. 87–88.

†Adapted from Samprit Chatterjee, Ali S. Hadi, and Bertram Price, *Regression Analysis by Example*, 3d ed., Wiley Interscience, New York, 2000, pp. 46–47.

TABLE 5.5 AVERAGE SALARY AND PER PUPIL SPENDING (DOLLARS), 1985

Observation	Salary	Spending	Observation	Salary	Spending
1	19,583	3346	27	22,795	3366
2	20,263	3114	28	21,570	2920
3	20,325	3554	29	22,080	2980
4	26,800	4642	30	22,250	3731
5	29,470	4669	31	20,940	2853
6	26,610	4888	32	21,800	2533
7	30,678	5710	33	22,934	2729
8	27,170	5536	34	18,443	2305
9	25,853	4168	35	19,538	2642
10	24,500	3547	36	20,460	3124
11	24,274	3159	37	21,419	2752
12	27,170	3621	38	25,160	3429
13	30,168	3782	39	22,482	3947
14	26,525	4247	40	20,969	2509
15	27,360	3982	41	27,224	5440
16	21,690	3568	42	25,892	4042
17	21,974	3155	43	22,644	3402
18	20,816	3059	44	24,640	2829
19	18,095	2967	45	22,341	2297
20	20,939	3285	46	25,610	2932
21	22,644	3914	47	26,015	3705
22	24,624	4517	48	25,788	4123
23	27,186	4349	49	29,132	3608
24	33,990	5020	50	41,480	8349
25	23,382	3594	51	25,845	3766
26	20,627	2821			

Source: National Education Association, as reported by *Albuquerque Tribune,* Nov. 7, 1986.

To find out if there is any relationship between teacher's pay and per pupil expenditure in public schools, the following model was suggested: $\text{Pay}_i = \beta_1 + \beta_2\ \text{Spend}_i + u_i$, where Pay stands for teacher's salary and Spend stands for per pupil expenditure.

a. Plot the data and eyeball a regression line.

b. Suppose on the basis of **a** you decide to estimate the above regression model. Obtain the estimates of the parameters, their standard errors, r^2, RSS, and ESS.

c. Interpret the regression. Does it make economic sense?

d. Establish a 95% confidence interval for β_2. Would you reject the hypothesis that the true slope coefficient is 3.0?

e. Obtain the mean and individual forecast value of Pay if per pupil spending is $5000. Also establish 95% confidence intervals for the true mean and individual values of Pay for the given spending figure.

f. How would you test the assumption of the normality of the error term? Show the test(s) you use.

5.10. Refer to exercise 3.20 and set up the ANOVA tables and test the hypothesis that there is no relationship between productivity and real wage

compensation. Do this for both the business and nonfarm business sectors.

5.11. Refer to exercise 1.7.
- **a.** Plot the data with impressions on the vertical axis and advertising expenditure on the horizontal axis. What kind of relationship do you observe?
- **b.** Would it be appropriate to fit a bivariate linear regression model to the data? Why or why not? If not, what type of regression model will you fit the data to? Do we have the necessary tools to fit such a model?
- **c.** Suppose you do not plot the data and simply fit the bivariate regression model to the data. Obtain the usual regression output. Save the results for a later look at this problem.

5.12. Refer to exercise 1.1.
- **a.** Plot the U.S. Consumer Price Index (CPI) against the Canadian CPI. What does the plot show?
- **b.** Suppose you want to predict the U.S. CPI on the basis of the Canadian CPI. Develop a suitable model.
- **c.** Test the hypothesis that there is no relationship between the two CPIs. Use $\alpha = 5\%$. If you reject the null hypothesis, does that mean the Canadian CPI "causes" the U.S. CPI? Why or why not?

5.13. Refer to exercise 3.22.
- **a.** Estimate the two regressions given there, obtaining standard errors and the other usual output.
- **b.** Test the hypothesis that the disturbances in the two regression models are normally distributed.
- **c.** In the gold price regression, test the hypothesis that $\beta_2 = 1$, that is, there is a one-to-one relationship between gold prices and CPI (i.e., gold is a perfect hedge). What is the p value of the estimated test statistic?
- **d.** Repeat step **c** for the NYSE Index regression. Is investment in the stock market a perfect hedge against inflation? What is the null hypothesis you are testing? What is its p value?
- **e.** Between gold and stock, which investment would you choose? What is the basis of your decision?

5.14. Table 5.6 gives data on GNP and four definitions of the money stock for the United States for 1970–1983. Regressing GNP on the various definitions of money, we obtain the results shown in Table 5.7.

The monetarists or quantity theorists maintain that nominal income (i.e., nominal GNP) is largely determined by changes in the quantity or the stock of money, although there is no consensus as to the "right" definition of money. Given the results in the preceding table, consider these questions:
- **a.** Which definition of money seems to be closely related to nominal GNP?
- **b.** Since the r^2 terms are uniformly high, does this fact mean that our choice for definition of money does not matter?
- **c.** If the Fed wants to control the money supply, which one of these money measures is a better target for that purpose? Can you tell from the regression results?

5.15. Suppose the equation of an **indifference curve** between two goods is

$$X_i Y_i = \beta_1 + \beta_2 X_i$$

TABLE 5.6 GNP AND FOUR MEASURES OF MONEY STOCK

Year	GNP, $ billion	Money stock measure, $ billion			
		M_1	M_2	M_3	L
1970	992.70	216.6	628.2	677.5	816.3
1971	1,077.6	230.8	712.8	776.2	903.1
1972	1,185.9	252.0	805.2	886.0	1,023.0
1973	1,326.4	265.9	861.0	985.0	1,141.7
1974	1,434.2	277.6	908.5	1,070.5	1,249.3
1975	1,549.2	291.2	1,023.3	1,174.2	1,367.9
1976	1,718.0	310.4	1,163.6	1,311.9	1,516.6
1977	1,918.3	335.4	1,286.7	1,472.9	1,704.7
1978	2,163.9	363.1	1,389.1	1,647.1	1,910.6
1979	2,417.8	389.1	1,498.5	1,804.8	2,117.1
1980	2,631.7	414.9	1,632.6	1,990.0	2,326.2
1981	2,957.8	441.9	1,796.6	2,238.2	2,599.8
1982	3,069.3	480.5	1,965.4	2,462.5	2,870.8
1983	3,304.8	525.4	2,196.3	2,710.4	3,183.1

Definitions:
M_1 = currency + demand deposits + travelers checks and other checkable deposits (OCDs)
$M_2 = M_1$ + overnight RPs and Eurodollars + MMMF (money market mutual fund) balances + MMDAs (money market deposit accounts) + savings and small deposits
$M_3 = M_2$ + large time deposits + term RPs + Institutional MMMF
L = M_3 + other liquid assets
Source: Economic Report of the President, 1985, GNP data from Table B-1, p. 232; money stock data from Table B-61, p. 303.

TABLE 5.7 GNP–MONEY STOCK REGRESSIONS, 1970–1983

1)	$\widehat{\text{GNP}}_t = -787.4723 + 8.0863\ M_{1t}$	(77.9664) (0.2197)	$r^2 = 0.9912$
2)	$\widehat{\text{GNP}}_t = -44.0626 + 1.5875\ M_{2t}$	(61.0134) (0.0448)	$r^2 = 0.9905$
3)	$\widehat{\text{GNP}}_t = 159.1366 + 1.2034\ M_{3t}$	(42.9882) (0.0262)	$r^2 = 0.9943$
4)	$\widehat{\text{GNP}}_t = 164.2071 + 1.0290\ L_t$	(44.7658) (0.0234)	$r^2 = 0.9938$

Note: The figures in parentheses are the estimated standard errors.

TABLE 5.8

Consumption of good X:	1	2	3	4	5
Consumption of good Y:	4	3.5	2.8	1.9	0.8

How would you estimate the parameters of this model? Apply the preceding model to the data in Table 5.8 and comment on your results.

5.16. Since 1986 the *Economist* has been publishing the Big Mac Index as a crude, and hilarious, measure of whether international currencies are at their "correct" exchange rate, as judged by the theory of **purchasing power parity (PPP).** The PPP holds that a unit of currency should be able

to buy the same bundle of goods in all countries. The proponents of PPP argue that, in the long run, currencies tend to move toward their PPP. The *Economist* uses McDonald's Big Mac as a representative bundle and gives the information in Table 5.9.

Consider the following regression model:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

where Y = actual exchange rate and X = implied PPP of the dollar.

TABLE 5.9 THE HAMBURGER STANDARD

	Big Mac prices		Implied PPP* of the dollar	Actual $ exchange rate April 17, 2001	Under (−)/ over (+) valuation against the dollar, %
	In local currency	In dollars			
United States†	$2.54	2.54	–	–	–
Argentina	Peso2.50	2.50	0.98	1.00	−2
Australia	A$3.00	1.52	1.18	1.98	−40
Brazil	Real3.60	1.64	1.42	2.19	−35
Britain	£1.99	2.85	1.28‡	1.43‡	12
Canada	C$3.33	2.14	1.31	1.56	−16
Chile	Peso1260	2.10	496	601	−17
China	Yuan9.90	1.20	3.90	8.28	−53
Czech Rep	Koruna56.00	1.43	22.0	39.0	−44
Denmark	DKr24.75	2.93	9.74	8.46	15
Euro area	€2.57	2.27	0.99§	0.88§	−11
France	FFr18.5	2.49	7.28	7.44	−2
Germany	DM5.10	2.30	2.01	2.22	−9
Italy	Lire4300	1.96	1693	2195	−23
Spain	Pta395	2.09	156	189	−18
Hong Kong	HK$10.70	1.37	4.21	7.80	−46
Hungary	Forint399	1.32	157	303	−48
Indonesia	Rupiah14700	1.35	5787	10855	−47
Japan	¥294	2.38	116	124	−6
Malaysia	M$4.52	1.19	1.78	3.80	−53
Mexico	Peso21.9	2.36	8.62	9.29	−7
New Zealand	NZ$3.60	1.46	1.42	2.47	−43
Philippines	Peso59.00	1.17	23.2	50.3	−54
Poland	Zloty5.90	1.46	2.32	4.03	−42
Russia	Rouble35.00	1.21	13.8	28.9	−52
Singapore	S$3.30	1.82	1.30	1.81	−28
South Africa	Rand9.70	1.19	3.82	8.13	−53
South Korea	Won3000	2.27	1181	1325	−11
Sweden	SKr24.0	2.33	9.45	10.28	−8
Switzerland	SFr6.30	3.65	2.48	1.73	44
Taiwan	NT$70.0	2.13	27.6	32.9	−16
Thailand	Baht55.0	1.21	21.7	45.5	−52

*Purchasing power parity: local price divided by price in the United States.
†Average of New York, Chicago, San Francisco, and Atlanta.
‡Dollars per pound.
§Dollars per euro.
Source: McDonald's; *The Economist*, April 21, 2001.

a. If the PPP holds, what values of β_1 and β_2 would you expect a priori?
b. Do the regression results support your expectation? What formal test do you use to test your hypothesis?
c. Should the *Economist* continue to publish the Big Mac Index? Why or why not?

5.17. Refer to the S.A.T. data given in exercise 2.16. Suppose you want to predict the male math (Y) scores on the basis of the female math scores (X) by running the following regression:

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

a. Estimate the preceding model.
b. From the estimated residuals, find out if the normality assumption can be sustained.
c. Now test the hypothesis that $\beta_2 = 1$, that is, there is a one-to-one correspondence between male and female math scores.
d. Set up the ANOVA table for this problem.

5.18. Repeat the exercise in the preceding problem but let Y and X denote the male and female verbal scores, respectively.

5.19. Table 5.10 gives annual data on the Consumer Price Index (CPI) and the Wholesale Price Index (WPI), also called Producer Price Index (PPI), for the U.S. economy for the period 1960–1999.
a. Plot the CPI on the vertical axis and the WPI on the horizontal axis. A priori, what kind of relationship do you expect between the two indexes? Why?

TABLE 5.10 CPI AND WPI, UNITED STATES, 1960–1999

Year	CPI	WPI	Year	CPI	WPI
1960	29.8	31.7	1980	86.3	93.8
1961	30.0	31.6	1981	94.0	98.8
1962	30.4	31.6	1982	97.6	100.5
1963	30.9	31.6	1983	101.3	102.3
1964	31.2	31.7	1984	105.3	103.5
1965	31.8	32.8	1985	109.3	103.6
1966	32.9	33.3	1986	110.5	99.70
1967	33.9	33.7	1987	115.4	104.2
1968	35.5	34.6	1988	120.5	109.0
1969	37.7	36.3	1989	126.1	113.0
1970	39.8	37.1	1990	133.8	118.7
1971	41.1	38.6	1991	137.9	115.9
1972	42.5	41.1	1992	141.9	117.6
1973	46.2	47.4	1993	145.8	118.6
1974	51.9	57.3	1994	149.7	121.9
1975	55.5	59.7	1995	153.5	125.7
1976	58.2	62.5	1996	158.6	128.8
1977	62.1	66.2	1997	161.3	126.7
1978	67.7	72.7	1998	163.9	122.7
1979	76.7	83.4	1999	168.3	128.0

Source: Economic Report of the President, 2000, pp. 373 and 379.

b. Suppose you want to predict one of these indexes on the basis of the other index. Which will you use as the regressand and which as the regressor? Why?
c. Run the regression you have decided in *b*. Show the standard output. Test the hypothesis that there is a one-to-one relationship between the two indexes.
d. From the residuals obtained from the regression in **c**, can you entertain the hypothesis that the true error term is normally distributed? Show the tests you use.

APPENDIX 5A

5A.1 PROBABILITY DISTRIBUTIONS RELATED TO THE NORMAL DISTRIBUTION

The ***t*, chi-square** (χ^2), and ***F*** probability distributions, whose salient features are discussed in **Appendix A,** are intimately related to the normal distribution. Since we will make heavy use of these probability distributions in the following chapters, we summarize their relationship with the normal distribution in the following theorem; the proofs, which are beyond the scope of this book, can be found in the references.[1]

Theorem 5.1. If $Z_1, Z_2, \ldots, Z_n$ are normally and independently distributed random variables such that $Z_i \sim N(\mu_i, \sigma_i^2)$, then the sum $Z = \sum k_i Z_i$, where k_i are constants not all zero, is also distributed normally with mean $\sum k_i \mu_i$ and variance $\sum k_i^2 \sigma_i^2$; that is, $Z \sim N(\sum k_i \mu_i, \sum k_i^2 \sigma_i^2)$. *Note:* μ denotes the mean value.

In short, linear combinations of normal variables are themselves normally distributed. For example, if Z_1 and Z_2 are normally and independently distributed as $Z_1 \sim N(10, 2)$ and $Z_2 \sim N(8, 1.5)$, then the linear combination $Z = 0.8Z_1 + 0.2Z_2$ is also normally distributed with mean $= 0.8(10) + 0.2(8) = 9.6$ and variance $= 0.64(2) + 0.04(1.5) = 1.34$, that is, $Z \sim (9.6, 1.34)$.

Theorem 5.2. If $Z_1, Z_2, \ldots, Z_n$ are normally distributed but are not independent, the sum $Z = \sum k_i Z_i$, where k_i are constants not all zero, is also normally distributed with mean $\sum k_i \mu_i$ and variance $[\sum k_i^2 \sigma_i^2 + 2 \sum k_i k_j \operatorname{cov}(Z_i, Z_j), i \neq j]$.

Thus, if $Z_1 \sim N(6, 2)$ and $Z_2 \sim N(7, 3)$ and $\operatorname{cov}(Z_1, Z_2) = 0.8$, then the linear combination $0.6Z_1 + 0.4Z_2$ is also normally distributed with mean $= 0.6(6) + 0.4(7) = 6.4$ and variance $= [0.36(2) + 0.16(3) + 2(0.6)(0.4)(0.8)] = 1.584$.

[1]For proofs of the various theorems, see Alexander M. Mood, Franklin A. Graybill, and Duane C. Bose, *Introduction to the Theory of Statistics,* 3d ed., McGraw-Hill, New York, 1974, pp. 239–249.

Theorem 5.3. If $Z_1, Z_2, \ldots, Z_n$ are normally and independently distributed random variables such that each $Z_i \sim N(0, 1)$, that is, a standardized normal variable, then $\sum Z_i^2 = Z_1^2 + Z_2^2 + \cdots + Z_n^2$ follows the chi-square distribution with n df. Symbolically, $\sum Z_i^2 \sim \chi_n^2$, where n denotes the degrees of freedom, df.

In short, "the sum of the squares of independent standard normal variables has a chi-square distribution with degrees of freedom equal to the number of terms in the sum."[2]

Theorem 5.4. If $Z_1, Z_2, \ldots, Z_n$ are independently distributed random variables each following chi-square distribution with k_i df, then the sum $\sum Z_i = Z_1 + Z_2 + \cdots + Z_n$ also follows a chi-square distribution with $k = \sum k_i$ df.

Thus, if Z_1 and Z_2 are independent χ^2 variables with df of k_1 and k_2, respectively, then $Z = Z_1 + Z_2$ is also a χ^2 variable with $(k_1 + k_2)$ degrees of freedom. This is called the **reproductive property** of the χ^2 distribution.

Theorem 5.5. If Z_1 is a standardized normal variable $[Z_1 \sim N(0, 1)]$ and another variable Z_2 follows the chi-square distribution with k df and is independent of Z_1, then the variable defined as

$$t = \frac{Z_1}{\sqrt{Z_2}/\sqrt{k}} = \frac{Z_1\sqrt{k}}{\sqrt{Z_2}} = \frac{\text{standard normal variable}}{\sqrt{\text{independent chi-square variable/df}}} \sim t_k$$

follows Student's t distribution with k df. *Note:* This distribution is discussed in **Appendix A** and is illustrated in Chapter 5.

Incidentally, note that as k, the df, increases indefinitely (i.e., as $k \to \infty$), the Student's t distribution approaches the standardized normal distribution.[3] As a matter of convention, the notation t_k means Student's t distribution or variable with k df.

Theorem 5.6. If Z_1 and Z_2 are independently distributed chi-square variables with k_1 and k_2 df, respectively, then the variable

$$F = \frac{Z_1/k_1}{Z_2/k_2} \sim F_{k_1,k_2}$$

has the F distribution with k_1 and k_2 degrees of freedom, where k_1 is known as the **numerator degrees of freedom** and k_2 the **denominator degrees of freedom.**

[2]Ibid., p. 243.

[3]For proof, see Henri Theil, *Introduction to Econometrics*, Prentice-Hall, Englewood Cliffs, N.J., 1978, pp. 237–245.

Again as a matter of convention, the notation F_{k_1,k_2} means an F variable with k_1 and k_2 degrees of freedom, the df in the numerator being quoted first.

In other words, Theorem 5.6 states that the F variable is simply the ratio of two independently distributed chi-square variables divided by their respective degrees of freedom.

Theorem 5.7. The square of (Student's) t variable with k df has an F distribution with $k_1 = 1$ df in the numerator and $k_2 = k$ df in the denominator.[4] That is,

$$F_{1,k} = t_k^2$$

Note that for this equality to hold, the numerator df of the F variable must be 1. Thus, $F_{1,4} = t_4^2$ or $F_{1,23} = t_{23}^2$ and so on.

As noted, we will see the practical utility of the preceding theorems as we progress.

Theorem 5.8. For large denominator df, the numerator df times the F value is approximately equal to the chi-square value with the numerator df. Thus,

$$m\,F_{m,n} = \chi_m^2 \qquad \text{as } n \to \infty$$

Theorem 5.9. For sufficiently large df, the chi-square distribution can be approximated by the standard normal distribution as follows:

$$Z = \sqrt{2\chi^2} - \sqrt{2k-1} \sim N(0, 1)$$

where k denotes df.

5A.2 DERIVATION OF EQUATION (5.3.2)

Let

$$Z_1 = \frac{\hat{\beta}_2 - \beta_2}{\text{se}\,(\hat{\beta}_2)} = \frac{(\hat{\beta}_2 - \beta_2)\sqrt{x_i^2}}{\sigma} \tag{1}$$

and

$$Z_2 = (n-2)\frac{\hat{\sigma}^2}{\sigma^2} \tag{2}$$

Provided σ is known, Z_1 follows the standardized normal distribution; that is, $Z_1 \sim N(0, 1)$. (Why?) Z_2, follows the χ^2 distribution with $(n-2)$ df.[5]

[4]For proof, see Eqs. (5.3.2) and (5.9.1).

[5]For proof, see Robert V. Hogg and Allen T. Craig, *Introduction to Mathematical Statistics,* 2d ed., Macmillan, New York, 1965, p. 144.

Furthermore, it can be shown that Z_2 is distributed independently of Z_1.[6] Therefore, by virtue of Theorem 5.5, the variable

$$t = \frac{Z_1\sqrt{n-2}}{\sqrt{Z_2}} \tag{3}$$

follows the t distribution with $n-2$ df. Substitution of (1) and (2) into (3) gives Eq. (5.3.2).

5A.3 DERIVATION OF EQUATION (5.9.1)

Equation (1) shows that $Z_1 \sim N(0, 1)$. Therefore, by Theorem 5.3, the preceding quantity

$$Z_1^2 = \frac{(\hat{\beta}_2 - \beta_2)^2 \sum x_i^2}{\sigma^2}$$

follows the χ^2 distribution with 1 df. As noted in Section 5A.1,

$$Z_2 = (n-2)\frac{\hat{\sigma}^2}{\sigma^2} = \frac{\sum \hat{u}_i^2}{\sigma^2}$$

also follows the χ^2 distribution with $n-2$ df. Moreover, as noted in Section 4.3, Z_2 is distributed independently of Z_1. Then from Theorem 5.6, it follows that

$$F = \frac{Z_1^2/1}{Z_2/(n-2)} = \frac{(\hat{\beta}_2 - \beta_2)^2(\sum x_i^2)}{\sum \hat{u}_i^2/(n-2)}$$

follows the F distribution with 1 and $n-2$ df, respectively. Under the null hypothesis H_0: $\beta_2 = 0$, the preceding F ratio reduces to Eq. (5.9.1).

5.A.4 DERIVATIONS OF EQUATIONS (5.10.2) AND (5.10.6)

Variance of Mean Prediction

Given $X_i = X_0$, the true mean prediction $E(Y_0 \mid X_0)$ is given by

$$E(Y_0 \mid X_0) = \beta_1 + \beta_2 X_0 \tag{1}$$

We estimate (1) from

$$\hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_0 \tag{2}$$

Taking the expectation of (2), given X_0, we get

$$\begin{aligned} E(\hat{Y}_0) &= E(\hat{\beta}_1) + E(\hat{\beta}_2)X_0 \\ &= \beta_1 + \beta_2 X_0 \end{aligned}$$

[6]For proof, see J. Johnston, *Econometric Methods,* McGraw-Hill, 3d ed., New York, 1984, pp. 181–182. (Knowledge of matrix algebra is required to follow the proof.)

because $\hat{\beta}_1$ and $\hat{\beta}_2$ are unbiased estimators. Therefore,

$$E(\hat{Y}_0) = E(Y_0 \mid X_0) = \beta_1 + \beta_2 X_0 \tag{3}$$

That is, $\hat{Y}_0$ is an unbiased predictor of $E(Y_0 \mid X_0)$.

Now using the property that $\text{var}\,(a+b) = \text{var}\,(a) + \text{var}\,(b) + 2\,\text{cov}\,(a,b)$, we obtain

$$\text{var}\,(\hat{Y}_0) = \text{var}\,(\hat{\beta}_1) + \text{var}\,(\hat{\beta}_2)X_0^2 + 2\ \text{cov}\,(\hat{\beta}_1\hat{\beta}_2)X_0 \tag{4}$$

Using the formulas for variances and covariance of $\hat{\beta}_1$ and $\hat{\beta}_2$ given in (3.3.1), (3.3.3), and (3.3.9) and manipulating terms, we obtain

$$\text{var}\,(\hat{Y}_0) = \sigma^2\left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2}\right] \qquad \textbf{= (5.10.2)}$$

Variance of Individual Prediction

We want to predict an individual Y corresponding to $X = X_0$; that is, we want to obtain

$$Y_0 = \beta_1 + \beta_2 X_0 + u_0 \tag{5}$$

We predict this as

$$\hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_0 \tag{6}$$

The prediction error, $Y_0 - \hat{Y}_0$, is

$$\begin{aligned} Y_0 - \hat{Y}_0 &= \beta_1 + \beta_2 X_0 + u_0 - (\hat{\beta}_1 + \hat{\beta}_2 X_0) \\ &= (\beta_1 - \hat{\beta}_1) + (\beta_2 - \hat{\beta}_2)X_0 + u_0 \end{aligned} \tag{7}$$

Therefore,

$$\begin{aligned} E(Y_0 - \hat{Y}_0) &= E(\beta_1 - \hat{\beta}_1) + E(\beta_2 - \hat{\beta}_2)X_0 - E(u_0) \\ &= 0 \end{aligned}$$

because $\hat{\beta}_1$, $\hat{\beta}_2$ are unbiased, X_0 is a fixed number, and $E(u_0)$ is zero by assumption.

Squaring (7) on both sides and taking expectations, we get $\text{var}\,(Y_0 - \hat{Y}_0) = \text{var}\,(\hat{\beta}_1) + X_0^2\,\text{var}\,(\hat{\beta}_2) + 2X_0\,\text{cov}\,(\beta_1, \beta_2) + \text{var}\,(u_o)$. Using the variance and covariance formulas for $\hat{\beta}_1$ and $\hat{\beta}_2$ given earlier, and noting that $\text{var}\,(u_0) = \sigma^2$, we obtain

$$\text{var}\,(Y_0 - \hat{Y}_0) = \sigma^2\left[1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2}\right] \qquad \textbf{= (5.10.6)}$$