4

Multiple Regression Analysis: Inference

This chapter continues our treatment of multiple regression analysis. We now turn to the problem of testing hypotheses about the parameters in the population regression model. We begin by finding the distributions of the OLS estimators under the added assumption that the population error is normally distributed. Sections 4.2 and 4.3 cover hypothesis testing about individual parameters, while Section 4.4 discusses how to test a single hypothesis involving more than one parameter. We focus on testing multiple restrictions in Section 4.5 and pay particular attention to determining whether a group of independent variables can be omitted from a model.

4.1 Sampling Distributions of the OLS Estimators

Up to this point, we have formed a set of assumptions under which OLS is unbiased; we have also derived and discussed the bias caused by omitted variables. In Section 3.4, we obtained the variances of the OLS estimators under the Gauss-Markov assumptions. In Section 3.5, we showed that this variance is smallest among linear unbiased estimators.

Knowing the expected value and variance of the OLS estimators is useful for describing the precision of the OLS estimators. However, in order to perform statistical inference, we need to know more than just the first two moments of $\hat{\beta}_j$; we need to know the full sampling distribution of the $\hat{\beta}_j$. Even under the Gauss-Markov assumptions, the distribution of $\hat{\beta}_j$ can have virtually any shape.

When we condition on the values of the independent variables in our sample, it is clear that the sampling distributions of the OLS estimators depend on the underlying distribution of the errors. To make the sampling distributions of the $\hat{\beta}_j$ tractable, we now assume that the unobserved error is *normally distributed* in the population. We call this the **normality assumption**.

Assumption MLR.6 (Normality)

The population error u is *independent* of the explanatory variables $x_1, x_2, \ldots, x_k$ and is normally distributed with zero mean and variance σ^2: $u \sim \text{Normal}(0,\sigma^2)$.

Assumption MLR.6 is much stronger than any of our previous assumptions. In fact, since u is independent of the x_j under MLR.6, $\text{E}(u|x_1, \ldots, x_k) = \text{E}(u) = 0$ and $\text{Var}(u|x_1, \ldots, x_k) = \text{Var}(u) = \sigma^2$. Thus, if we make Assumption MLR.6, then we are necessarily assuming MLR.4 and MLR.5. To emphasize that we are assuming more than before, we will refer to the full set of Assumptions MLR.1 through MLR.6.

For cross-sectional regression applications, Assumptions MLR.1 through MLR.6 are called the **classical linear model (CLM) assumptions**. Thus, we will refer to the model under these six assumptions as the **classical linear model**. It is best to think of the CLM assumptions as containing all of the Gauss-Markov assumptions *plus* the assumption of a normally distributed error term.

Under the CLM assumptions, the OLS estimators $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ have a stronger efficiency property than they would under the Gauss-Markov assumptions. It can be shown that the OLS estimators are the **minimum variance unbiased estimators**, which means that OLS has the smallest variance among unbiased estimators; we no longer have to restrict our comparison to estimators that are linear in the y_i. This property of OLS under the CLM assumptions is discussed further in Appendix E.

A succinct way to summarize the population assumptions of the CLM is

$$y|\mathbf{x} \sim \text{Normal}(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k, \sigma^2),$$

where $\mathbf{x}$ is again shorthand for $(x_1, \ldots, x_k)$. Thus, conditional on $\mathbf{x}$, y has a normal distribution with mean linear in $x_1, \ldots, x_k$ and a constant variance. For a single independent variable x, this situation is shown in Figure 4.1.

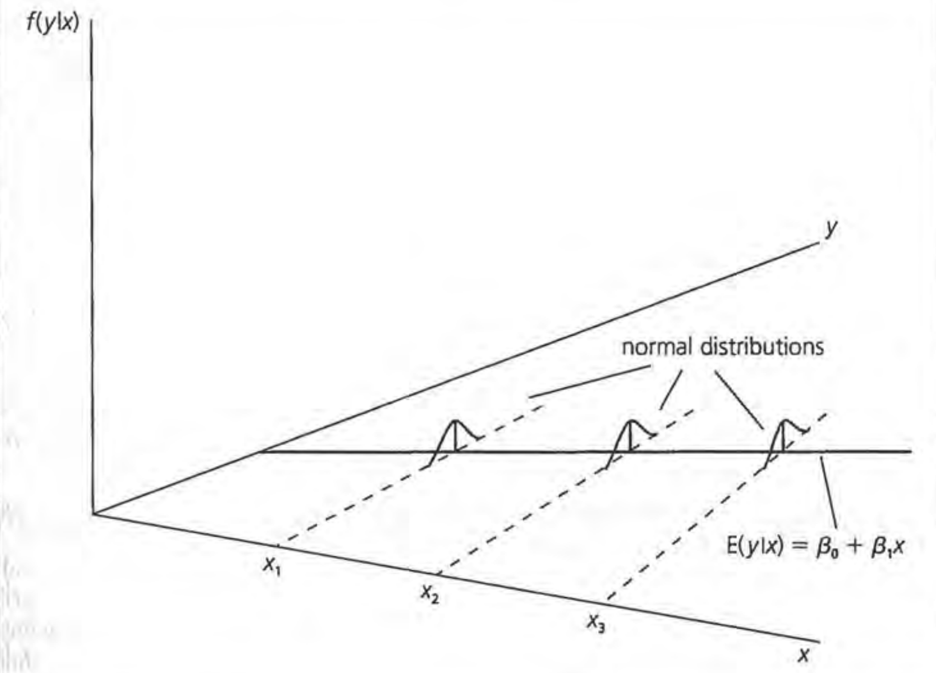
The argument justifying the normal distribution for the errors usually runs something like this: Because u is the sum of many different unobserved factors affecting y, we can invoke the central limit theorem (see Appendix C) to conclude that u has an approximate normal distribution. This argument has some merit, but it is not without weaknesses. First, the factors in u can have very different distributions in the population (for example, ability and quality of schooling in the error in a wage equation). Although the central limit theorem (CLT) can still hold in such cases, the normal approximation can be poor depending on how many factors appear in u and how different are their distributions.

A more serious problem with the CLT argument is that it assumes that all unobserved factors affect y in a separate, additive fashion. Nothing guarantees that this is so. If u is a complicated function of the unobserved factors, then the CLT argument does not really apply.

In any application, whether normality of u can be assumed is really an empirical matter. For example, there is no theorem that says *wage* conditional on *educ*, *exper*, and *tenure* is normally distributed. If anything, simple reasoning suggests that the opposite is true: since *wage* can never be less than zero, it cannot, strictly speaking, have a normal distribution. Further, because there are minimum wage laws, some fraction of the

FIGURE 4.1

The homoskedastic normal distribution with a single explanatory variable.



population earns exactly the minimum wage, which also violates the normality assumption. Nevertheless, as a practical matter, we can ask whether the conditional wage distribution is "close" to being normal. Past empirical evidence suggests that normality is *not* a good assumption for wages.

Often, using a transformation, especially taking the log, yields a distribution that is closer to normal. For example, something like log(*price*) tends to have a distribution that looks more normal than the distribution of *price*. Again, this is an empirical issue. We will discuss the consequences of nonnormality for statistical inference in Chapter 5.

There are some examples where MLR.6 is clearly false. Whenever y takes on just a few values it cannot have anything close to a normal distribution. The dependent variable in Example 3.5 provides a good example. The variable *narr86*, the number of times a young man was arrested in 1986, takes on a small range of integer values and is zero for most men. Thus, *narr86* is far from being normally distributed. What can be done in these cases? As we will see in Chapter 5—and this is important—nonnormality of the errors is not a serious problem with large sample sizes. For now, we just make the normality assumption.

Normality of the error term translates into normal sampling distributions of the OLS estimators:

Theorem 4.1 (Normal Sampling Distributions)

Under the CLM assumptions MLR.1 through MLR.6, conditional on the sample values of the independent variables,

$$\hat{\beta}_j \sim \text{Normal}[\beta_j, \text{Var}(\hat{\beta}_j)], \tag{4.1}$$

where $\text{Var}(\hat{\beta}_j)$ was given in Chapter 3 [equation (3.51)]. Therefore,

$$(\hat{\beta}_j - \beta_j)/\text{sd}(\hat{\beta}_j) \sim \text{Normal}(0,1).$$

The proof of (4.1) is not that difficult, given the properties of normally distributed random variables in Appendix B. Each $\hat{\beta}_j$ can be written as $\hat{\beta}_j = \beta_j + \sum_{i=1}^{n} w_{ij}u_i$, where $w_{ij} = \hat{r}_{ij}/\text{SSR}_j$, $\hat{r}_{ij}$ is the i^{th} residual from the regression of the x_j on all the other independent variables, and SSR_j is the sum of squared residuals from this regression [see equation (3.62)]. Since the w_{ij} depend only on the independent variables, they can be treated as nonrandom. Thus, $\hat{\beta}_j$ is just a linear combination of the errors in the sample, $\{u_i: i = 1, 2, \ldots, n\}$. Under Assumption MLR.6 (and the random sampling Assumption MLR.2), the errors are independent, identically distributed Normal$(0,\sigma^2)$ random variables. An important fact about independent normal random variables is that a linear combination of such random variables is normally distributed (see Appendix B). This basically completes the proof. In Section 3.3, we showed that $\text{E}(\hat{\beta}_j) = \beta_j$, and we derived $\text{Var}(\hat{\beta}_j)$ in Section 3.4; there is no need to re-derive these facts.

QUESTION 4.1

Suppose that u is independent of the explanatory variables, and it takes on the values -2, -1, 0, 1, and 2 with equal probability of 1/5. Does this violate the Gauss-Markov assumptions? Does this violate the CLM assumptions?

The second part of this theorem follows immediately from the fact that when we standardize a normal random variable by subtracting off its mean and dividing by its standard deviation, we end up with a standard normal random variable.

The conclusions of Theorem 4.1 can be strengthened. In addition to (4.1), any linear combination of the $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ is also normally distributed, and any subset of the $\hat{\beta}_j$ has a *joint* normal distribution. These facts underlie the testing results in the remainder of this chapter. In Chapter 5, we will show that the normality of the OLS estimators is still *approximately* true in large samples even without normality of the errors.

4.2 Testing Hypotheses about a Single Population Parameter: The *t* Test

This section covers the very important topic of testing hypotheses about any single parameter in the population regression function. The population model can be written as

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + u, \tag{4.2}$$

and we assume that it satisfies the CLM assumptions. We know that OLS produces unbiased estimators of the β_j. In this section, we study how to test hypotheses about a particular β_j. For a full understanding of hypothesis testing, one must remember that the β_j are unknown features of the population, and we will never know them with certainty. Nevertheless, we can *hypothesize* about the value of β_j and then use statistical inference to test our hypothesis.

In order to construct hypotheses tests, we need the following result:

Theorem 4.2 (*t* Distribution for the Standardized Estimators)

Under the CLM assumptions MLR.1 through MLR.6,

$$(\hat{\beta}_j - \beta_j)/\text{se}(\hat{\beta}_j) \sim t_{n-k-1}, \tag{4.3}$$

where $k + 1$ is the number of unknown parameters in the population model $y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + u$ (k slope parameters and the intercept β_0).

This result differs from Theorem 4.1 in some notable respects. Theorem 4.1 showed that, under the CLM assumptions, $(\hat{\beta}_j - \beta_j)/\text{sd}(\hat{\beta}_j) \sim \text{Normal}(0,1)$. The t distribution in (4.3) comes from the fact that the constant σ in $\text{sd}(\hat{\beta}_j)$ has been replaced with the random variable $\hat{\sigma}$. The proof that this leads to a t distribution with $n - k - 1$ degrees of freedom is not especially insightful. Essentially, the proof shows that (4.3) can be written as the ratio of the standard normal random variable $(\hat{\beta}_j - \beta_j)/\text{sd}(\hat{\beta}_j)$ over the square root of $\hat{\sigma}^2/\sigma^2$. These random variables can be shown to be independent, and $(n - k - 1)\hat{\sigma}^2/\sigma^2 \sim \chi^2_{n-k-1}$. The result then follows from the definition of a t random variable (see Section B.5).

Theorem 4.2 is important in that it allows us to test hypotheses involving the β_j. In most applications, our primary interest lies in testing the **null hypothesis**

$$\text{H}_0\text{: } \beta_j = 0, \tag{4.4}$$

where j corresponds to any of the k independent variables. It is important to understand what (4.4) means and to be able to describe this hypothesis in simple language for a particular application. Since β_j measures the partial effect of x_j on (the expected value of) y, after controlling for all other independent variables, (4.4) means that, once $x_1, x_2, \dots, x_{j-1}, x_{j+1}, \dots, x_k$ have been accounted for, x_j has *no effect* on the expected value of y. We cannot state the null hypothesis as "x_j does have a partial effect on y" because this is true for any value of β_j other than zero. Classical testing is suited for testing *simple hypotheses* like (4.4).

As an example, consider the wage equation

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + u.$$

The null hypothesis H_0: $\beta_2 = 0$ means that, once education and tenure have been accounted for, the number of years in the workforce (*exper*) has no effect on hourly wage. This is an economically interesting hypothesis. If it is true, it implies that a person's work history prior to the current employment does not affect wage. If $\beta_2 > 0$, then prior work experience contributes to productivity, and hence to wage.

You probably remember from your statistics course the rudiments of hypothesis testing for the mean from a normal population. (This is reviewed in Appendix C.) The mechanics of testing (4.4) in the multiple regression context are very similar. The hard part is obtaining the coefficient estimates, the standard errors, and the critical values, but most of this work is done automatically by econometrics software. Our job is to learn how regression output can be used to test hypotheses of interest.

The statistic we use to test (4.4) (against any alternative) is called "the" **t statistic** or "the" **t ratio** of $\hat{\beta}_j$ and is defined as

$$t_{\hat{\beta}_j} \equiv \hat{\beta}_j/\text{se}(\hat{\beta}_j). \tag{4.5}$$

We have put "the" in quotation marks because, as we will see shortly, a more general form of the t statistic is needed for testing other hypotheses about β_j. For now, it is important to know that (4.5) is suitable only for testing (4.4). For particular applications, it is helpful to index t statistics using the name of the independent variable; for example, t_{educ} would be the t statistic for $\hat{\beta}_{educ}$.

The t statistic for $\hat{\beta}_j$ is simple to compute given $\hat{\beta}_j$ and its standard error. In fact, most regression packages do the division for you and report the t statistic along with each coefficient and its standard error.

Before discussing how to use (4.5) formally to test H_0: $\beta_j = 0$, it is useful to see why $t_{\hat{\beta}_j}$ has features that make it reasonable as a test statistic to detect $\beta_j \neq 0$. First, since se($\hat{\beta}_j$) is always positive, $t_{\hat{\beta}_j}$ has the same sign as $\hat{\beta}_j$: if $\hat{\beta}_j$ is positive, then so is $t_{\hat{\beta}_j}$, and if $\hat{\beta}_j$ is negative, so is $t_{\hat{\beta}_j}$. Second, for a given value of se($\hat{\beta}_j$), a larger value of $\hat{\beta}_j$ leads to larger values of $t_{\hat{\beta}_j}$. If $\hat{\beta}_j$ becomes more negative, so does $t_{\hat{\beta}_j}$.

Since we are testing H_0: $\beta_j = 0$, it is only natural to look at our unbiased estimator of β_j, $\hat{\beta}_j$, for guidance. In any interesting application, the point estimate $\hat{\beta}_j$ will *never* exactly be zero, whether or not H_0 is true. The question is: How far is $\hat{\beta}_j$ from zero? A sample value of $\hat{\beta}_j$ very far from zero provides evidence against H_0: $\beta_j = 0$. However, we must recognize that there is a sampling error in our estimate $\hat{\beta}_j$, so the size of $\hat{\beta}_j$ must be weighed against its sampling error. Since the standard error of $\hat{\beta}_j$ is an estimate of the standard deviation of $\hat{\beta}_j$, $t_{\hat{\beta}_j}$ measures how many estimated standard deviations $\hat{\beta}_j$ is away from zero. This is precisely what we do in testing whether the mean of a population is zero, using the standard t statistic from introductory statistics. Values of $t_{\hat{\beta}_j}$ sufficiently far from zero will result in a rejection of H_0. The precise rejection rule depends on the alternative hypothesis and the chosen significance level of the test.

Determining a rule for rejecting (4.4) at a given significance level—that is, the probability of rejecting H_0 when it is true—requires knowing the sampling distribution of $t_{\hat{\beta}_j}$ when H_0 is true. From Theorem 4.2, we know this to be t_{n-k-1}. This is the key theoretical result needed for testing (4.4).

Before proceeding, it is important to remember that we are testing hypotheses about the *population* parameters. We are *not* testing hypotheses about the estimates from a particular sample. Thus, it never makes sense to state a null hypothesis as "H_0: $\hat{\beta}_1 = 0$" or, even worse, as "H_0: .237 = 0" when the estimate of a parameter is .237 in the sample. We are testing whether the unknown population value, β_1, is zero.

Some treatments of regression analysis define the t statistic as the *absolute value* of (4.5), so that the t statistic is always positive. This practice has the drawback of making testing against one-sided alternatives clumsy. Throughout this text, the t statistic always has the same sign as the corresponding OLS coefficient estimate.

Testing against One-Sided Alternatives

In order to determine a rule for rejecting H_0, we need to decide on the relevant **alternative hypothesis**. First, consider a **one-sided alternative** of the form

$$H_1\colon \beta_j > 0. \tag{4.6}$$

This means that we do not care about alternatives to H_0 of the form H_1: $\beta_j < 0$; for some reason, perhaps on the basis of introspection or economic theory, we are ruling out population values of β_j less than zero. (Another way to think about this is that the null hypothesis is actually H_0: $\beta_j \leq 0$; in either case, the statistic $t_{\hat{\beta}_j}$ is used as the test statistic.)

How should we choose a rejection rule? We must first decide on a **significance level** or the probability of rejecting H_0 when it is in fact true. For concreteness, suppose we have decided on a 5% significance level, as this is the most popular choice. Thus, we are willing to mistakenly reject H_0 when it is true 5% of the time. Now, while $t_{\hat{\beta}_j}$ has a t distribution under H_0—so that it has zero mean—under the alternative $\beta_j > 0$, the expected value of $t_{\hat{\beta}_j}$ is positive. Thus, we are looking for a "sufficiently large" positive value of $t_{\hat{\beta}_j}$ in order to reject H_0: $\beta_j = 0$ in favor of H_1: $\beta_j > 0$. Negative values of $t_{\hat{\beta}_j}$ provide no evidence in favor of H_1.

The definition of "sufficiently large," with a 5% significance level, is the 95$^{\text{th}}$ percentile in a t distribution with $n - k - 1$ degrees of freedom; denote this by c. In other words, the **rejection rule** is that H_0 is rejected in favor of H_1 at the 5% significance level if

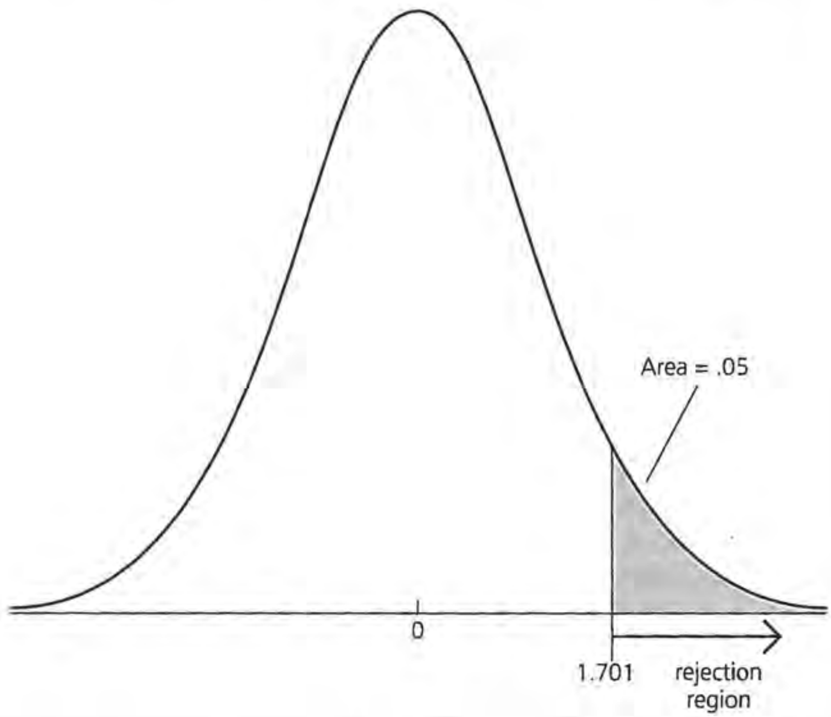
$$t_{\hat{\beta}_j} > c. \tag{4.7}$$

By our choice of the **critical value** c, rejection of H_0 will occur for 5% of all random samples when H_0 is true.

The rejection rule in (4.7) is an example of a **one-tailed test**. In order to obtain c, we only need the significance level and the degrees of freedom. For example, for a 5% level test and with $n - k - 1 = 28$ degrees of freedom, the critical value is $c = 1.701$. If $t_{\hat{\beta}_j} <$ 1.701, then we fail to reject H_0 in favor of (4.6) at the 5% level. Note that a negative value for $t_{\hat{\beta}_j}$, no matter how large in absolute value, leads to a failure in rejecting H_0 in favor of (4.6). (See Figure 4.2.)

FIGURE 4.2

5% rejection rule for the alternative $H_1: \beta_j > 0$ with 28 *df.*



The same procedure can be used with other significance levels. For a 10% level test and if $df = 21$, the critical value is $c = 1.323$. For a 1% significance level and if $df = 21$, $c = 2.518$. All of these critical values are obtained directly from Table G.2. You should note a pattern in the critical values: as the significance level falls, the critical value increases, so that we require a larger and larger value of $t_{\hat{\beta}_j}$ in order to reject H_0. Thus, if H_0 is rejected at, say, the 5% level, then it is automatically rejected at the 10% level as well. It makes no sense to reject the null hypothesis at, say, the 5% level and then to redo the test to determine the outcome at the 10% level.

As the degrees of freedom in the t distribution gets large, the t distribution approaches the standard normal distribution. For example, when $n - k - 1 = 120$, the 5% critical value for the one-sided alternative (4.7) is 1.658, compared with the standard normal value of 1.645. These are close enough for practical purposes; for degrees of freedom greater than 120, one can use the standard normal critical values.

EXAMPLE 4.1

(Hourly Wage Equation)

Using the data in WAGE1.RAW gives the estimated equation

$$\widehat{\log(wage)} = \underset{(.104)}{.284} + \underset{(.007)}{.092}\, educ + \underset{(.0017)}{.0041}\, exper + \underset{(.003)}{.022}\, tenure$$
$$n = 526, R^2 = .316,$$

where standard errors appear in parentheses below the estimated coefficients. We will follow this convention throughout the text. This equation can be used to test whether the return to *exper*, controlling for *educ* and *tenure*, is zero in the population, against the alternative that it is positive. Write this as H_0: $\beta_{exper} = 0$ versus H_1: $\beta_{exper} > 0$. (In applications, indexing a parameter by its associated variable name is a nice way to label parameters, since the numerical indices that we use in the general model are arbitrary and can cause confusion.) Remember that β_{exper} denotes the unknown population parameter. It is nonsense to write "H_0: .0041 = 0" or "H_0: $\hat{\beta}_{exper} = 0$."

Since we have 522 degrees of freedom, we can use the standard normal critical values. The 5% critical value is 1.645, and the 1% critical value is 2.326. The *t* statistic for $\hat{\beta}_{exper}$ is

$$t_{exper} = .0041/.0017 \approx 2.41,$$

and so $\hat{\beta}_{exper}$, or *exper*, is statistically significant even at the 1% level. We also say that "$\hat{\beta}_{exper}$ is statistically greater than zero at the 1% significance level."

The estimated return for another year of experience, holding tenure and education fixed, is not especially large. For example, adding three more years increases log(*wage*) by 3(.0041) = .0123, so wage is only about 1.2% higher. Nevertheless, we have persuasively shown that the partial effect of experience *is* positive in the population.

The one-sided alternative that the parameter is less than zero,

$$H_1: \beta_j < 0, \tag{4.8}$$

also arises in applications. The rejection rule for alternative (4.8) is just the mirror image of the previous case. Now, the critical value comes from the left tail of the *t* distribution. In practice, it is easiest to think of the rejection rule as

$$t_{\hat{\beta}_j} < -c, \tag{4.9}$$

where *c* is the critical value for the alternative H_1: $\beta_j > 0$. For simplicity, we always assume *c* is positive, since this is how critical values are reported in *t* tables, and so the critical value $-c$ is a negative number.

QUESTION 4.2

Let community loan approval rates be determined by

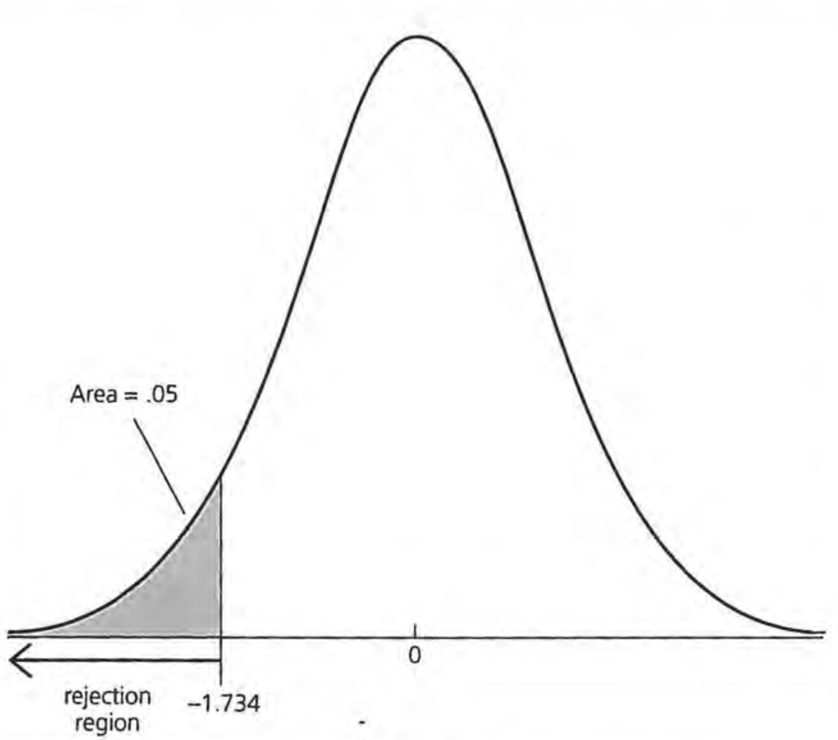
$$apprate = \beta_0 + \beta_1 percmin + \beta_2 avginc + \beta_3 avgwlth + \beta_4 avgdebt + u,$$

where *percmin* is the percent minority in the community, *avginc* is average income, *avgwlth* is average wealth, and *avgdebt* is some measure of average debt obligations. How do you state the null hypothesis that there is *no* difference in loan rates across neighborhoods due to racial and ethnic composition, when average income, average wealth, and average debt have been controlled for? How do you state the alternative that there is discrimination against minorities in loan approval rates?

For example, if the significance level is 5% and the degrees of freedom is 18, then $c = 1.734$, and so H_0: $\beta_j = 0$ is rejected in favor of H_1: $\beta_j < 0$ at the 5% level if $t_{\hat{\beta}_j} < -1.734$. It is important to remember that, to reject H_0 against the negative alternative (4.8), we must get a negative t statistic. A positive t ratio, no matter how large, provides no evidence in favor of (4.8). The rejection rule is illustrated in Figure 4.3.

FIGURE 4.3

5% rejection rule for the alternative H_1: $\beta_j < 0$ with 18 *df*.



EXAMPLE 4.2

(Student Performance and School Size)

There is much interest in the effect of school size on student performance. (See, for example, *The New York Times Magazine*, 5/28/95.) One claim is that, everything else being equal, students at smaller schools fare better than those at larger schools. This hypothesis is assumed to be true even after accounting for differences in class sizes across schools.

The file MEAP93.RAW contains data on 408 high schools in Michigan for the year 1993. We can use these data to test the null hypothesis that school size has no effect on standardized test scores against the alternative that size has a negative effect. Performance is measured by the percentage of students receiving a passing score on the Michigan Educational Assessment Program (MEAP) standardized tenth-grade math test (*math*10). School size is measured by student enrollment (*enroll*). The null hypothesis is H_0: $\beta_{enroll} = 0$, and the alternative is H_1: $\beta_{enroll} < 0$. For now, we will control for two other factors, average annual teacher compensation (*totcomp*) and the number of staff per one thousand students (*staff*). Teacher compensation is a measure of teacher quality, and staff size is a rough measure of how much attention students receive.

The estimated equation, with standard errors in parentheses, is

$$\begin{aligned} \widehat{math10} &= \underset{(6.113)}{2.274} + \underset{(.00010)}{.00046}\, totcomp + \underset{(.040)}{.048}\, staff - \underset{(.00022)}{.00020}\, enroll \\ n &= 408,\ R^2 = .0541. \end{aligned}$$

The coefficient on *enroll*, $-.00020$, is in accordance with the conjecture that larger schools hamper performance: higher enrollment leads to a lower percentage of students with a passing tenth-grade math score. (The coefficients on *totcomp* and *staff* also have the signs we expect.) The fact that *enroll* has an estimated coefficient different from zero could just be due to sampling error; to be convinced of an effect, we need to conduct a *t* test.

Since $n - k - 1 = 408 - 4 = 404$, we use the standard normal critical value. At the 5% level, the critical value is -1.65; the *t* statistic on *enroll* must be *less* than -1.65 to reject H_0 at the 5% level.

The *t* statistic on *enroll* is $-.00020/.00022 \approx -.91$, which is larger than -1.65: we *fail* to reject H_0 in favor of H_1 at the 5% level. In fact, the 15% critical value is -1.04, and since $-.91 > -1.04$, we fail to reject H_0 even at the 15% level. We conclude that *enroll* is not statistically significant at the 15% level.

The variable *totcomp* is statistically significant even at the 1% significance level because its *t* statistic is 4.6. On the other hand, the *t* statistic for *staff* is 1.2, and so we cannot reject H_0: $\beta_{staff} = 0$ against H_1: $\beta_{staff} > 0$ even at the 10% significance level. (The critical value is $c = 1.28$ from the standard normal distribution.)

To illustrate how changing functional form can affect our conclusions, we also estimate the model with all independent variables in logarithmic form. This allows, for example, the school size effect to diminish as school size increases. The estimated equation is

$$\begin{aligned} \widehat{math10} &= \underset{(48.70)}{-207.66} + \underset{(4.06)}{21.16}\, \log(totcomp) + \underset{(4.19)}{3.98}\, \log(staff) - \underset{(0.69)}{1.29}\, \log(enroll) \\ n &= 408,\ R^2 = .0654. \end{aligned}$$

The t statistic on log(*enroll*) is about -1.87; since this is below the 5% critical value -1.65, we reject H_0: $\beta_{\log(enroll)} = 0$ in favor of H_1: $\beta_{\log(enroll)} < 0$ at the 5% level.

In Chapter 2, we encountered a model where the dependent variable appeared in its original form (called *level* form), while the independent variable appeared in log form (called *level-log* model). The interpretation of the parameters is the same in the multiple regression context, except, of course, that we can give the parameters a ceteris paribus interpretation. Holding *totcomp* and *staff* fixed, we have $\Delta\widehat{math10} = -1.29[\Delta\log(enroll)]$, so that

$$\Delta\widehat{math10} \approx -(1.29/100)(\%\Delta enroll) \approx -.013(\%\Delta enroll).$$

Once again, we have used the fact that the change in log(*enroll*), when multiplied by 100, is approximately the percentage change in *enroll*. Thus, if enrollment is 10% higher at a school, $\widehat{math10}$ is predicted to be $.013(10) = 0.13$ percentage points lower (*math*10 is measured as a percent).

Which model do we prefer: the one using the level of *enroll* or the one using log(*enroll*)? In the level-level model, enrollment does not have a statistically significant effect, but in the level-log model it does. This translates into a higher R-squared for the level-log model, which means we explain more of the variation in *math*10 by using *enroll* in logarithmic form (6.5% to 5.4%). The level-log model is preferred, as it more closely captures the relationship between *math*10 and *enroll*. We will say more about using R-squared to choose functional form in Chapter 6.

Two-Sided Alternatives

In applications, it is common to test the null hypothesis H_0: $\beta_j = 0$ against a **two-sided alternative**; that is,

$$H_1: \beta_j \neq 0. \tag{4.10}$$

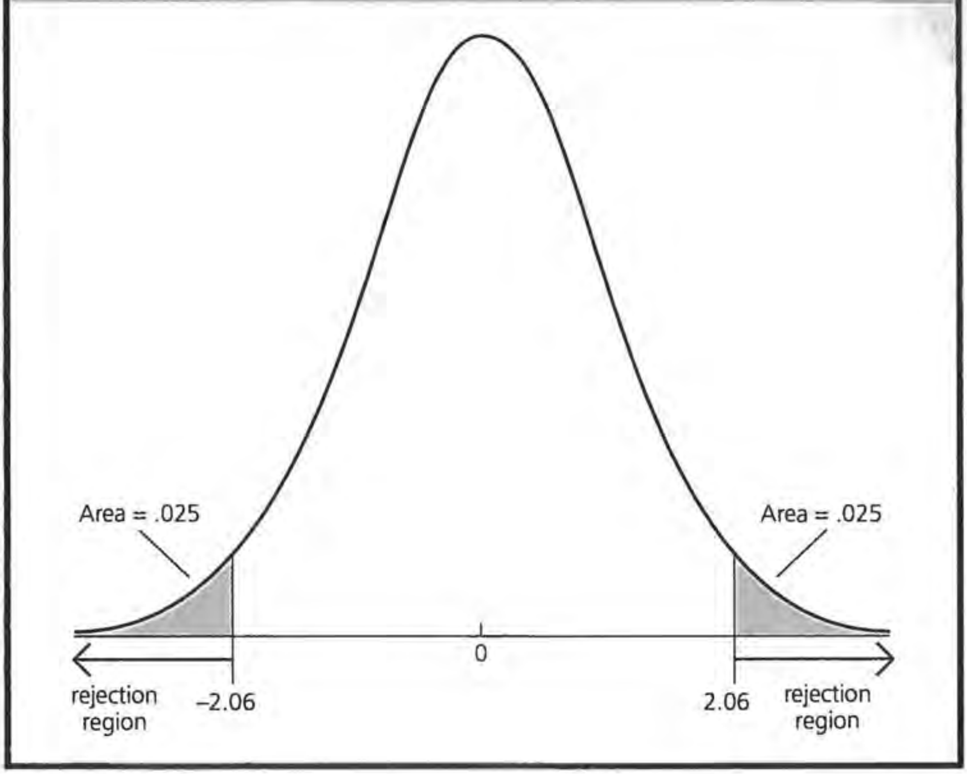
Under this alternative, x_j has a ceteris paribus effect on y without specifying whether the effect is positive or negative. This is the relevant alternative when the sign of β_j is not well determined by theory (or common sense). Even when we know whether β_j is positive or negative under the alternative, a two-sided test is often prudent. At a minimum, using a two-sided alternative prevents us from looking at the estimated equation and then basing the alternative on whether $\hat{\beta}_j$ is positive or negative. Using the regression estimates to help us formulate the null or alternative hypotheses is not allowed because classical statistical inference presumes that we state the null and alternative about the population before looking at the data. For example, we should not first estimate the equation relating math performance to enrollment, note that the estimated effect is negative, and then decide the relevant alternative is H_1: $\beta_{enroll} < 0$.

When the alternative is two-sided, we are interested in the *absolute value* of the t statistic. The rejection rule for H_0: $\beta_j = 0$ against (4.10) is

$$|t_{\hat{\beta}_j}| > c, \tag{4.11}$$

FIGURE 4.4

5% rejection rule for the alternative H_1: $\beta_j \neq 0$ with 25 *df*.



where $|\cdot|$ denotes absolute value and c is an appropriately chosen critical value. To find c, we again specify a significance level, say 5%. For a **two-tailed test,** c is chosen to make the area in each tail of the t distribution equal 2.5%. In other words, c is the 97.5$^{\text{th}}$ percentile in the t distribution with $n - k - 1$ degrees of freedom. When $n - k - 1 = 25$, the 5% critical value for a two-sided test is $c = 2.060$. Figure 4.4 provides an illustration of this distribution.

When a specific alternative is not stated, it is usually considered to be two-sided. In the remainder of this text, the default will be a two-sided alternative, and 5% will be the default significance level. When carrying out empirical econometric analysis, it is always a good idea to be explicit about the alternative and the significance level. If H_0 is rejected in favor of (4.10) at the 5% level, we usually say that "x_j is **statistically significant,** or statistically different from zero, at the 5% level." If H_0 is not rejected, we say that "x_j is **statistically insignificant** at the 5% level."

EXAMPLE 4.3

(Determinants of College GPA)

We use GPA1.RAW to estimate a model explaining college GPA (*colGPA*), with the average number of lectures missed per week (*skipped*) as an additional explanatory variable. The estimated model is

$$\begin{aligned}\widehat{colGPA} = \underset{(.33)}{1.39} + \underset{(.094)}{.412}\, hsGPA + \underset{(.011)}{.015}\, ACT - \underset{(.026)}{.083}\ skipped\\ n = 141, R^2 = .234.\end{aligned}$$

We can easily compute *t* statistics to see which variables are statistically significant, using a two-sided alternative in each case. The 5% critical value is about 1.96, since the degrees of freedom (141 − 4 = 137) is large enough to use the standard normal approximation. The 1% critical value is about 2.58.

The *t* statistic on *hsGPA* is 4.38, which is significant at very small significance levels. Thus, we say that "*hsGPA* is statistically significant at any *conventional* significance level." The *t* statistic on *ACT* is 1.36, which is not statistically significant at the 10% level against a two-sided alternative. The coefficient on *ACT* is also practically small: a 10-point increase in *ACT*, which is large, is predicted to increase *colGPA* by only .15 point. Thus, the variable *ACT* is practically, as well as statistically, insignificant.

The coefficient on *skipped* has a *t* statistic of −.083/.026 = −3.19, so *skipped* is statistically significant at the 1% significance level (3.19 > 2.58). This coefficient means that another lecture missed per week lowers predicted *colGPA* by about .083. Thus, holding *hsGPA* and *ACT* fixed, the predicted difference in *colGPA* between a student who misses no lectures per week and a student who misses five lectures per week is about .42. Remember that this says nothing about specific students, but pertains to average students across the population.

In this example, for each variable in the model, we could argue that a one-sided alternative is appropriate. The variables *hsGPA* and *skipped* are very significant using a two-tailed test and have the signs that we expect, so there is no reason to do a one-tailed test. On the other hand, against a one-sided alternative ($\beta_3 > 0$), *ACT* is significant at the 10% level but not at the 5% level. This does not change the fact that the coefficient on *ACT* is pretty small.

Testing Other Hypotheses about β_j

Although H_0: $\beta_j = 0$ is the most common hypothesis, we sometimes want to test whether β_j is equal to some other given constant. Two common examples are $\beta_j = 1$ and $\beta_j = -1$. Generally, if the null is stated as

$$H_0\colon \beta_j = a_j, \tag{4.12}$$

where a_j is our hypothesized value of β_j, then the appropriate t statistic is

$$t = (\hat{\beta}_j - a_j)/\mathrm{se}(\hat{\beta}_j).$$

As before, t measures how many estimated standard deviations $\hat{\beta}_j$ is away from the hypothesized value of β_j. The general t statistic is usefully written as

$$t = \frac{(estimate - hypothesized\ value)}{standard\ error}. \quad (4.13)$$

Under (4.12), this t statistic is distributed as t_{n-k-1} from Theorem 4.2. The usual t statistic is obtained when $a_j = 0$.

We can use the general t statistic to test against one-sided or two-sided alternatives. For example, if the null and alternative hypotheses are H_0: $\beta_j = 1$ and H_1: $\beta_j > 1$, then we find the critical value for a one-sided alternative *exactly* as before: the difference is in how we compute the t statistic, not in how we obtain the appropriate c. We reject H_0 in favor of H_1 if $t > c$. In this case, we would say that "$\hat{\beta}_j$ is statistically greater than one" at the appropriate significance level.

EXAMPLE 4.4

(Campus Crime and Enrollment)

Consider a simple model relating the annual number of crimes on college campuses (*crime*) to student enrollment (*enroll*):

$$\log(crime) = \beta_0 + \beta_1 \log(enroll) + u.$$

This is a constant elasticity model, where β_1 is the elasticity of crime with respect to enrollment. It is not much use to test H_0: $\beta_1 = 0$, as we expect the total number of crimes to increase as the size of the campus increases. A more interesting hypothesis to test would be that the elasticity of crime with respect to enrollment is one: H_0: $\beta_1 = 1$. This means that a 1% increase in enrollment leads to, on average, a 1% increase in crime. A noteworthy alternative is H_1: $\beta_1 > 1$, which implies that a 1% increase in enrollment increases campus crime by *more* than 1%. If $\beta_1 > 1$, then, in a relative sense—not just an absolute sense—crime is more of a problem on larger campuses. One way to see this is to take the exponential of the equation:

$$crime = \exp(\beta_0) enroll^{\beta_1} \exp(u).$$

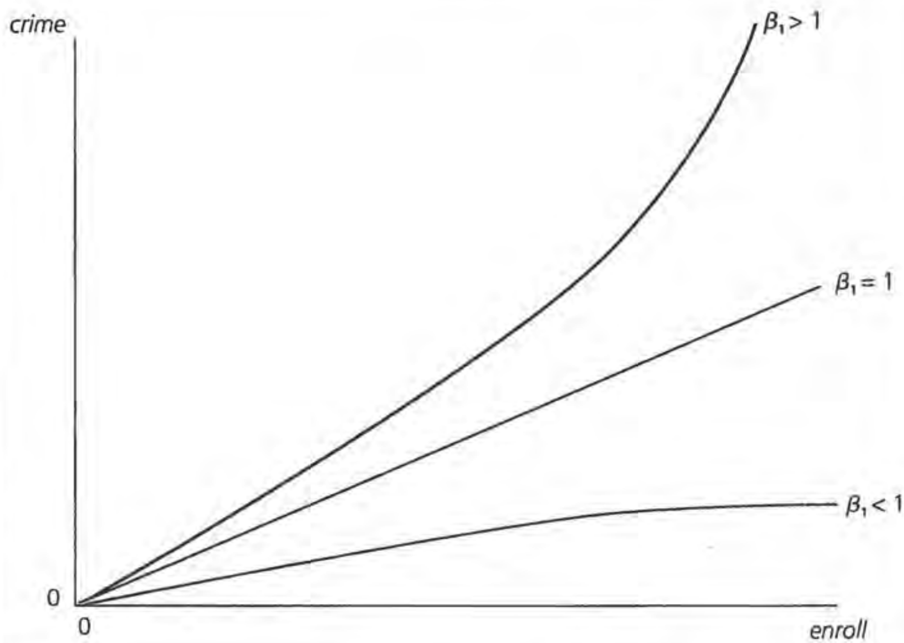
(See Appendix A for properties of the natural logarithm and exponential functions.) For $\beta_0 = 0$ and $u = 0$, this equation is graphed in Figure 4.5 for $\beta_1 < 1$, $\beta_1 = 1$, and $\beta_1 > 1$.

We test $\beta_1 = 1$ against $\beta_1 > 1$ using data on 97 colleges and universities in the United States for the year 1992, contained in the data file CAMPUS.RAW. The data come from the FBI's *Uniform Crime Reports*, and the average number of campus crimes in the sample is about 394, while the average enrollment is about 16,076. The estimated equation (with estimates and standard errors rounded to two decimal places) is

$$\begin{aligned} \widehat{\log(crime)} &= \underset{(1.03)}{-6.63} + \underset{(0.11)}{1.27} \log(enroll) \\ n &= 97, R^2 = .585. \end{aligned} \quad (4.14)$$

FIGURE 4.5

Graph of $crime = enroll^{\beta_1}$ for $\beta_1 < 1$, $\beta_1 = 1$, and $\beta_1 > 1$.



The estimated elasticity of *crime* with respect to *enroll*, 1.27, is in the direction of the alternative $\beta_1 > 1$. But is there enough evidence to conclude that $\beta_1 > 1$? We need to be careful in testing this hypothesis, especially because the statistical output of standard regression packages is much more complex than the simplified output reported in equation (4.14). Our first instinct might be to construct "the" t statistic by taking the coefficient on log(*enroll*) and dividing it by its standard error, which is the t statistic reported by a regression package. But this is the *wrong* statistic for testing H_0: $\beta_1 = 1$. The correct t statistic is obtained from (4.13): we subtract the hypothesized value, unity, from the estimate and divide the result by the standard error of $\hat{\beta}_1$: $t = (1.27 - 1)/.11 = .27/.11 \approx 2.45$. The one-sided 5% critical value for a t distribution with $97 - 2 = 95$ df is about 1.66 (using $df = 120$), so we clearly reject $\beta_1 = 1$ in favor of $\beta_1 > 1$ at the 5% level. In fact, the 1% critical value is about 2.37, and so we reject the null in favor of the alternative at even the 1% level.

We should keep in mind that this analysis holds no other factors constant, so the elasticity of 1.27 is not necessarily a good estimate of ceteris paribus effect. It could be that larger enrollments are correlated with other factors that cause higher crime: larger schools might be located in higher crime areas. We could control for this by collecting data on crime rates in the local city.

For a two-sided alternative, for example H_0: $\beta_j = -1$, H_1: $\beta_j \neq -1$, we still compute the t statistic as in (4.13): $t = (\hat{\beta}_j + 1)/\mathrm{se}(\hat{\beta}_j)$ (notice how subtracting -1 means adding 1). The rejection rule is the usual one for a two-sided test: reject H_0 if $|t| > c$, where c is a two-tailed critical value. If H_0 is rejected, we say that "$\hat{\beta}_j$ is statistically different from negative one" at the appropriate significance level.

EXAMPLE 4.5

(Housing Prices and Air Pollution)

For a sample of 506 communities in the Boston area, we estimate a model relating median housing price (*price*) in the community to various community characteristics: *nox* is the amount of nitrogen oxide in the air, in parts per million; *dist* is a weighted distance of the community from five employment centers, in miles; *rooms* is the average number of rooms in houses in the community; and *stratio* is the average student-teacher ratio of schools in the community. The population model is

$$\log(price) = \beta_0 + \beta_1 \log(nox) + \beta_2 \log(dist) + \beta_3 rooms + \beta_4 stratio + u.$$

Thus, β_1 is the elasticity of *price* with respect to *nox*. We wish to test H_0: $\beta_1 = -1$ against the alternative H_1: $\beta_1 \neq -1$. The t statistic for doing this test is $t = (\hat{\beta}_1 + 1)/\mathrm{se}(\hat{\beta}_1)$.

Using the data in HPRICE2.RAW, the estimated model is

$$\begin{aligned} \widehat{\log(price)} = {} & \underset{(0.32)}{11.08} - \underset{(.117)}{.954} \log(nox) - \underset{(.043)}{.134} \log(dist) + \underset{(.019)}{.255}\, rooms - \underset{(.006)}{.052}\, stratio \\ & n = 506,\ R^2 = .581. \end{aligned}$$

The slope estimates all have the anticipated signs. Each coefficient is statistically different from zero at very small significance levels, including the coefficient on log(*nox*). But we do not want to test that $\beta_1 = 0$. The null hypothesis of interest is H_0: $\beta_1 = -1$, with corresponding t statistic $(-.954 + 1)/.117 = .393$. There is little need to look in the t table for a critical value when the t statistic is this small: the estimated elasticity is not statistically different from -1 even at very large significance levels. Controlling for the factors we have included, there is little evidence that the elasticity is different from -1.

Computing *p*-Values for *t* Tests

So far, we have talked about how to test hypotheses using a classical approach: after stating the alternative hypothesis, we choose a significance level, which then determines a critical value. Once the critical value has been identified, the value of the t statistic is compared with the critical value, and the null is either rejected or not rejected at the given significance level.

Even after deciding on the appropriate alternative, there is a component of arbitrariness to the classical approach, which results from having to choose a significance level ahead of time. Different researchers prefer different significance levels, depending on the particular application. There is no "correct" significance level.

Committing to a significance level ahead of time can hide useful information about the outcome of a hypothesis test. For example, suppose that we wish to test the null hypothesis that a parameter is zero against a two-sided alternative, and with 40 degrees of freedom we obtain a t statistic equal to 1.85. The null hypothesis is not rejected at the 5% level, since the t statistic is less than the two-tailed critical value of $c = 2.021$. A researcher whose agenda is not to reject the null could simply report this outcome along with the estimate: the null hypothesis is not rejected at the 5% level. Of course, if the t statistic, or the coefficient and its standard error, are reported, then we can also determine that the null hypothesis would be rejected at the 10% level, since the 10% critical value is $c = 1.684$.

Rather than testing at different significance levels, it is more informative to answer the following question: Given the observed value of the t statistic, what is the *smallest* significance level at which the null hypothesis would be rejected? This level is known as the **p-value** for the test (see Appendix C). In the previous example, we know the p-value is greater than .05, since the null is not rejected at the 5% level, and we know that the p-value is less than .10, since the null is rejected at the 10% level. We obtain the actual p-value by computing the probability that a t random variable, with 40 *df*, is larger than 1.85 in absolute value. That is, the p-value is the significance level of the test when we use the value of the test statistic, 1.85 in the above example, as the critical value for the test. This p-value is shown in Figure 4.6.

Because a p-value is a probability, its value is always between zero and one. In order to compute p-values, we either need extremely detailed printed tables of the t distribution—which is not very practical—or a computer program that computes areas under the probability density function of the t distribution. Most modern regression packages have this capability. Some packages compute p-values routinely with each OLS regression, but only for certain hypotheses. If a regression package reports a p-value along with the standard OLS output, it is almost certainly the p-value for testing the null hypothesis H_0: $\beta_j = 0$ against the two-sided alternative. The p-value in this case is

$$P(|T| > |t|), \tag{4.15}$$

where, for clarity, we let T denote a t distributed random variable with $n - k - 1$ degrees of freedom and let t denote the numerical value of the test statistic.

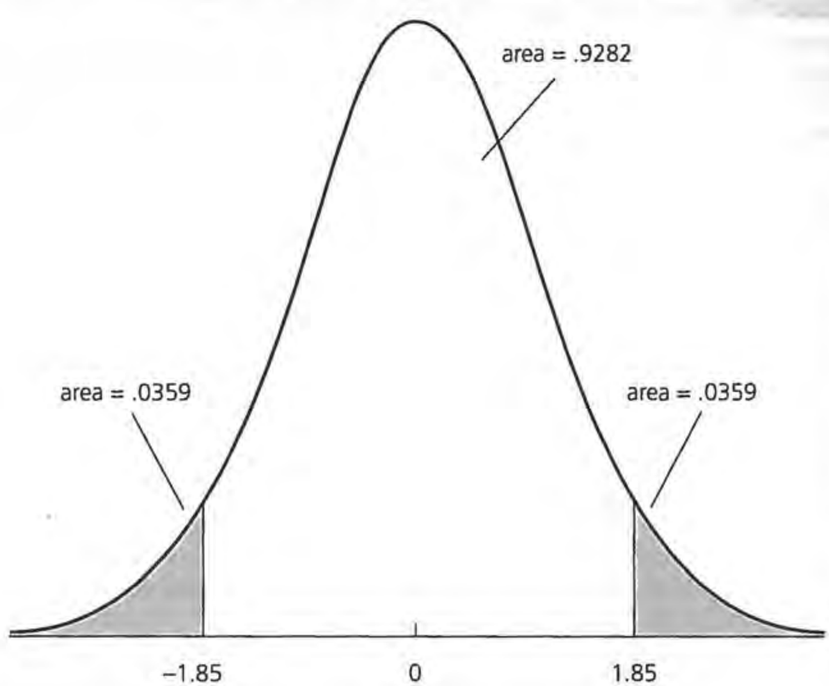
The p-value nicely summarizes the strength or weakness of the empirical evidence against the null hypothesis. Perhaps its most useful interpretation is the following: the p-value is the probability of observing a t statistic as extreme as we did *if the null hypothesis is true*. This means that *small* p-values are evidence *against* the null; large p-values provide little evidence against H_0. For example, if the p-value $= .50$ (reported always as a decimal, not a percent), then we would observe a value of the t statistic as extreme as we did in 50% of all random samples when the null hypothesis is true; this is pretty weak evidence against H_0.

In the example with $df = 40$ and $t = 1.85$, the p-value is computed as

$$p\text{-value} = P(|T| > 1.85) = 2P(T > 1.85) = 2(.0359) = .0718,$$

where $P(T > 1.85)$ is the area to the right of 1.85 in a t distribution with 40 *df*. (This value was computed using the econometrics package Stata; it is not available in Table G.2.) This

FIGURE 4.6
Obtaining the p-value against a two-sided alternative, when $t = 1.85$ and $df = 40$.



means that, if the null hypothesis is true, we would observe an absolute value of the t statistic as large as 1.85 about 7.2 percent of the time. This provides some evidence against the null hypothesis, but we would not reject the null at the 5% significance level.

The previous example illustrates that once the p-value has been computed, a classical test can be carried out at any desired level. If α denotes the significance level of the test (in decimal form), then H_0 is rejected if p-value $< \alpha$; otherwise, H_0 is not rejected at the $100 \cdot \alpha\%$ level.

Computing p-values for one-sided alternatives is also quite simple. Suppose, for example, that we test H_0: $\beta_j = 0$ against H_1: $\beta_j > 0$. If $\hat{\beta}_j < 0$, then computing a p-value is not important: we know that the p-value is greater than .50, which will never cause us to reject H_0 in favor of H_1. If $\hat{\beta}_j > 0$, then $t > 0$ and the p-value is just the probability that a t random variable with the appropriate df exceeds the value t. Some regression packages only compute p-values for two-sided alternatives. But it is simple to obtain the one-sided p-value: just divide the two-sided p-value by 2.

If the alternative is H_1: $\beta_j < 0$, it makes sense to compute a p-value if $\hat{\beta}_j < 0$ (and hence $t < 0$): p-value $= P(T < t) = P(T > |t|)$ because the t distribution is symmetric about zero. Again, this can be obtained as one-half of the p-value for the two-tailed test.

QUESTION 4.3

Suppose you estimate a regression model and obtain $\hat{\beta}_1 = .56$ and p-value $= .086$ for testing H_0: $\beta_1 = 0$ against H_1: $\beta_1 \neq 0$. What is the p-value for testing H_0: $\beta_1 = 0$ against H_1: $\beta_1 > 0$?

Because you will quickly become familiar with the magnitudes of t statistics that lead to statistical significance, especially for large sample sizes, it is not always crucial to report p-values for t statistics. But it does not hurt to report them. Further, when we discuss F testing in Section 4.5, we will see that it is important to compute p-values, because critical values for F tests are not so easily memorized.

A Reminder on the Language of Classical Hypothesis Testing

When H_0 is not rejected, we prefer to use the language "we fail to reject H_0 at the x% level," rather than "H_0 is accepted at the x% level." We can use Example 4.5 to illustrate why the former statement is preferred. In this example, the estimated elasticity of *price* with respect to *nox* is $-.954$, and the t statistic for testing H_0: $\beta_{nox} = -1$ is $t = .393$; therefore, we cannot reject H_0. But there are many other values for β_{nox} (more than we can count) that cannot be rejected. For example, the t statistic for H_0: $\beta_{nox} = -.9$ is $(-.954 + .9)/.117 = -.462$, and so this null is not rejected either. Clearly $\beta_{nox} = -1$ and $\beta_{nox} = -.9$ cannot both be true, so it makes no sense to say that we "accept" either of these hypotheses. All we can say is that the data do not allow us to reject either of these hypotheses at the 5% significance level.

Economic, or Practical, versus Statistical Significance

Because we have emphasized *statistical significance* throughout this section, now is a good time to remember that we should pay attention to the magnitude of the *coefficient* estimates in addition to the size of the t statistics. The statistical significance of a variable x_j is determined entirely by the size of $t_{\hat{\beta}_j}$, whereas the **economic significance** or **practical significance** of a variable is related to the size (and sign) of $\hat{\beta}_j$.

Recall that the t statistic for testing H_0: $\beta_j = 0$ is defined by dividing the estimate by its standard error: $t_{\hat{\beta}_j} = \hat{\beta}_j/\text{se}(\hat{\beta}_j)$. Thus, $t_{\hat{\beta}_j}$ can indicate statistical significance either because $\hat{\beta}_j$ is "large" or because $\text{se}(\hat{\beta}_j)$ is "small." It is important in practice to distinguish between these reasons for statistically significant t statistics. Too much focus on statistical significance can lead to the false conclusion that a variable is "important" for explaining y even though its estimated effect is modest.

EXAMPLE 4.6

[Participation Rates in 401(k) Plans]

In Example 3.3, we used the data on 401(k) plans to estimate a model describing participation rates in terms of the firm's match rate and the age of the plan. We now include a measure of firm size, the total number of firm employees (*totemp*). The estimated equation is

$$\widehat{prate} = 80.29 + 5.44\ mrate + .269\ age - .00013\ totemp$$
$$(0.78)\ (0.52)\qquad (.045)\qquad (.00004)$$
$$n = 1{,}534, R^2 = .100.$$

The smallest t statistic in absolute value is that on the variable *totemp*: $t = -.00013/.00004 = -3.25$, and this is statistically significant at very small significance levels. (The two-tailed p-value for this t statistic is about .001.) Thus, all of the variables are statistically significant at rather small significance levels.

How big, in a practical sense, is the coefficient on *totemp*? Holding *mrate* and *age* fixed, if a firm grows by 10,000 employees, the participation rate falls by 10,000(.00013) = 1.3 percentage points. This is a huge increase in number of employees with only a modest effect on the participation rate. Thus, although firm size does affect the participation rate, the effect is not practically very large.

The previous example shows that it is especially important to interpret the magnitude of the coefficient, in addition to looking at t statistics, when working with large samples. With large sample sizes, parameters can be estimated very precisely: standard errors are often quite small relative to the coefficient estimates, which usually results in statistical significance.

Some researchers insist on using smaller significance levels as the sample size increases, partly as a way to offset the fact that standard errors are getting smaller. For example, if we feel comfortable with a 5% level when n is a few hundred, we might use the 1% level when n is a few thousand. Using a smaller significance level means that economic and statistical significance are more likely to coincide, but there are no guarantees: in the previous example, even if we use a significance level as small as .1% (one-tenth of one percent), we would still conclude that *totemp* is statistically significant.

Most researchers are also willing to entertain larger significance levels in applications with small sample sizes, reflecting the fact that it is harder to find significance with smaller sample sizes (the critical values are larger in magnitude, and the estimators are less precise). Unfortunately, whether or not this is the case can depend on the researcher's underlying agenda.

EXAMPLE 4.7

(Effect of Job Training on Firm Scrap Rates)

The scrap rate for a manufacturing firm is the number of defective items—products that must be discarded—out of every 100 produced. Thus, for a given number of items produced, a decrease in the scrap rate reflects higher worker productivity

We can use the scrap rate to measure the effect of worker training on productivity. Using the data in JTRAIN.RAW, but only for the year 1987 and for nonunionized firms, we obtain the following estimated equation:

$$\widehat{\log(scrap)} = 12.46 - .029\ hrsemp - .962\log(sales) + .761\log(employ)$$
$$(5.69)\ (.023)\qquad (.453)\qquad (.407)$$
$$n = 29, R^2 = .262.$$

The variable *hrsemp* is annual hours of training per employee, *sales* is annual firm sales (in dollars), and *employ* is the number of firm employees. For 1987, the average scrap rate in the sample is about 4.6 and the average of *hrsemp* is about 8.9.

The main variable of interest is *hrsemp*. One more hour of training per employee lowers log(*scrap*) by .029, which means the scrap rate is about 2.9% lower. Thus, if *hrsemp* increases by 5—each employee is trained 5 more hours per year—the scrap rate is estimated to fall by 5(2.9) = 14.5%. This seems like a reasonably large effect, but whether the additional training is worthwhile to the firm depends on the cost of training and the benefits from a lower scrap rate. We do not have the numbers needed to do a cost benefit analysis, but the estimated effect seems nontrivial.

What about the *statistical significance* of the training variable? The *t* statistic on *hrsemp* is $-.029/.023 = -1.26$, and now you probably recognize this as not being large enough in magnitude to conclude that *hrsemp* is statistically significant at the 5% level. In fact, with $29 - 4 = 25$ degrees of freedom for the one-sided alternative, H_1: $\beta_{hrsemp} < 0$, the 5% critical value is about -1.71. Thus, using a strict 5% level test, we must conclude that *hrsemp* is not statistically significant, even using a one-sided alternative.

Because the sample size is pretty small, we might be more liberal with the significance level. The 10% critical value is -1.32, and so *hrsemp* is almost significant against the one-sided alternative at the 10% level. The *p*-value is easily computed as $P(T_{25} < -1.26) = .110$. This may be a low enough *p*-value to conclude that the estimated effect of training is not just due to sampling error, but some economists would have different opinions on this.

Remember that large standard errors can also be a result of multicollinearity (high correlation among some of the independent variables), even if the sample size seems fairly large. As we discussed in Section 3.4, there is not much we can do about this problem other than to collect more data or change the scope of the analysis by dropping or combining certain independent variables. As in the case of a small sample size, it can be hard to precisely estimate partial effects when some of the explanatory variables are highly correlated. (Section 4.5 contains an example.)

We end this section with some guidelines for discussing the economic and statistical significance of a variable in a multiple regression model:

1. Check for statistical significance. If the variable is statistically significant, discuss the magnitude of the coefficient to get an idea of its practical or economic importance. This latter step can require some care, depending on how the independent and dependent variables appear in the equation. (In particular, what are the units of measurement? Do the variables appear in logarithmic form?)
2. If a variable is not statistically significant at the usual levels (10%, 5%, or 1%), you might still ask if the variable has the expected effect on y and whether that effect is practically large. If it is large, you should compute a p-value for the t statistic. For small sample sizes, you can sometimes make a case for p-values as large as .20 (but there are no hard rules). With large p-values, that is, small t statistics, we are treading on thin ice because the practically large estimates may be due to sampling error: a different random sample could result in a very different estimate.

3. It is common to find variables with small t statistics that have the "wrong" sign. For practical purposes, these can be ignored: we conclude that the variables are statistically insignificant. A significant variable that has the unexpected sign and a practically large effect is much more troubling and difficult to resolve. One must usually think more about the model and the nature of the data in order to solve such problems. Often, a counterintuitive, significant estimate results from the omission of a key variable or from one of the important problems we will discuss in Chapters 9 and 15.

4.3 Confidence Intervals

Under the classical linear model assumptions, we can easily construct a **confidence interval (CI)** for the population parameter β_j. Confidence intervals are also called *interval estimates* because they provide a range of likely values for the population parameter, and not just a point estimate.

Using the fact that $(\hat{\beta}_j - \beta_j)/\mathrm{se}(\hat{\beta}_j)$ has a t distribution with $n - k - 1$ degrees of freedom [see (4.3)], simple manipulation leads to a CI for the unknown β_j: a *95% confidence interval*, given by

$$\hat{\beta}_j \pm c \cdot \mathrm{se}(\hat{\beta}_j), \tag{4.16}$$

where the constant c is the 97.5$^{\text{th}}$ percentile in a t_{n-k-1} distribution. More precisely, the lower and upper bounds of the confidence interval are given by

$$\underline{\beta}_j \equiv \hat{\beta}_j - c \cdot \mathrm{se}(\hat{\beta}_j)$$

and

$$\bar{\beta}_j \equiv \hat{\beta}_j + c \cdot \mathrm{se}(\hat{\beta}_j),$$

respectively.

At this point, it is useful to review the meaning of a confidence interval. If random samples were obtained over and over again, with $\underline{\beta}_j$, and $\bar{\beta}_j$ computed each time, then the (unknown) population value β_j would lie in the interval $(\underline{\beta}_j, \bar{\beta}_j)$ for 95% of the samples. Unfortunately, for the single sample that we use to construct the CI, we do not know whether β_j is actually contained in the interval. We hope we have obtained a sample that is one of the 95% of all samples where the interval estimate contains β_j, but we have no guarantee.

Constructing a confidence interval is very simple when using current computing technology. Three quantities are needed: $\hat{\beta}_j$, $\mathrm{se}(\hat{\beta}_j)$, and c. The coefficient estimate and its standard error are reported by any regression package. To obtain the value c, we must know the degrees of freedom, $n - k - 1$, and the level of confidence—95% in this case. Then, the value for c is obtained from the t_{n-k-1} distribution.

As an example, for $df = n - k - 1 = 25$, a 95% confidence interval for any β_j is given by $[\hat{\beta}_j - 2.06 \cdot \mathrm{se}(\hat{\beta}_j), \hat{\beta}_j + 2.06 \cdot \mathrm{se}(\hat{\beta}_j)]$.

When $n - k - 1 > 120$, the t_{n-k-1} distribution is close enough to normal to use the 97.5$^{\text{th}}$ percentile in a standard normal distribution for constructing a 95% CI: $\hat{\beta}_j \pm 1.96 \cdot \mathrm{se}(\hat{\beta}_j)$. In fact, when $n - k - 1 > 50$, the value of c is so close to 2 that we can use a simple *rule of thumb* for a 95% confidence interval: $\hat{\beta}_j$ plus or minus two of its standard errors. For small degrees of freedom, the exact percentiles should be obtained from the t tables.

It is easy to construct confidence intervals for any other level of confidence. For example, a 90% CI is obtained by choosing c to be the 95th percentile in the t_{n-k-1} distribution. When $df = n - k - 1 = 25$, $c = 1.71$, and so the 90% CI is $\hat{\beta}_j \pm 1.71 \cdot \text{se}(\hat{\beta}_j)$, which is necessarily narrower than the 95% CI. For a 99% CI, c is the 99.5th percentile in the t_{25} distribution. When $df = 25$, the 99% CI is roughly $\hat{\beta}_j \pm 2.79 \cdot \text{se}(\hat{\beta}_j)$, which is inevitably wider than the 95% CI.

Many modern regression packages save us from doing any calculations by reporting a 95% CI along with each coefficient and its standard error. Once a confidence interval is constructed, it is easy to carry out two-tailed hypotheses tests. If the null hypothesis is H_0: $\beta_j = a_j$, then H_0 is rejected against H_1: $\beta_j \neq a_j$ at (say) the 5% significance level if, and only if, a_j is *not* in the 95% confidence interval.

EXAMPLE 4.8

(Model of R&D Expenditures)

Economists studying industrial organization are interested in the relationship between firm size—often measured by annual sales—and spending on research and development (R&D). Typically, a constant elasticity model is used. One might also be interested in the ceteris paribus effect of the profit margin—that is, profits as a percentage of sales—on R&D spending. Using the data in RDCHEM.RAW, on 32 U.S. firms in the chemical industry, we estimate the following equation (with standard errors in parentheses below the coefficients):

$$\widehat{\log(rd)} = \underset{(.47)}{-4.38} + \underset{(.060)}{1.084} \log(sales) + \underset{(.0218)}{.0217}\, profmarg$$
$$n = 32, R^2 = .918.$$

The estimated elasticity of R&D spending with respect to firm sales is 1.084, so that, holding profit margin fixed, a 1 percent increase in sales is associated with a 1.084 percent increase in R&D spending. (Incidentally, R&D and sales are both measured in millions of dollars, but their units of measurement have no effect on the elasticity estimate.) We can construct a 95% confidence interval for the sales elasticity once we note that the estimated model has $n - k - 1 = 32 - 2 - 1 = 29$ degrees of freedom. From Table G.2, we find the 97.5th percentile in a t_{29} distribution: $c = 2.045$. Thus, the 95% confidence interval for $\beta_{\log(sales)}$ is $1.084 \pm .060(2.045)$, or about .961,1.21) That zero is well outside this interval is hardly surprising: we expect R&D spending to increase with firm size. More interesting is that unity is included in the 95% confidence interval for $\beta_{\log(sales)}$, which means that we cannot reject H_0:$\beta_{\log(sales)} = 1$ against H_1:$\beta_{\log(sales)} \neq 1$ at the 5% significance level. In other words, the estimated R&D-sales elasticity is not statistically different from 1 at the 5% level. (The estimate is not practically different from 1, either.)

The estimated coefficient on *profmarg* is also positive, and the 95% confidence interval for the population parameter, $\beta_{profmarg}$, is $.0217 \pm .0128(2.045)$, or about $(-.0045,.0479)$. In this case, zero is included in the 95% confidence interval, so we fail to reject H_0:$\beta_{profmarg} = 0$ against

$H_1: \beta_{profmarg} \neq 0$ at the 5% level. Nevertheless, the t statistic is about 1.70, which gives a two-sided p-value of about .10, and so we would conclude that *profmarg* is statistically significant at the 10% level against the two-sided alternative, or at the 5% level against the one-sided alternative $H_1: \beta_{profmarg} > 0$. Plus, the economic size of the profit margin coefficient is not trivial: holding *sales* fixed, a one percentage point increase in *profmarg* is estimated to increase R&D spending by $100(.0217) \approx 2.2$ percent. A complete analysis of this example goes beyond simply stating whether a particular value, zero in this case, is or is not in the 95% confidence interval.

You should remember that a confidence interval is only as good as the underlying assumptions used to construct it. If we have omitted important factors that are correlated with the explanatory variables, then the coefficient estimates are not reliable: OLS is biased. If heteroskedasticity is present—for instance, in the previous example, if the variance of $\log(rd)$ depends on any of the explanatory variables—then the standard error is not valid as an estimate of $\mathrm{sd}(\hat{\beta}_j)$ (as we discussed in Section 3.4), and the confidence interval computed using these standard errors will not truly be a 95% CI. We have also used the normality assumption on the errors in obtaining these CIs, but, as we will see in Chapter 5, this is not as important for applications involving hundreds of observations.

4.4 Testing Hypotheses about a Single Linear Combination of the Parameters

The previous two sections have shown how to use classical hypothesis testing or confidence intervals to test hypotheses about a single β_j at a time. In applications, we must often test hypotheses involving more than one of the population parameters. In this section, we show how to test a single hypothesis involving more than one of the β_j. Section 4.5 shows how to test multiple hypotheses.

To illustrate the general approach, we will consider a simple model to compare the returns to education at junior colleges and four-year colleges; for simplicity, we refer to the latter as "universities." (Kane and Rouse [1995] provide a detailed analysis of the returns to two- and four-year colleges.) The population includes working people with a high school degree, and the model is

$$\log(wage) = \beta_0 + \beta_1 jc + \beta_2 univ + \beta_3 exper + u, \tag{4.17}$$

where *jc* is number of years attending a two-year college, *univ* is number of years at a four-year college, and *exper* is months in the workforce. Note that any combination of junior college and four-year college is allowed, including $jc = 0$ and $univ = 0$.

The hypothesis of interest is whether one year at a junior college is worth one year at a university: this is stated as

$$H_0: \beta_1 = \beta_2. \tag{4.18}$$

Under H_0, another year at a junior college and another year at a university lead to the same ceteris paribus percentage increase in *wage*. For the most part, the alternative of interest is one-sided: a year at a junior college is worth less than a year at a university. This is stated as

$$H_1: \beta_1 < \beta_2. \tag{4.19}$$

The hypotheses in (4.18) and (4.19) concern *two* parameters, β_1 and β_2, a situation we have not faced yet. We cannot simply use the individual t statistics for $\hat{\beta}_1$ and $\hat{\beta}_2$ to test H_0. However, conceptually, there is no difficulty in constructing a t statistic for testing (4.18). In order to do so, we rewrite the null and alternative as H_0: $\beta_1 - \beta_2 = 0$ and H_1: $\beta_1 - \beta_2 < 0$, respectively. The t statistic is based on whether the estimated difference $\hat{\beta}_1 - \hat{\beta}_2$ is sufficiently less than zero to warrant rejecting (4.18) in favor of (4.19). To account for the sampling error in our estimators, we standardize this difference by dividing by the standard error:

$$t = \frac{\hat{\beta}_1 - \hat{\beta}_2}{\mathrm{se}(\hat{\beta}_1 - \hat{\beta}_2)}. \tag{4.20}$$

Once we have the t statistic in (4.20), testing proceeds as before. We choose a significance level for the test and, based on the *df*, obtain a critical value. Because the alternative is of the form in (4.19), the rejection rule is of the form $t < -c$, where c is a positive value chosen from the appropriate t distribution. Or, we compute the t statistic and then compute the p-value (see Section 4.2).

The only thing that makes testing the equality of two different parameters more difficult than testing about a single β_j is obtaining the standard error in the denominator of (4.20). Obtaining the numerator is trivial once we have performed the OLS regression. Using the data in TWOYEAR.RAW, which comes from Kane and Rouse (1995), we estimate equation (4.17):

$$\begin{aligned} \widehat{\log(wage)} = {} & 1.472 + .0667\ jc + .0769\ univ + .0049\ exper \\ & (.021)\quad (.0068)\qquad (.0023)\qquad\quad (.0002) \\ & n = 6{,}763,\ R^2 = .222. \end{aligned} \tag{4.21}$$

It is clear from (4.21) that *jc* and *univ* have both economically and statistically significant effects on wage. This is certainly of interest, but we are more concerned about testing whether the estimated *difference* in the coefficients is statistically significant. The difference is estimated as $\hat{\beta}_1 - \hat{\beta}_2 = -.0102$, so the return to a year at a junior college is about one percentage point less than a year at a university. Economically, this is not a trivial difference. The difference of $-.0102$ is the numerator of the t statistic in (4.20).

Unfortunately, the regression results in equation (4.21) do *not* contain enough information to obtain the standard error of $\hat{\beta}_1 - \hat{\beta}_2$. It might be tempting to claim that $\mathrm{se}(\hat{\beta}_1 - \hat{\beta}_2) = \mathrm{se}(\hat{\beta}_1) - \mathrm{se}(\hat{\beta}_2)$, but this is not true. In fact, if we reversed the roles of $\hat{\beta}_1$ and $\hat{\beta}_2$, we would wind up with a negative standard error of the difference using the difference in standard errors. Standard errors must *always* be positive because they are estimates of standard deviations. Although the standard error of the difference $\hat{\beta}_1 - \hat{\beta}_2$ certainly depends on

$\text{se}(\hat{\beta}_1)$ and $\text{se}(\hat{\beta}_2)$, it does so in a somewhat complicated way. To find $\text{se}(\hat{\beta}_1 - \hat{\beta}_2)$, we first obtain the variance of the difference. Using the results on variances in Appendix B, we have

$$\text{Var}(\hat{\beta}_1 - \hat{\beta}_2) = \text{Var}(\hat{\beta}_1) + \text{Var}(\hat{\beta}_2) - 2\,\text{Cov}(\hat{\beta}_1,\hat{\beta}_2). \tag{4.22}$$

Observe carefully how the two variances are *added* together, and twice the covariance is then subtracted. The standard deviation of $\hat{\beta}_1 - \hat{\beta}_2$ is just the square root of (4.22), and, since $[\text{se}(\hat{\beta}_1)]^2$ is an unbiased estimator of $\text{Var}(\hat{\beta}_1)$, and similarly for $[\text{se}(\hat{\beta}_2)]^2$, we have

$$\text{se}(\hat{\beta}_1 - \hat{\beta}_2) = \left\{[\text{se}(\hat{\beta}_1)]^2 + [\text{se}(\hat{\beta}_2)]^2 - 2s_{12}\right\}^{1/2}, \tag{4.23}$$

where s_{12} denotes an estimate of $\text{Cov}(\hat{\beta}_1,\hat{\beta}_2)$. We have not displayed a formula for $\text{Cov}(\hat{\beta}_1,\hat{\beta}_2)$. Some regression packages have features that allow one to obtain s_{12}, in which case one can compute the standard error in (4.23) and then the t statistic in (4.20). Appendix E shows how to use matrix algebra to obtain s_{12}.

Some of the more sophisticated econometrics programs include special commands that can be used for testing hypotheses about linear combinations. Here, we cover an approach that is simple to compute in virtually any statistical package. Rather than trying to compute $\text{se}(\hat{\beta}_1 - \hat{\beta}_2)$ from (4.23), it is much easier to estimate a different model that directly delivers the standard error of interest. Define a new parameter as the difference between β_1 and β_2: $\theta_1 = \beta_1 - \beta_2$. Then, we want to test

$$\text{H}_0\text{: } \theta_1 = 0 \text{ against H}_1\text{: } \theta_1 < 0. \tag{4.24}$$

The t statistic in (4.20) in terms of $\hat{\theta}_1$ is just $t = \hat{\theta}_1/\text{se}(\hat{\theta}_1)$. The challenge is finding $\text{se}(\hat{\theta}_1)$.

We can do this by rewriting the model so that θ_1 appears directly on one of the independent variables. Because $\theta_1 = \beta_1 - \beta_2$, we can also write $\beta_1 = \theta_1 + \beta_2$. Plugging this into (4.17) and rearranging gives the equation

$$\begin{aligned} \log(wage) &= \beta_0 + (\theta_1 + \beta_2)jc + \beta_2 univ + \beta_3 exper + u \\ &= \beta_0 + \theta_1 jc + \beta_2(jc + univ) + \beta_3 exper + u. \end{aligned} \tag{4.25}$$

The key insight is that the parameter we are interested in testing hypotheses about, θ_1, now multiplies the variable *jc*. The intercept is still β_0, and *exper* still shows up as being multiplied by β_3. More importantly, there is a new variable multiplying β_2, namely $jc + univ$. Thus, if we want to directly estimate θ_1 and obtain the standard error $\hat{\theta}_1$, then we must construct the new variable $jc + univ$ and include it in the regression model in place of *univ*. In this example, the new variable has a natural interpretation: it is *total* years of college, so define $totcoll = jc + univ$ and write (4.25) as

$$\log(wage) = \beta_0 + \theta_1 jc + \beta_2 totcoll + \beta_3 exper + u. \tag{4.26}$$

The parameter β_1 has disappeared from the model, while θ_1 appears explicitly. This model is really just a different way of writing the original model. The only reason we have defined this new model is that, when we estimate it, the coefficient on *jc* is $\hat{\theta}_1$, and, more importantly, se($\hat{\theta}_1$) is reported along with the estimate. The t statistic that we want is the one reported by any regression package on the variable *jc* (*not* the variable *totcoll*).

When we do this with the 6,763 observations used earlier, the result is

$$\begin{aligned}\widehat{\log(wage)} = \underset{(.021)}{1.472} - \underset{(.0069)}{.0102}\, jc + \underset{(.0023)}{.0769}\, totcoll + \underset{(.0002)}{.0049}\, exper \\ n = 6{,}763,\ R^2 = .222.\end{aligned} \tag{4.27}$$

The only number in this equation that we could not get from (4.21) is the standard error for the estimate $-.0102$, which is .0069. The t statistic for testing (4.18) is $-.0102/.0069 = -1.48$. Against the one-sided alternative (4.19), the p-value is about .070, so there is some, but not strong, evidence against (4.18).

The intercept and slope estimate on *exper*, along with their standard errors, are the same as in (4.21). This fact *must* be true, and it provides one way of checking whether the transformed equation has been properly estimated. The coefficient on the new variable, *totcoll*, is the same as the coefficient on *univ* in (4.21), and the standard error is also the same. We know that this must happen by comparing (4.17) and (4.25).

It is quite simple to compute a 95% confidence interval for $\theta_1 = \beta_1 - \beta_2$. Using the standard normal approximation, the CI is obtained as usual: $\hat{\theta}_1 \pm 1.96\ \text{se}(\hat{\theta}_1)$, which in this case leads to $-.0102 \pm .0135$.

The strategy of rewriting the model so that it contains the parameter of interest works in all cases and is easy to implement. (See Problems 4.12 and 4.14 for other examples.)

4.5 Testing Multiple Linear Restrictions: The *F* Test

The t statistic associated with any OLS coefficient can be used to test whether the corresponding unknown parameter in the population is equal to any given constant (which is usually, but not always, zero). We have just shown how to test hypotheses about a single linear combination of the β_j by rearranging the equation and running a regression using transformed variables. But so far, we have only covered hypotheses involving a *single* restriction. Frequently, we wish to test *multiple* hypotheses about the underlying parameters $\beta_0, \beta_1, \ldots, \beta_k$. We begin with the leading case of testing whether a set of independent variables has no partial effect on a dependent variable.

Testing Exclusion Restrictions

We already know how to test whether a particular variable has no partial effect on the dependent variable: use the t statistic. Now, we want to test whether a *group* of variables has no effect on the dependent variable. More precisely, the null hypothesis is that a set of variables has no effect on y, once another set of variables has been controlled.

As an illustration of why testing significance of a group of variables is useful, we consider the following model that explains major league baseball players' salaries:

$$\log(salary) = \beta_0 + \beta_1 years + \beta_2 gamesyr + \beta_3 bavg + \beta_4 hrunsyr + \beta_5 rbisyr + u, \tag{4.28}$$

where *salary* is the 1993 total salary, *years* is years in the league, *gamesyr* is average games played per year, *bavg* is career batting average (for example, $bavg = 250$), *hrunsyr* is home runs per year, and *rbisyr* is runs batted in per year. Suppose we want to test the null hypothesis that, once years in the league and games per year have been controlled for, the statistics measuring performance—*bavg*, *hrunsyr*, and *rbisyr*—have no effect on salary. Essentially, the null hypothesis states that productivity as measured by baseball statistics has no effect on salary.

In terms of the parameters of the model, the null hypothesis is stated as

$$\text{H}_0\text{: } \beta_3 = 0, \beta_4 = 0, \beta_5 = 0. \tag{4.29}$$

The null (4.29) constitutes three **exclusion restrictions**: if (4.29) is true, then *bavg*, *hrunsyr*, and *rbisyr* have no effect on log(*salary*) after *years* and *gamesyr* have been controlled for and therefore should be excluded from the model. This is an example of a set of **multiple restrictions** because we are putting more than one restriction on the parameters in (4.28); we will see more general examples of multiple restrictions later. A test of multiple restrictions is called a **multiple hypotheses test** or a **joint hypotheses test**.

What should be the alternative to (4.29)? If what we have in mind is that "performance statistics matter, even after controlling for years in the league and games per year," then the appropriate alternative is simply

$$\text{H}_1\text{: H}_0 \text{ is not true.} \tag{4.30}$$

The alternative (4.30) holds if at least one of β_3, β_4, or β_5 is different from zero. (Any or all could be different from zero.) The test we study here is constructed to detect any violation of H_0. It is also valid when the alternative is something like H_1: $\beta_3 > 0$, or $\beta_4 > 0$, or $\beta_5 > 0$, but it will not be the best possible test under such alternatives. We do not have the space or statistical background necessary to cover tests that have more power under multiple one-sided alternatives.

How should we proceed in testing (4.29) against (4.30)? It is tempting to test (4.29) by using the *t* statistics on the variables *bavg*, *hrunsyr*, and *rbisyr* to determine whether each variable is *individually* significant. This option is not appropriate. A particular *t* statistic tests a hypothesis that puts no restrictions on the other parameters. Besides, we would have three outcomes to contend with—one for each *t* statistic. What would constitute rejection of (4.29) at, say, the 5% level? Should all three or only one of the three *t* statistics be required to be significant at the 5% level? These are hard questions, and fortunately we do not have to answer them. Furthermore, using separate *t* statistics to test a multiple hypothesis like (4.29) can be very misleading. We need a way to test the exclusion restrictions *jointly*.

To illustrate these issues, we estimate equation (4.28) using the data in MLB1.RAW. This gives

$$
\begin{aligned}
\widehat{\log(salary)} = {} & \underset{(0.29)}{11.19} + \underset{(.0121)}{.0689}\, years + \underset{(.0026)}{.0126}\, gamesyr \\
& + \underset{(.00110)}{.00098}\, bavg + \underset{(.0161)}{.0144}\, hrunsyr + \underset{(.0072)}{.0108}\, rbisyr \\
& n = 353,\ \mathrm{SSR} = 183.186,\ R^2 = .6278,
\end{aligned}
\tag{4.31}
$$

where SSR is the sum of squared residuals. (We will use this later.) We have left several terms after the decimal in SSR and R-squared to facilitate future comparisons. Equation (4.31) reveals that, whereas *years* and *gamesyr* are statistically significant, none of the variables *bavg*, *hrunsyr*, and *rbisyr* has a statistically significant t statistic against a two-sided alternative, at the 5% significance level. (The t statistic on *rbisyr* is the closest to being significant; its two-sided p-value is .134.) Thus, based on the three t statistics, it appears that we cannot reject H_0.

This conclusion turns out to be wrong. In order to see this, we must derive a test of multiple restrictions whose distribution is known and tabulated. The sum of squared residuals now turns out to provide a very convenient basis for testing multiple hypotheses. We will also show how the R-squared can be used in the special case of testing for exclusion restrictions.

Knowing the sum of squared residuals in (4.31) tells us nothing about the truth of the hypothesis in (4.29). However, the factor that will tell us something is how much the SSR increases when we drop the variables *bavg*, *hrunsyr*, and *rbisyr* from the model. Remember that, because the OLS estimates are chosen to minimize the sum of squared residuals, the SSR *always* increases when variables are dropped from the model; this is an algebraic fact. The question is whether this increase is large enough, *relative* to the SSR in the model with all of the variables, to warrant rejecting the null hypothesis.

The model without the three variables in question is simply

$$
\log(salary) = \beta_0 + \beta_1 years + \beta_2 gamesyr + u. \tag{4.32}
$$

In the context of hypothesis testing, equation (4.32) is the **restricted model** for testing (4.29); model (4.28) is called the **unrestricted model**. The restricted model always has fewer parameters than the unrestricted model.

When we estimate the restricted model using the data in MLB1.RAW, we obtain

$$
\begin{aligned}
\widehat{\log(salary)} = {} & \underset{(.11)}{11.22} + \underset{(.0125)}{.0713}\, years + \underset{(.0013)}{.0202}\, gamesyr \\
& n = 353,\ \mathrm{SSR} = 198.311,\ R^2 = .5971.
\end{aligned}
\tag{4.33}
$$

As we surmised, the SSR from (4.33) is greater than the SSR from (4.31), and the R-squared from the restricted model is less than the R-squared from the unrestricted model. What we need to decide is whether the increase in the SSR in going from the unrestricted model to the restricted model (183.186 to 198.311) is large enough to warrant rejection of (4.29). As with all testing, the answer depends on the significance level of the test. But we cannot carry out the test at a chosen significance level until we have a statistic whose distribution is known, and can be tabulated, under H_0. Thus, we need a way to combine the information in the two SSRs to obtain a test statistic with a known distribution under H_0.

Because it is no more difficult, we might as well derive the test for the general case. Write the *unrestricted* model with k independent variables as

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + u; \tag{4.34}$$

the number of parameters in the unrestricted model is $k + 1$. (Remember to add one for the intercept.) Suppose that we have q exclusion restrictions to test: that is, the null hypothesis states that q of the variables in (4.34) have zero coefficients. For notational simplicity, assume that it is the last q variables in the list of independent variables: $x_{k-q+1}, \ldots, x_k$. (The order of the variables, of course, is arbitrary and unimportant.) The null hypothesis is stated as

$$H_0: \beta_{k-q+1} = 0, \ldots, \beta_k = 0, \tag{4.35}$$

which puts q exclusion restrictions on the model (4.34). The alternative to (4.35) is simply that it is false; this means that at least one of the parameters listed in (4.35) is different from zero. When we impose the restrictions under H_0, we are left with the restricted model:

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_{k-q} x_{k-q} + u. \tag{4.36}$$

In this subsection, we assume that both the unrestricted and restricted models contain an intercept, since that is the case most widely encountered in practice.

Now, for the test statistic itself. Earlier, we suggested that looking at the relative increase in the SSR when moving from the unrestricted to the restricted model should be informative for testing the hypothesis (4.35). The ***F* statistic** (or *F ratio*) is defined by

$$F \equiv \frac{(\mathrm{SSR}_r - \mathrm{SSR}_{ur})/q}{\mathrm{SSR}_{ur}/(n - k - 1)}, \tag{4.37}$$

where SSR_r is the sum of squared residuals from the restricted model and SSR_{ur} is the sum of squared residuals from the unrestricted model.

You should immediately notice that, since SSR_r can be no smaller than SSR_{ur}, the F statistic is *always* nonnegative (and almost always strictly positive). Thus, if you

QUESTION 4.4

Consider relating individual performance on a standardized test, *score*, to a variety of other variables. School factors include average class size, per student expenditures, average teacher compensation, and total school enrollment. Other variables specific to the student are family income, mother's education, father's education, and number of siblings. The model is

$$score = \beta_0 + \beta_1 classize + \beta_2 expend + \beta_3 tchcomp + \beta_4 enroll + \beta_5 faminc + \beta_6 motheduc + \beta_7 fatheduc + \beta_8 siblings + u.$$

State the null hypothesis that student-specific variables have no effect on standardized test performance, once school-related factors have been controlled for. What are k and q for this example? Write down the restricted version of the model.

compute a negative F statistic, then something is wrong; the order of the SSRs in the numerator of F has usually been reversed. Also, the SSR in the denominator of F is the SSR from the *unrestricted* model. The easiest way to remember where the SSRs appear is to think of F as measuring the relative increase in SSR when moving from the unrestricted to the restricted model.

The difference in SSRs in the numerator of F is divided by q, which is the number of restrictions imposed in moving from the unrestricted to the restricted model (q independent variables are dropped). Therefore, we can write

$$q = \textbf{numerator degrees of freedom} = df_r - df_{ur}, \tag{4.38}$$

which also shows that q is the difference in degrees of freedom between the restricted and unrestricted models. (Recall that df = number of observations − number of estimated parameters.) Since the restricted model has fewer parameters—and each model is estimated using the same n observations—df_r is always greater than df_{ur}.

The SSR in the denominator of F is divided by the degrees of freedom in the unrestricted model:

$$n - k - 1 = \textbf{denominator degrees of freedom} = df_{ur}. \tag{4.39}$$

In fact, the denominator of F is just the unbiased estimator of $\sigma^2 = \text{Var}(u)$ in the unrestricted model.

In a particular application, computing the F statistic is easier than wading through the somewhat cumbersome notation used to describe the general case. We first obtain the degrees of freedom in the unrestricted model, df_{ur}. Then, we count how many variables are excluded in the restricted model; this is q. The SSRs are reported with every OLS regression, and so forming the F statistic is simple.

In the major league baseball salary regression, $n = 353$, and the full model (4.28) contains six parameters. Thus, $n - k - 1 = df_{ur} = 353 - 6 = 347$. The restricted model (4.32) contains three fewer independent variables than (4.28), and so $q = 3$. Thus, we have all of the ingredients to compute the F statistic; we hold off doing so until we know what to do with it.

In order to use the F statistic, we must know its sampling distribution under the null in order to choose critical values and rejection rules. It can be shown that, under H_0 (and assuming the CLM assumptions hold), F is distributed as an F random variable with $(q, n - k - 1)$ degrees of freedom. We write this as

$$F \sim F_{q,n-k-1}.$$

The distribution of $F_{q,n-k-1}$ is readily tabulated and available in statistical tables (see Table G.3) and, even more importantly, in statistical software.

We will not derive the F distribution because the mathematics is very involved. Basically, it can be shown that equation (4.37) is actually the ratio of two independent chi-square random variables, divided by their respective degrees of freedom. The numerator chi-square random variable has q degrees of freedom, and the chi-square in the denominator has $n - k - 1$ degrees of freedom. This is the definition of an F distributed random variable (see Appendix B).

It is pretty clear from the definition of F that we will reject H_0 in favor of H_1 when F is sufficiently "large." How large depends on our chosen significance level. Suppose that we have decided on a 5% level test. Let c be the 95th percentile in the $F_{q,n-k-1}$ distribution. This critical value depends on q (the numerator df) and $n - k - 1$ (the denominator df). It is important to keep the numerator and denominator degrees of freedom straight.

The 10%, 5%, and 1% critical values for the F distribution are given in Table G.3. The rejection rule is simple. Once c has been obtained, we reject H_0 in favor of H_1 at the chosen significance level if

$$F > c. \tag{4.40}$$

With a 5% significance level, $q = 3$, and $n - k - 1 = 60$, the critical value is $c = 2.76$. We would reject H_0 at the 5% level if the computed value of the F statistic exceeds 2.76. The 5% critical value and rejection region are shown in Figure 4.7. For the same degrees of freedom, the 1% critical value is 4.13.

In most applications, the numerator degrees of freedom (q) will be notably smaller than the denominator degrees of freedom ($n - k - 1$). Applications where $n - k - 1$ is small are unlikely to be successful because the parameters in the unrestricted model will probably not be precisely estimated. When the denominator df reaches about 120, the F distribution is no longer sensitive to it. (This is entirely analogous to the t distribution being well approximated by the standard normal distribution as the df gets large.) Thus, there is an entry in the table for the denominator $df = \infty$, and this is what we use with large samples (because $n - k - 1$ is then large). A similar statement holds for a very large numerator df, but this rarely occurs in applications.

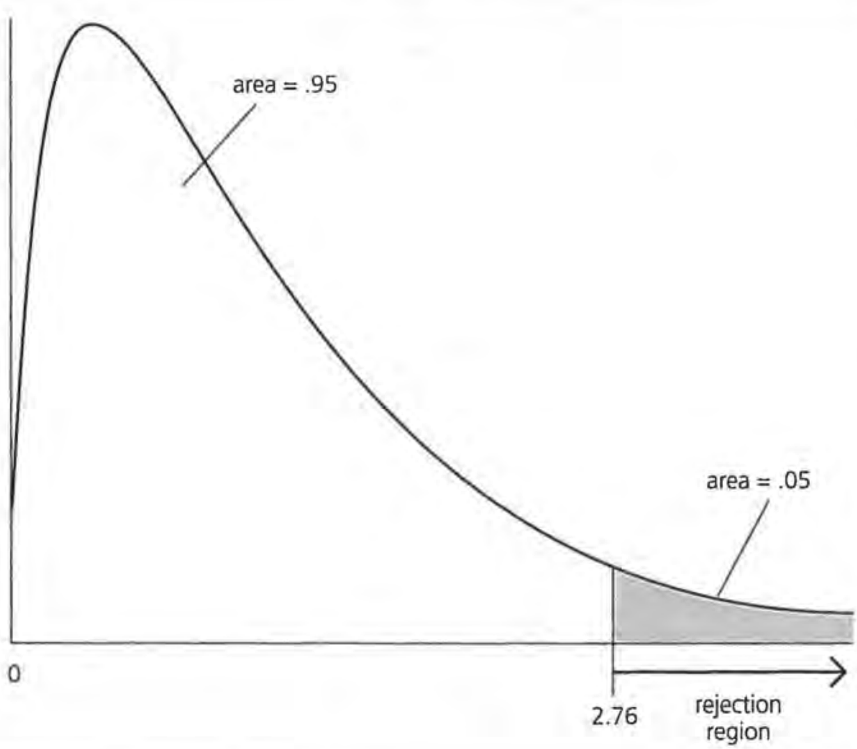
If H_0 is rejected, then we say that $x_{k-q+1}, \ldots, x_k$ are **jointly statistically significant** (or just *jointly significant*) at the appropriate significance level. This test alone does not allow us to say which of the variables has a partial effect on y; they may all affect y or maybe only one affects y. If the null is not rejected, then the variables are **jointly insignificant**, which often justifies dropping them from the model.

For the major league baseball example with three numerator degrees of freedom and 347 denominator degrees of freedom, the 5% critical value is 2.60, and the 1% critical value is 3.78. We reject H_0 at the 1% level if F is above 3.78; we reject at the 5% level if F is above 2.60.

We are now in a position to test the hypothesis that we began this section with: after controlling for *years* and *gamesyr*, the variables *bavg*, *hrunsyr*, and *rbisyr* have no effect on players' salaries. In practice, it is easiest to first compute $(\text{SSR}_r - \text{SSR}_{ur})/\text{SSR}_{ur}$ and to multiply the result by $(n - k - 1)/q$; the reason the formula is stated as in (4.37) is that

FIGURE 4.7

The 5% critical value and rejection region in an $F_{3,60}$ distribution.



it makes it easier to keep the numerator and denominator degrees of freedom straight. Using the SSRs in (4.31) and (4.33), we have

$$F = \frac{(198.311 - 183.186)}{183.186} \cdot \frac{347}{3} \approx 9.55.$$

This number is well above the 1% critical value in the F distribution with 3 and 347 degrees of freedom, and so we soundly reject the hypothesis that *bavg*, *hrunsyr*, and *rbisyr* have no effect on salary.

The outcome of the joint test may seem surprising in light of the insignificant t statistics for the three variables. What is happening is that the two variables *hrunsyr* and *rbisyr* are highly correlated, and this multicollinearity makes it difficult to uncover the partial effect of each variable; this is reflected in the individual t statistics. The F statistic tests whether these variables (including *bavg*) are *jointly* significant, and multicollinearity between *hrunsyr* and *rbisyr* is much less relevant for testing this hypothesis.

In Problem 4.16, you are asked to reestimate the model while dropping *rbisyr*, in which case *hrunsyr* becomes very significant. The same is true for *rbisyr* when *hrunsyr* is dropped from the model.

The F statistic is often useful for testing exclusion of a group of variables when the variables in the group are highly correlated. For example, suppose we want to test whether firm performance affects the salaries of chief executive officers. There are many ways to measure firm performance, and it probably would not be clear ahead of time which measures would be most important. Since measures of firm performance are likely to be highly correlated, hoping to find individually significant measures might be asking too much due to multicollinearity. But an F test can be used to determine whether, as a group, the firm performance variables affect salary.

Relationship between F and t Statistics

We have seen in this section how the F statistic can be used to test whether a group of variables should be included in a model. What happens if we apply the F statistic to the case of testing significance of a *single* independent variable? This case is certainly not ruled out by the previous development. For example, we can take the null to be H_0: $\beta_k = 0$ and $q = 1$ (to test the single exclusion restriction that x_k can be excluded from the model). From Section 4.2, we know that the t statistic on β_k can be used to test this hypothesis. The question, then, is do we have two separate ways of testing hypotheses about a single coefficient? The answer is no. It can be shown that the F statistic for testing exclusion of a single variable is equal to the *square* of the corresponding t statistic. Since t^2_{n-k-1} has an $F_{1,n-k-1}$ distribution, the two approaches lead to exactly the same outcome, provided that the alternative is two-sided. The t statistic is more flexible for testing a single hypothesis because it can be used to test against one-sided alternatives. Since t statistics are also easier to obtain than F statistics, there is really no reason to use an F statistic to test hypotheses about a single parameter.

We have already seen in the salary regressions for major league baseball players that two (or more) variables that each have insignificant t statistics can be jointly very significant. It is also possible that, in a group of several explanatory variables, one variable has a significant t statistic, but the group of variables is jointly insignificant at the usual significance levels. What should we make of this kind of outcome? For concreteness, suppose that in a model with many explanatory variables we cannot reject the null hypothesis that β_1, β_2, β_3, β_4, and β_5 are all equal to zero at the 5% level, yet the t statistic for $\hat{\beta}_1$ is significant at the 5% level. Logically, we cannot have $\beta_1 \neq 0$ but also have β_1, β_2, β_3, β_4, and β_5 all equal to zero! But as a matter of testing, it is possible that we can group a bunch of insignificant variables with a significant variable and conclude that the entire set of variables is jointly insignificant. (Such possible conflicts between a t test and a joint F test give another example of why we should not "accept" null hypotheses; we should only fail to reject them.) The F statistic is intended to detect whether a set of coefficients is different from zero, but it is never the best test for determining whether a single coefficient is different from zero. The t test is best suited for testing a single hypothesis. (In statistical terms, an F statistic for joint restrictions including $\beta_1 = 0$ will have less power for detecting $\beta_1 \neq 0$ than the usual t statistic. See Section C.6 in Appendix C for a discussion of the power of a test.)

Unfortunately, the fact that we can sometimes hide a statistically significant variable along with some insignificant variables could lead to abuse if regression results are not carefully reported. For example, suppose that, in a study of the determinants of loan-acceptance rates at the city level, x_1 is the fraction of black households in the city. Suppose that the variables x_2, x_3, x_4, and x_5 are the fractions of households headed by different age groups. In explaining loan rates, we would include measures of income, wealth, credit ratings, and so on. Suppose that age of household head has no effect on loan approval rates, once other variables are controlled for. Even if race has a marginally significant effect, it is possible that the race and age variables could be jointly insignificant. Someone wanting to conclude that race is not a factor could simply report something like "Race and age variables were added to the equation, but they were jointly insignificant at the 5% level." Hopefully, peer review prevents these kinds of misleading conclusions, but you should be aware that such outcomes are possible.

Often, when a variable is very statistically significant and it is tested jointly with another set of variables, the set will be jointly significant. In such cases, there is no logical inconsistency in rejecting both null hypotheses.

The *R*-Squared Form of the *F* Statistic

For testing exclusion restrictions, it is often more convenient to have a form of the F statistic that can be computed using the R-squareds from the restricted and unrestricted models. One reason for this is that the R-squared is always between zero and one, whereas the SSRs can be very large depending on the unit of measurement of y, making the calculation based on the SSRs tedious. Using the fact that $\text{SSR}_r = \text{SST}(1 - R_r^2)$ and $\text{SSR}_{ur} = \text{SST}(1 - R_{ur}^2)$, we can substitute into (4.37) to obtain

$$F = \frac{(R_{ur}^2 - R_r^2)/q}{(1 - R_{ur}^2)/(n - k - 1)} = \frac{(R_{ur}^2 - R_r^2)/q}{(1 - R_{ur}^2)/df_{ur}} \tag{4.41}$$

(note that the SST terms cancel everywhere). This is called the ***R*-squared form of the *F* statistic**. [At this point, you should be cautioned that although equation (4.41) is very convenient for testing exclusion restrictions, it cannot be applied for testing all linear restrictions. As we will see when we discuss testing general linear restrictions, the sum of squared residuals form of the F statistic is sometimes needed.]

Because the R-squared is reported with almost all regressions (whereas the SSR is not), it is easy to use the R-squareds from the unrestricted and restricted models to test for exclusion of some variables. Particular attention should be paid to the order of the R-squareds in the numerator: the *unrestricted* R-squared comes first [contrast this with the SSRs in (4.37)]. Because $R_{ur}^2 > R_r^2$, this shows again that F will always be positive.

In using the R-squared form of the test for excluding a set of variables, it is important to *not* square the R-squared before plugging it into formula (4.41); the squaring has already been done. All regressions report R^2, and these numbers are plugged directly into (4.41). For the baseball salary example, we can use (4.41) to obtain the F statistic:

$$F = \frac{(.6278 - .5971)}{(1 - .6278)} \cdot \frac{347}{3} \approx 9.54,$$

which is very close to what we obtained before. (The difference is due to rounding error.)

EXAMPLE 4.9

(Parents' Education in a Birth Weight Equation)

As another example of computing an F statistic, consider the following model to explain child birth weight in terms of various factors:

$$\begin{aligned} bwght = \beta_0 + \beta_1 cigs + \beta_2 parity + \beta_3 faminc + \\ \beta_4 motheduc + \beta_5 fatheduc + u, \end{aligned} \tag{4.42}$$

where *bwght* is birth weight, in pounds, *cigs* is average number of cigarettes the mother smoked per day during pregnancy, *parity* is the birth order of this child, *faminc* is annual family income, *motheduc* is years of schooling for the mother, and *fatheduc* is years of schooling for the father. Let us test the null hypothesis that, after controlling for *cigs*, *parity*, and *faminc*, parents' education has no effect on birth weight. This is stated as H_0: $\beta_4 = 0$, $\beta_5 = 0$, and so there are $q = 2$ exclusion restrictions to be tested. There are $k + 1 = 6$ parameters in the unrestricted model (4.42), so the *df* in the unrestricted model is $n - 6$, where n is the sample size.

We will test this hypothesis using the data in BWGHT.RAW. This data set contains information on 1,388 births, but we must be careful in counting the observations used in testing the null hypothesis. It turns out that information on at least one of the variables *motheduc* and *fatheduc* is missing for 197 births in the sample; these observations cannot be included when estimating the unrestricted model. Thus, we really have $n = 1{,}191$ observations, and so there are $1{,}191 - 6 = 1{,}185$ *df* in the unrestricted model. We must be sure to use these *same* 1,191 observations when estimating the restricted model (not the full 1,388 observations that are available). Generally, when estimating the restricted model to compute an F test, we must use the same observations to estimate the unrestricted model; otherwise, the test is not valid. When there are no missing data, this will not be an issue.

The numerator *df* is 2, and the denominator *df* is 1,185; from Table G.3, the 5% critical value is $c = 3.0$. Rather than report the complete results, for brevity, we present only the R-squareds. The R-squared for the full model turns out to be $R^2_{ur} = .0387$. When *motheduc* and *fatheduc* are dropped from the regression, the R-squared falls to $R^2_r = .0364$. Thus, the F statistic is $F = [(.0387 - .0364)/(1 - .0387)](1{,}185/2) = 1.42$; since this is well below the 5% critical value, we fail to reject H_0. In other words, *motheduc and fatheduc* are jointly insignificant in the birth weight equation.

Computing p-Values for F Tests

For reporting the outcomes of F tests, p-values are especially useful. Since the F distribution depends on the numerator and denominator *df*, it is difficult to get a feel for how strong or weak the evidence is against the null hypothesis simply by looking at the value of the F statistic and one or two critical values.

In the F testing context, the p-value is defined as

$$p\text{-value} = \mathrm{P}(\mathcal{F} > F), \tag{4.43}$$

QUESTION 4.5

The data in ATTEND.RAW were used to estimate the two equations

$$\widehat{atndrte} = \underset{(2.87)}{47.13} + \underset{(1.09)}{13.37}\ priGPA$$
$$n = 680, R^2 = .183,$$

and

$$\widehat{atndrte} = \underset{(3.88)}{75.70} + \underset{(1.08)}{17.26}\ priGPA - \underset{(?)}{1.72}\ ACT$$
$$n = 680, R^2 = .291,$$

where, as always, standard errors are in parentheses; the standard error for *ACT* is missing in the second equation. What is the t statistic for the coefficient on *ACT*? (*Hint:* First compute the F statistic for significance of *ACT*.)

where, for emphasis, we let $\mathcal{F}$ denote an F random variable with $(q, n - k - 1)$ degrees of freedom, and F is the actual value of the test statistic. The p-value still has the same interpretation as it did for t statistics: it is the probability of observing a value of F at least as large as we did, *given* that the null hypothesis is true. A small p-value is evidence against H_0. For example, p-value $= .016$ means that the chance of observing a value of F as large as we did when the null hypothesis was true is only 1.6%; we usually reject H_0 in such cases. If the p-value $= .314$, then the chance of observing a value of the F statistic as large as we did under the null hypothesis is 31.4%. Most would find this to be pretty weak evidence against H_0.

As with t testing, once the p-value has been computed, the F test can be carried out at any significance level. For example, if the p-value $= .024$, we reject H_0 at the 5% significance level but not at the 1% level.

The p-value for the F test in Example 4.9 is .238, and so the null hypothesis that $\beta_{motheduc}$ and $\beta_{fatheduc}$ are both zero is not rejected at even the 20% significance level.

Many econometrics packages have a built-in feature for testing multiple exclusion restrictions. These packages have several advantages over calculating the statistics by hand: we will less likely make a mistake, p-values are computed automatically, and the problem of missing data, as in Example 4.9, is handled without any additional work on our part.

The F Statistic for Overall Significance of a Regression

A special set of exclusion restrictions is routinely tested by most regression packages. These restrictions have the same interpretation, regardless of the model. In the model with k independent variables, we can write the null hypothesis as

$$H_0: x_1, x_2, \ldots, x_k \text{ do not help to explain } y.$$

This null hypothesis is, in a way, very pessimistic. It states that *none* of the explanatory variables has an effect on y. Stated in terms of the parameters, the null is that all slope parameters are zero:

$$H_0: \beta_1 = \beta_2 = \ldots = \beta_k = 0, \qquad \textbf{(4.44)}$$

and the alternative is that at least one of the β_j is different from zero. Another useful way of stating the null is that H_0: $E(y|x_1, x_2, \ldots, x_k) = E(y)$, so that knowing the values of x_1, $x_2, \ldots, x_k$ does not affect the expected value of y.

There are k restrictions in (4.44), and when we impose them, we get the restricted model

$$y = \beta_0 + u; \tag{4.45}$$

all independent variables have been dropped from the equation. Now, the R-squared from estimating (4.45) is zero; none of the variation in y is being explained because there are no explanatory variables. Therefore, the F statistic for testing (4.44) can be written as

$$\frac{R^2/k}{(1 - R^2)/(n - k - 1)}, \tag{4.46}$$

where R^2 is just the usual R-squared from the regression of y on $x_1, x_2, \ldots, x_k$.

Most regression packages report the F statistic in (4.46) automatically, which makes it tempting to use this statistic to test general exclusion restrictions. You must avoid this temptation. The F statistic in (4.41) is used for general exclusion restrictions; it depends on the R-squareds from the restricted and unrestricted models. The special form of (4.46) is valid only for testing joint exclusion of *all* independent variables. This is sometimes called determining the **overall significance of the regression.**

If we fail to reject (4.44), then there is no evidence that any of the independent variables help to explain y. This usually means that we must look for other variables to explain y. For Example 4.9, the F statistic for testing (4.44) is about 9.55 with $k = 5$ and $n - k - 1 = 1{,}185$ *df.* The p-value is zero to four places after the decimal point, so that (4.44) is rejected very strongly. Thus, we conclude that the variables in the *bwght* equation *do* explain some variation in *bwght*. The amount explained is not large: only 3.87%. But the seemingly small R-squared results in a highly significant F statistic. That is why we must compute the F statistic to test for joint significance and not just look at the size of the R-squared.

Occasionally, the F statistic for the hypothesis that all independent variables are jointly insignificant is the focus of a study. Problem 4.10 asks you to use stock return data to test whether stock returns over a four-year horizon are predictable based on information known only at the beginning of the period. Under the *efficient markets hypothesis*, the returns should not be predictable; the null hypothesis is precisely (4.44).

Testing General Linear Restrictions

Testing exclusion restrictions is by far the most important application of F statistics. Sometimes, however, the restrictions implied by a theory are more complicated than just excluding some independent variables. It is still straightforward to use the F statistic for testing.

As an example, consider the following equation:

$$\begin{aligned}\log(price) = \beta_0 &+ \beta_1\log(assess) + \beta_2\log(lotsize) \\ &+ \beta_3\log(sqrft) + \beta_4 bdrms + u,\end{aligned} \tag{4.47}$$

where *price* is house price, *assess* is the assessed housing value (before the house was sold), *lotsize* is size of the lot, in feet, *sqrft* is square footage, and *bdrms* is number of bedrooms. Now, suppose we would like to test whether the assessed housing price is a rational valuation. If this is the case, then a 1% change in *assess* should be associated with a 1% change in *price*; that is, $\beta_1 = 1$. In addition, *lotsize*, *sqrft*, and *bdrms* should not help to explain log(*price*), once the assessed value has been controlled for. Together, these hypotheses can be stated as

$$H_0: \beta_1 = 1, \beta_2 = 0, \beta_3 = 0, \beta_4 = 0. \tag{4.48}$$

There are four restrictions here to be tested; three are exclusion restrictions, but $\beta_1 = 1$ is not. How can we test this hypothesis using the F statistic?

As in the exclusion restriction case, we estimate the unrestricted model, (4.47) in this case, and then impose the restrictions in (4.48) to obtain the restricted model. It is the second step that can be a little tricky. But all we do is plug in the restrictions. If we write (4.47) as

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + u, \tag{4.49}$$

then the restricted model is $y = \beta_0 + x_1 + u$. Now, in order to impose the restriction that the coefficient on x_1 is unity, we must estimate the following model:

$$y - x_1 = \beta_0 + u. \tag{4.50}$$

This is just a model with an intercept (β_0) but with a different dependent variable than in (4.49). The procedure for computing the F statistic is the same: estimate (4.50), obtain the SSR (SSR_r), and use this with the unrestricted SSR from (4.49) in the F statistic (4.37). We are testing $q = 4$ restrictions, and there are $n - 5$ *df* in the unrestricted model. The F statistic is simply $[(SSR_r - SSR_{ur})/SSR_{ur}][(n - 5)/4]$.

Before illustrating this test using a data set, we must emphasize one point: we cannot use the R-squared form of the F statistic for this example because the dependent variable in (4.50) is different from the one in (4.49). This means the total sum of squares from the two regressions will be different, and (4.41) is no longer equivalent to (4.37). As a general rule, the SSR form of the F statistic should be used if a different dependent variable is needed in running the restricted regression.

The estimated unrestricted model using the data in HPRICE1.RAW is

$$\begin{aligned}
\widehat{\log(price)} = &\underset{(.570)}{.264} + \underset{(.151)}{1.043}\ \log(assess) + \underset{(.0386)}{.0074}\ \log(lotsize) \\
&- \underset{(.1384)}{.1032}\ \log(sqrft) + \underset{(.0221)}{.0338}\ bdrms \\
&n = 88,\ SSR = 1.822,\ R^2 = .773.
\end{aligned}$$

If we use separate t statistics to test each hypothesis in (4.48), we fail to reject each one. But rationality of the assessment is a joint hypothesis, so we should test the restrictions jointly. The SSR from the restricted model turns out to be $SSR_r = 1.880$, and so the F statistic is $[(1.880 - 1.822)/1.822](83/4) = .661$. The 5% critical value in an F distribution with (4,83) *df* is about 2.50, and so we fail to reject H_0. There is essentially no evidence against the hypothesis that the assessed values are rational.

4.6 Reporting Regression Results

We end this chapter by providing a few guidelines on how to report multiple regression results for relatively complicated empirical projects. This should help you to read published works in the applied social sciences, while also preparing you to write your own empirical papers. We will expand on this topic in the remainder of the text by reporting results from various examples, but many of the key points can be made now.

Naturally, the estimated OLS coefficients should always be reported. For the key variables in an analysis, you should *interpret* the estimated coefficients (which often requires knowing the units of measurement of the variables). For example, is an estimate an elasticity, or does it have some other interpretation that needs explanation? The economic or practical importance of the estimates of the key variables should be discussed.

The standard errors should always be included along with the estimated coefficients. Some authors prefer to report the t statistics rather than the standard errors (and sometimes just the absolute value of the t statistics). Although nothing is really wrong with this, there is some preference for reporting standard errors. First, it forces us to think carefully about the null hypothesis being tested; the null is not always that the population parameter is zero. Second, having standard errors makes it easier to compute confidence intervals.

The R-squared from the regression should always be included. We have seen that, in addition to providing a goodness-of-fit measure, it makes calculation of F statistics for exclusion restrictions simple. Reporting the sum of squared residuals and the standard error of the regression is sometimes a good idea, but it is not crucial. The number of observations used in estimating any equation should appear near the estimated equation.

If only a couple of models are being estimated, the results can be summarized in equation form, as we have done up to this point. However, in many papers, several equations are estimated with many different sets of independent variables. We may estimate the same equation for different groups of people, or even have equations explaining different dependent variables. In such cases, it is better to summarize the results in one or more tables. The dependent variable should be indicated clearly in the table, and the independent variables should be listed in the first column. Standard errors (or t statistics) can be put in parentheses below the estimates.

EXAMPLE 4.10

(Salary-Pension Tradeoff for Teachers)

Let *totcomp* denote average total annual compensation for a teacher, including salary and all fringe benefits (pension, health insurance, and so on). Extending the standard wage equation, total compensation should be a function of productivity and perhaps other characteristics. As is standard, we use logarithmic form:

$$\log(totcomp) = f(productivity\ characteristics, other\ factors),$$

where $f(\cdot)$ is some function (unspecified for now). Write

$$totcomp = salary + benefits = salary\left(1 + \frac{benefits}{salary}\right).$$

This equation shows that total compensation is the product of two terms: *salary* and $1 + b/s$, where b/s is shorthand for the "benefits to salary ratio." Taking the log of this equation gives $\log(totcomp) = \log(salary) + \log(1 + b/s)$. Now, for "small" b/s, $\log(1 + b/s) \approx b/s$; we will use this approximation. This leads to the econometric model

$$\log(salary) = \beta_0 + \beta_1(b/s) + other\ factors.$$

Testing the salary-benefits tradeoff then is the same as a test of H_0: $\beta_1 = -1$ against H_1: $\beta_1 \neq -1$.

We use the data in MEAP93.RAW to test this hypothesis. These data are averaged at the school level, and we do not observe very many other factors that could affect total compensation. We will include controls for size of the school (*enroll*), staff per thousand students (*staff*), and measures such as the school dropout and graduation rates. The average b/s in the sample is about .205, and the largest value is .450.

The estimated equations are given in Table 4.1, where standard errors are given in parentheses below the coefficient estimates. The key variable is b/s, the benefits-salary ratio.

TABLE 4.1

Testing the Salary-Benefits Tradeoff

Dependent Variable: log(*salary*)			
Independent Variables	**(1)**	**(2)**	**(3)**
b/s	−.825 (.200)	−.605 (.165)	−.589 (.165)
log(*enroll*)	——	.0874 (.0073)	.0881 (.0073)
log(*staff*)	——	−.222 (.050)	−.218 (.050)
droprate	——	——	−.00028 (.00161)
gradrate	——	——	.00097 (.00066)
intercept	10.523 (0.042)	10.884 (0.252)	10.738 (0.258)
Observations *R*-Squared	408 .040	408 .353	408 .361

QUESTION 4.6

How does adding *droprate* and *gradrate* affect the estimate of the salary-benefits tradeoff? Are these variables jointly significant at the 5% level? What about the 10% level?

From the first column in Table 4.1, we see that, without controlling for any other factors, the OLS coefficient for *b/s* is $-.825$. The t statistic for testing the null hypothesis H_0: $\beta_1 = -1$ is $t = (-.825 + 1)/.200 = .875$, and so the simple regression fails to reject H_0. After adding controls for school size and staff size (which roughly captures the number of students taught by each teacher), the estimate of the *b/s* coefficient becomes $-.605$. Now, the test of $\beta_1 = -1$ gives a t statistic of about 2.39; thus, H_0 is rejected at the 5% level against a two-sided alternative. The variables log(*enroll*) and log(*staff*) are very statistically significant.

SUMMARY

In this chapter, we have covered the very important topic of statistical inference, which allows us to infer something about the population model from a random sample. We summarize the main points:

1. Under the classical linear model assumptions MLR.1 through MLR.6, the OLS estimators are normally distributed.
2. Under the CLM assumptions, the t statistics have t distributions under the null hypothesis.
3. We use t statistics to test hypotheses about a single parameter against one- or two-sided alternatives, using one- or two-tailed tests, respectively. The most common null hypothesis is H_0: $\beta_j = 0$, but we sometimes want to test other values of β_j under H_0.
4. In classical hypothesis testing, we first choose a significance level, which, along with the *df* and alternative hypothesis, determines the critical value against which we compare the t statistic. It is more informative to compute the p-value for a t test—the smallest significance level for which the null hypothesis is rejected—so that the hypothesis can be tested at any significance level.
5. Under the CLM assumptions, confidence intervals can be constructed for each β_j. These CIs can be used to test any null hypothesis concerning β_j against a two-sided alternative.
6. Single hypothesis tests concerning more than one β_j can always be tested by rewriting the model to contain the parameter of interest. Then, a standard t statistic can be used.
7. The F statistic is used to test multiple exclusion restrictions, and there are two equivalent forms of the test. One is based on the SSRs from the restricted and unrestricted models. A more convenient form is based on the R-squareds from the two models.
8. When computing an F statistic, the numerator *df* is the number of restrictions being tested, while the denominator *df* is the degrees of freedom in the unrestricted model.

9. The alternative for F testing is two-sided. In the classical approach, we specify a significance level which, along with the numerator *df* and the denominator *df*, determines the critical value. The null hypothesis is rejected when the statistic, F, exceeds the critical value, c. Alternatively, we can compute a p-value to summarize the evidence against H_0.
10. General multiple linear restrictions can be tested using the sum of squared residuals form of the F statistic.
11. The F statistic for the overall significance of a regression tests the null hypothesis that *all* slope parameters are zero, with the intercept unrestricted. Under H_0, the explanatory variables have no effect on the expected value of y.

The Classical Linear Model Assumptions

Now is a good time to review the full set of classical linear model (CLM) assumptions for cross-sectional regression. Following each assumption is a comment about its role in multiple regression analysis.

Assumption MLR.1 (Linear in Parameters)

The model in the population can be written as

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u,$$

where $\beta_0, \beta_1, \ldots, \beta_k$ are the unknown parameters (constants) of interest and u is an unobservable random error or disturbance term.

Assumption MLR.1 describes the population relationship we hope to estimate, and explicitly sets out the β_j—the ceteris paribus population effects of the x_j on y—as the parameters of interest.

Assumption MLR.2 (Random Sampling)

We have a random sample of n observations, $\{(x_{i1}, x_{i2}, \ldots, x_{ik}, y_i): i = 1, \ldots, n\}$, following the population model in Assumption MLR.1.

This random sampling assumption means that we have data that can be used to estimate the β_j, and that the data have been chosen to be representative of the population described in Assumption MLR.1.

Assumption MLR.3 (No Perfect Collinearity)

In the sample (and therefore in the population), none of the independent variables is constant, and there are no exact *linear* relationships among the independent variables.

Once we have a sample of data, we need to know that we can use the data to compute the OLS estimates, the $\hat{\beta}_j$ This is the role of Assumption MLR.3: if we have sample variation in each independent variable and no exact linear relationships among the independent variables, we can compute the $\hat{\beta}_j$.

Assumption MLR.4 (Zero Conditional Mean)

The error u has an expected value of zero given any values of the explanatory variables. In other words, $E(u|x_1, x_2, \ldots, x_k) = 0$.

As we discussed in the text, assuming that the unobservables are, on average, unrelated to the explanatory variables is key to deriving the first statistical property of each OLS estimator: its unbiasedness for the corresponding population parameter. Of course, all of the previous assumptions are used to show unbiasedness.

Assumption MLR.5 (Homoskedasticity)

The error u has the same variance given any values of the explanatory variables. In other words,

$$\text{Var}(u|x_1,x_2,\ldots,x_k) = \sigma^2.$$

Compared with Assumption MLR.4, the homoskedasticity assumption is of secondary importance; in particular, Assumption MLR.5 has no bearing on the unbiasedness of the $\hat{\beta}_j$. Still, homoskedasticity has two important implications: (i) We can derive formulas for the sampling variances whose components are easy to characterize; (ii) We can conclude, under the Gauss-Markov assumptions MLR.1 to MLR.5, that the OLS estimators have smallest variance among *all* linear, unbiased estimators.

Assumption MLR.6 (Normality)

The population error u is *independent* of the explanatory variables x_1, x_2, ..., x_k and is normally distributed with zero mean and variance σ^2: $u \sim \text{Normal}(0, \sigma^2)$.

In this chapter, we added Assumption MLR.6 to obtain the exact sampling distributions of t statistics and F statistics, so that we can carry out exact hypotheses tests. In the next chapter, we will see that MLR.6 can be dropped if we have a reasonably large sample size. Assumption MLR.6 does imply a stronger efficiency property of OLS: the OLS estimators have smallest variance among *all* unbiased estimators; the comparison group is no longer restricted to estimators linear in the $\{y_i: i = 1,2,\ldots,n\}$.

KEY TERMS

Alternative Hypothesis
Classical Linear Model
Classical Linear Model (CLM) Assumptions
Confidence Interval (CI)
Critical Value
Denominator Degrees of Freedom
Economic Significance
Exclusion Restrictions
F Statistic
Joint Hypotheses Test
Jointly Insignificant
Jointly Statistically Significant
Minimum Variance Unbiased Estimators
Multiple Hypotheses Test
Multiple Restrictions
Normality Assumption
Null Hypothesis
Numerator Degrees of Freedom
One-Sided Alternative
One-Tailed Test
Overall Significance of the Regression
p-Value
Practical Significance
R-squared Form of the F Statistic
Rejection Rule
Restricted Model
Significance Level
Statistically Insignificant
Statistically Significant
t Ratio
t Statistic
Two-Sided Alternative
Two-Tailed Test
Unrestricted Model

PROBLEMS

4.1 Which of the following can cause the usual OLS t statistics to be invalid (that is, not to have t distributions under H_0)?

(i) Heteroskedasticity.
(ii) A sample correlation coefficient of .95 between two independent variables that are in the model.
(iii) Omitting an important explanatory variable.

4.2 Consider an equation to explain salaries of CEOs in terms of annual firm sales, return on equity (*roe*, in percent form), and return on the firm's stock (*ros*, in percent form):

$$\log(salary) = \beta_0 + \beta_1 \log(sales) + \beta_2 roe + \beta_3 ros + u.$$

(i) In terms of the model parameters, state the null hypothesis that, after controlling for *sales* and *roe*, *ros* has no effect on CEO salary. State the alternative that better stock market performance increases a CEO's salary.
(ii) Using the data in CEOSAL1.RAW, the following equation was obtained by OLS:

$$\widehat{\log(salary)} = \underset{(.32)}{4.32} + \underset{(.035)}{.280} \log(sales) + \underset{(.0041)}{.0174}\, roe + \underset{(.00054)}{.00024}\, ros$$
$$n = 209, R^2 = .283.$$

By what percentage is *salary* predicted to increase if *ros* increases by 50 points? Does *ros* have a practically large effect on *salary*?
(iii) Test the null hypothesis that *ros* has no effect on *salary* against the alternative that *ros* has a positive effect. Carry out the test at the 10% significance level.
(iv) Would you include *ros* in a final model explaining CEO compensation in terms of firm performance? Explain.

4.3 The variable *rdintens* is expenditures on research and development (R&D) as a percentage of sales. Sales are measured in millions of dollars. The variable *profmarg* is profits as a percentage of sales.

Using the data in RDCHEM.RAW for 32 firms in the chemical industry, the following equation is estimated:

$$\widehat{rdintens} = \underset{(1.369)}{.472} + \underset{(.216)}{.321} \log(sales) + \underset{(.046)}{.050}\, profmarg$$
$$n = 32, R^2 = .099.$$

(i) Interpret the coefficient on log(*sales*). In particular, if *sales* increases by 10%, what is the estimated percentage point change in *rdintens*? Is this an economically large effect?
(ii) Test the hypothesis that R&D intensity does not change with *sales* against the alternative that it does increase with sales. Do the test at the 5% and 10% levels.
(iii) Interpret the coefficient on *profmarg*. Is it economically large?
(iv) Does *profmarg* have a statistically significant effect on *rdintens*?

4.4 Are rent rates influenced by the student population in a college town? Let *rent* be the average monthly rent paid on rental units in a college town in the United States. Let *pop* denote the total city population, *avginc* the average city income, and *pctstu* the student population as a percentage of the total population. One model to test for a relationship is

$$\log(rent) = \beta_0 + \beta_1 \log(pop) + \beta_2 \log(avginc) + \beta_3 pctstu + u.$$

(i) State the null hypothesis that size of the student body relative to the population has no ceteris paribus effect on monthly rents. State the alternative that there is an effect.

(ii) What signs do you expect for β_1 and β_2?

(iii) The equation estimated using 1990 data from RENTAL.RAW for 64 college towns is

$$\begin{aligned}\widehat{\log(rent)} = \underset{(.844)}{.043} + \underset{(.039)}{.066} \log(pop) + \underset{(.081)}{.507} \log(avginc) + \underset{(.0017)}{.0056}\, pctstu\\ n = 64,\ R^2 = .458.\end{aligned}$$

What is wrong with the statement: "A 10% increase in population is associated with about a 6.6% increase in rent"?

(iv) Test the hypothesis stated in part (i) at the 1% level.

4.5 Consider the estimated equation from Example 4.3, which can be used to study the effects of skipping class on college GPA:

$$\begin{aligned}\widehat{colGPA} = \underset{(.33)}{1.39} + \underset{(.094)}{.412}\, hsGPA + \underset{(.011)}{.015}\, ACT - \underset{(.026)}{.083}\, skipped\\ n = 141,\ R^2 = .234.\end{aligned}$$

(i) Using the standard normal approximation, find the 95% confidence interval for β_{hsGPA}.

(ii) Can you reject the hypothesis H_0: $\beta_{hsGPA} = .4$ against the two-sided alternative at the 5% level?

(iii) Can you reject the hypothesis H_0: $\beta_{hsGPA} = 1$ against the two-sided alternative at the 5% level?

4.6 In Section 4.5, we used as an example testing the rationality of assessments of housing prices. There, we used a log-log model in *price* and *assess* [see equation (4.47)]. Here, we use a level-level formulation.

(i) In the simple regression model

$$price = \beta_0 + \beta_1 assess + u,$$

the assessment is rational if $\beta_1 = 1$ and $\beta_0 = 0$. The estimated equation is

$$\begin{aligned}\widehat{price} = \underset{(16.27)}{-14.47} + \underset{(.049)}{.976}\, assess\\ n = 88,\ \text{SSR} = 165{,}644.51,\ R^2 = .820.\end{aligned}$$

First, test the hypothesis that H_0: $\beta_0 = 0$ against the two-sided alternative. Then, test H_0: $\beta_1 = 1$ against the two-sided alternative. What do you conclude?

(ii) To test the joint hypothesis that $\beta_0 = 0$ and $\beta_1 = 1$, we need the SSR in the restricted model. This amounts to computing $\sum_{i=1}^{n} (price_i - assess_i)^2$, where $n = 88$, since the residuals in the restricted model are just $price_i - assess_i$. (No estimation is needed for the restricted model because both parameters are specified under H_0.) This turns out to yield SSR = 209,448.99. Carry out the F test for the joint hypothesis.

(iii) Now, test H_0: $\beta_2 = 0$, $\beta_3 = 0$, and $\beta_4 = 0$ in the model

$$price = \beta_0 + \beta_1 assess + \beta_2 lotsize + \beta_3 sqrft + \beta_4 bdrms + u.$$

The R-squared from estimating this model using the same 88 houses is .829.

(iv) If the variance of *price* changes with *assess*, *lotsize*, *sqrft*, or *bdrms*, what can you say about the F test from part (iii)?

4.7 In Example 4.7, we used data on nonunionized manufacturing firms to estimate the relationship between the scrap rate and other firm characteristics. We now look at this example more closely and use all available firms.

(i) The population model estimated in Example 4.7 can be written as

$$\log(scrap) = \beta_0 + \beta_1 hrsemp + \beta_2 \log(sales) + \beta_3 \log(employ) + u.$$

Using the 43 observations available for 1987, the estimated equation is

$$\begin{aligned}\widehat{\log(scrap)} = \underset{(4.57)}{11.74} - \underset{(.019)}{.042}\, hrsemp - \underset{(.370)}{.951} \log(sales) + \underset{(.360)}{.992} \log(employ)\end{aligned}$$
$$n = 43,\ R^2 = .310.$$

Compare this equation to that estimated using only the 29 nonunionized firms in the sample.

(ii) Show that the population model can also be written as

$$\log(scrap) = \beta_0 + \beta_1 hrsemp + \beta_2 \log(sales/employ) + \theta_3 \log(employ) + u,$$

where $\theta_3 = \beta_2 + \beta_3$. [*Hint:* Recall that $\log(x_2/x_3) = \log(x_2) - \log(x_3)$.] Interpret the hypothesis H_0: $\theta_3 = 0$.

(iii) When the equation from part (ii) is estimated, we obtain

$$\widehat{\log(scrap)} = \underset{(4.57)}{11.74} - \underset{(.019)}{.042}\, hrsemp - \underset{(.370)}{.951} \log(sales/employ) + \underset{(.205)}{.041} \log(employ)$$
$$n = 43,\ R^2 = .310.$$

Controlling for worker training and for the sales-to-employee ratio, do bigger firms have larger statistically significant scrap rates?

(iv) Test the hypothesis that a 1% increase in *sales/employ* is associated with a 1% drop in the scrap rate.

4.8 Consider the multiple regression model with three independent variables, under the classical linear model assumptions MLR.1 through MLR.6:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u.$$

You would like to test the null hypothesis H_0: $\beta_1 - 3\beta_2 = 1$.

(i) Let $\hat{\beta}_1$ and $\hat{\beta}_2$ denote the OLS estimators of β_1 and β_2. Find $\mathrm{Var}(\hat{\beta}_1 - 3\hat{\beta}_2)$ in terms of the variances of $\hat{\beta}_1$ and $\hat{\beta}_2$ and the covariance between them. What is the standard error of $\hat{\beta}_1 - 3\hat{\beta}_2$?

(ii) Write the t statistic for testing H_0: $\beta_1 - 3\beta_2 = 1$.

(iii) Define $\theta_1 = \beta_1 - 3\beta_2$ and $\hat{\theta}_1 = \hat{\beta}_1 - 3\hat{\beta}_2$. Write a regression equation involving β_0, θ_1, β_2, and β_3 that allows you to directly obtain $\hat{\theta}_1$ and its standard error.

4.9 In Problem 3.3, we estimated the equation

$$\begin{aligned}\widehat{sleep} = \underset{(112.28)}{3{,}638.25} - \underset{(.017)}{.148}\ totwrk - \underset{(5.88)}{11.13}\ educ + \underset{(1.45)}{2.20}\ age \\ n = 706, R^2 = .113,\end{aligned}$$

where we now report standard errors along with the estimates.

(i) Is either *educ* or *age* individually significant at the 5% level against a two-sided alternative? Show your work.

(ii) Dropping *educ* and *age* from the equation gives

$$\begin{aligned}\widehat{sleep} = \underset{(38.91)}{3{,}586.38} - \underset{(.017)}{.151}\ totwrk \\ n = 706, R^2 = .103.\end{aligned}$$

Are *educ* and *age* jointly significant in the original equation at the 5% level? Justify your answer.

(iii) Does including *educ* and *age* in the model greatly affect the estimated tradeoff between sleeping and working?

(iv) Suppose that the sleep equation contains heteroskedasticity. What does this mean about the tests computed in parts (i) and (ii)?

4.10 Regression analysis can be used to test whether the market efficiently uses information in valuing stocks. For concreteness, let *return* be the total return from holding a firm's stock over the four-year period from the end of 1990 to the end of 1994. The *efficient markets hypothesis* says that these returns should not be systematically related to information known in 1990. If firm characteristics known at the beginning of the period help to predict stock returns, then we could use this information in choosing stocks.

For 1990, let *dkr* be a firm's debt to capital ratio, let *eps* denote the earnings per share, let *netinc* denote net income, and let *salary* denote total compensation for the CEO.

(i) Using the data in RETURN.RAW, the following equation was estimated:

$$\begin{aligned}\widehat{return} = \underset{(6.89)}{-14.37} + \underset{(.201)}{.321}\ dkr + \underset{(.078)}{.043}\ eps - \underset{(.0047)}{.0051}\ netinc + \underset{(.0022)}{.0035}\ salary \\ n = 142, R^2 = .0395.\end{aligned}$$

Test whether the explanatory variables are jointly significant at the 5% level. Is any explanatory variable individually significant?

(ii) Now, reestimate the model using the log form for *netinc* and *salary*:

$$\widehat{return} = \underset{(39.37)}{-36.30} + \underset{(.203)}{.327}\ dkr + \underset{(.080)}{.069}\ eps - \underset{(3.39)}{4.74}\ \log(netinc) + \underset{(6.31)}{7.24}\ \log(salary)$$
$$n = 142,\ R^2 = .0330.$$

Do any of your conclusions from part (i) change?

(iii) How come we do not also use the logs of *dkr* and *eps* in part (ii)?

(iv) Overall, is the evidence for predictability of stock returns strong or weak?

4.11 The following table was created using the data in CEOSAL2.RAW:

Dependent Variable: log(*salary*)			
Independent Variables	**(1)**	**(2)**	**(3)**
log(*sales*)	.224 (.027)	.158 (.040)	.188 (.040)
log(*mktval*)	——	.112 (.050)	.100 (.049)
profmarg	——	−.0023 (.0022)	−.0022 (.0021)
ceoten	——	——	.0171 (.0055)
comten	——	——	−.0092 (.0033)
intercept	4.94 (0.20)	4.62 (0.25)	4.57 (0.25)
Observations *R*-Squared	177 .281	177 .304	177 .353

The variable *mktval* is market value of the firm, *profmarg* is profit as a percentage of sales, *ceoten* is years as CEO with the current company, and *comten* is total years with the company.

(i) Comment on the effect of *profmarg* on CEO salary.

(ii) Does market value have a significant effect? Explain.

(iii) Interpret the coefficients on *ceoten* and *comten*. Are the variables statistically significant?

(iv) What do you make of the fact that longer tenure with the company, holding the other factors fixed, is associated with a lower salary?

COMPUTER EXERCISES

C4.1 The following model can be used to study whether campaign expenditures affect election outcomes:

$$voteA = \beta_0 + \beta_1\log(expendA) + \beta_2\log(expendB) + \beta_3 prtystrA + u,$$

where *voteA* is the percentage of the vote received by Candidate A, *expendA* and *expendB* are campaign expenditures by Candidates A and B, and *prtystrA* is a measure of party strength for Candidate A (the percentage of the most recent presidential vote that went to A's party).

(i) What is the interpretation of β_1?

(ii) In terms of the parameters, state the null hypothesis that a 1% increase in A's expenditures is offset by a 1% increase in B's expenditures.

(iii) Estimate the given model using the data in VOTE1.RAW and report the results in usual form. Do A's expenditures affect the outcome? What about B's expenditures? Can you use these results to test the hypothesis in part (ii)?

(iv) Estimate a model that directly gives the t statistic for testing the hypothesis in part (ii). What do you conclude? (Use a two-sided alternative.)

C4.2 Use the data in LAWSCH85.RAW for this exercise.

(i) Using the same model as Problem 3.4, state and test the null hypothesis that the rank of law schools has no ceteris paribus effect on median starting salary.

(ii) Are features of the incoming class of students—namely, *LSAT* and *GPA*—individually or jointly significant for explaining *salary*? (Be sure to account for missing data on *LSAT* and *GPA*.)

(iii) Test whether the size of the entering class (*clsize*) or the size of the faculty (*faculty*) needs to be added to this equation; carry out a single test. (Be careful to account for missing data on *clsize* and *faculty*.)

(iv) What factors might influence the rank of the law school that are not included in the salary regression?

C4.3 Refer to Problem 3.14. Now, use the log of the housing price as the dependent variable:

$$\log(price) = \beta_0 + \beta_1 sqrft + \beta_2 bdrms + u.$$

(i) You are interested in estimating and obtaining a confidence interval for the percentage change in *price* when a 150-square-foot bedroom is added to a house. In decimal form, this is $\theta_1 = 150\beta_1 + \beta_2$. Use the data in HPRICE1.RAW to estimate θ_1.

(ii) Write β_2 in terms of θ_1 and β_1 and plug this into the log(*price*) equation.

(iii) Use part (ii) to obtain a standard error for $\hat{\theta}_1$ and use this standard error to construct a 95% confidence interval.

C4.4 In Example 4.9, the restricted version of the model can be estimated using all 1,388 observations in the sample. Compute the R-squared from the regression of *bwght* on *cigs*,