The first example is a simple variation of the wage equation introduced in Chapter 2 for obtaining the effect of education on hourly wage:

$$wage = \beta_0 + \beta_1 educ + \beta_2 exper + u, \tag{3.1}$$

where *exper* is years of labor market experience. Thus, *wage* is determined by the two explanatory or independent variables, education and experience, and by other unobserved factors, which are contained in u. We are still primarily interested in the effect of *educ* on *wage*, holding fixed all other factors affecting *wage*; that is, we are interested in the parameter β_1.

Compared with a simple regression analysis relating *wage* to *educ*, equation (3.1) effectively takes *exper* out of the error term and puts it explicitly in the equation. Because *exper* appears in the equation, its coefficient, β_2, measures the ceteris paribus effect of *exper* on *wage*, which is also of some interest.

Not surprisingly, just as with simple regression, we will have to make assumptions about how u in (3.1) is related to the independent variables, *educ* and *exper*. However, as we will see in Section 3.2, there is one thing of which we can be confident: since (3.1) contains experience explicitly, we will be able to measure the effect of education on wage, holding experience fixed. In a simple regression analysis—which puts *exper* in the error term—we would have to assume that experience is uncorrelated with education, a tenuous assumption.

As a second example, consider the problem of explaining the effect of per student spending (*expend*) on the average standardized test score (*avgscore*) at the high school level. Suppose that the average test score depends on funding, average family income (*avginc*), and other unobservables:

$$avgscore = \beta_0 + \beta_1 expend + \beta_2 avginc + u. \tag{3.2}$$

The coefficient of interest for policy purposes is β_1, the ceteris paribus effect of *expend* on *avgscore*. By including *avginc* explicitly in the model, we are able to control for its effect on *avgscore*. This is likely to be important because average family income tends to be correlated with per student spending: spending levels are often determined by both property and local income taxes. In simple regression analysis, *avginc* would be included in the error term, which would likely be correlated with *expend*, causing the OLS estimator of β_1 in the two-variable model to be biased.

In the two previous similar examples, we have shown how observable factors other than the variable of primary interest [*educ* in equation (3.1) and *expend* in equation (3.2)] can be included in a regression model. Generally, we can write a model with two independent variables as

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u, \tag{3.3}$$

where β_0 is the intercept, β_1 measures the change in y with respect to x_1, holding other factors fixed, and β_2 measures the change in y with respect to x_2, holding other factors fixed.

Multiple regression analysis is also useful for generalizing functional relationships between variables. As an example, suppose family consumption (*cons*) is a quadratic function of family income (*inc*):

$$cons = \beta_0 + \beta_1 inc + \beta_2 inc^2 + u, \tag{3.4}$$

where u contains other factors affecting consumption. In this model, consumption depends on only one observed factor, income; so it might seem that it can be handled in a simple regression framework. But the model falls outside simple regression because it contains two functions of income, inc and inc^2 (and therefore three parameters, β_0, β_1, and β_2). Nevertheless, the consumption function is easily written as a regression model with two independent variables by letting $x_1 = inc$ and $x_2 = inc^2$.

Mechanically, there will be *no* difference in using the method of ordinary least squares (introduced in Section 3.2) to estimate equations as different as (3.1) and (3.4). Each equation can be written as (3.3), which is all that matters for computation. There is, however, an important difference in how one *interprets* the parameters. In equation (3.1), β_1 is the ceteris paribus effect of *educ* on *wage*. The parameter β_1 has no such interpretation in (3.4). In other words, it makes no sense to measure the effect of *inc* on *cons* while holding inc^2 fixed, because if *inc* changes, then so must inc^2! Instead, the change in consumption with respect to the change in income—the marginal propensity to consume—is approximated by

$$\frac{\Delta cons}{\Delta inc} \approx \beta_1 + 2\beta_2 inc.$$

See Appendix A for the calculus needed to derive this equation. In other words, the marginal effect of income on consumption depends on β_2 as well as on β_1 and the level of income. This example shows that, in any particular application, the definitions of the independent variables are crucial. But for the theoretical development of multiple regression, we can be vague about such details. We will study examples like this more completely in Chapter 6.

In the model with two independent variables, the key assumption about how u is related to x_1 and x_2 is

$$\mathrm{E}(u|x_1,x_2) = 0. \tag{3.5}$$

The interpretation of condition (3.5) is similar to the interpretation of Assumption SLR.3 for simple regression analysis. It means that, for any values of x_1 and x_2 in the population, the average unobservable is equal to zero. As with simple regression, the important part of the assumption is that the expected value of u is the same for all combinations of x_1 and x_2; that this common value is zero is no assumption at all as long as the intercept β_0 is included in the model (see Section 2.1).

How can we interpret the zero conditional mean assumption in the previous examples? In equation (3.1), the assumption is $\mathrm{E}(u|educ,exper) = 0$. This implies that other factors affecting *wage* are not related on average to *educ* and *exper*. Therefore, if we think innate ability is part of u, then we will need average ability levels to be the same across all combinations of education and experience in the working population. This may or may not be true, but, as we will see in Section 3.3, this is the question we need to ask in order to determine whether the method of ordinary least squares produces unbiased estimators.

The example measuring student performance [equation (3.2)] is similar to the wage equation. The zero conditional mean assumption is $E(u|expend,avginc) = 0$, which means that other factors affecting test scores—school or student characteristics—are, on average, unrelated to per student funding and average family income.

When applied to the quadratic consumption function in (3.4), the zero conditional mean assumption has a slightly different interpretation. Written literally, equation (3.5) becomes $E(u|inc,inc^2) = 0$. Since inc^2 is known when inc is known, including inc^2 in the expectation is redundant: $E(u|inc,inc^2) = 0$ is the same as $E(u|inc) = 0$. Nothing is wrong with putting inc^2 along with inc in the expectation when stating the assumption, but $E(u|inc) = 0$ is more concise.

QUESTION 3.1

A simple model to explain city murder rates (*murdrate*) in terms of the probability of conviction (*prbconv*) and average sentence length (*avgsen*) is

$$murdrate = \beta_0 + \beta_1 prbconv + \beta_2 avgsen + u.$$

What are some factors contained in u? Do you think the key assumption (3.5) is likely to hold?

The Model with *k* Independent Variables

Once we are in the context of multiple regression, there is no need to stop with two independent variables. Multiple regression analysis allows many observed factors to affect y. In the wage example, we might also include amount of job training, years of tenure with the current employer, measures of ability, and even demographic variables like number of siblings or mother's education. In the school funding example, additional variables might include measures of teacher quality and school size.

The general **multiple linear regression model** (also called the multiple regression model) can be written in the population as

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_k x_k + u, \tag{3.6}$$

where β_0 is the **intercept**, β_1 is the parameter associated with x_1, β_2 is the parameter associated with x_2, and so on. Since there are k independent variables and an intercept, equation (3.6) contains $k + 1$ (unknown) population parameters. For shorthand purposes, we will sometimes refer to the parameters other than the intercept as **slope parameters**, even though this is not always literally what they are. [See equation (3.4), where neither β_1 nor β_2 is itself a slope, but together they determine the slope of the relationship between consumption and income.]

The terminology for multiple regression is similar to that for simple regression and is given in Table 3.1. Just as in simple regression, the variable u is the **error term** or **disturbance**. It contains factors other than $x_1, x_2, \ldots, x_k$ that affect y. No matter how many explanatory variables we include in our model, there will always be factors we cannot include, and these are collectively contained in u.

When applying the general multiple regression model, we must know how to interpret the parameters. We will get plenty of practice now and in subsequent chapters, but it is

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2. \tag{3.14}$$

The intercept $\hat{\beta}_0$ in equation (3.14) is the predicted value of y when $x_1 = 0$ and $x_2 = 0$. Sometimes, setting x_1 and x_2 both equal to zero is an interesting scenario; in other cases, it will not make sense. Nevertheless, the intercept is always needed to obtain a prediction of y from the OLS regression line, as (3.14) makes clear.

The estimates $\hat{\beta}_1$ and $\hat{\beta}_2$ have **partial effect**, or **ceteris paribus**, interpretations. From equation (3.14), we have

$$\Delta\hat{y} = \hat{\beta}_1 \Delta x_1 + \hat{\beta}_2 \Delta x_2,$$

so we can obtain the predicted change in y given the changes in x_1 and x_2. (Note how the intercept has nothing to do with the changes in y.) In particular, when x_2 is held fixed, so that $\Delta x_2 = 0$, then

$$\Delta\hat{y} = \hat{\beta}_1 \Delta x_1,$$

holding x_2 fixed. The key point is that, by including x_2 in our model, we obtain a coefficient on x_1 with a ceteris paribus interpretation. This is why multiple regression analysis is so useful. Similarly,

$$\Delta\hat{y} = \hat{\beta}_2 \Delta x_2,$$

holding x_1 fixed.

EXAMPLE 3.1

(Determinants of College GPA)

The variables in GPA1.RAW include college grade point average (*colGPA*), high school GPA (*hsGPA*), and achievement test score (*ACT*) for a sample of 141 students from a large university; both college and high school GPAs are on a four-point scale. We obtain the following OLS regression line to predict college GPA from high school GPA and achievement test score:

$$\widehat{colGPA} = 1.29 + .453\ hsGPA + .0094\ ACT. \tag{3.15}$$

How do we interpret this equation? First, the intercept 1.29 is the predicted college GPA if *hsGPA* and *ACT* are both set as zero. Since no one who attends college has either a zero high school GPA or a zero on the achievement test, the intercept in this equation is not, by itself, meaningful.

More interesting estimates are the slope coefficients on *hsGPA* and *ACT*. As expected, there is a positive partial relationship between *colGPA* and *hsGPA*: holding *ACT* fixed, another point on *hsGPA* is associated with .453 of a point on the college GPA, or almost half a point. In other words, if we choose two students, A and B, and these students have the same ACT score, but the high school GPA of Student A is one point higher than the high school GPA of Student B, then we predict Student A to have a college GPA .453 higher than that of Student B. (This says nothing about any two actual people, but it is our best prediction.)

(W)

The sign on *ACT* implies that, while holding *hsGPA* fixed, a change in the ACT score of 10 points—a very large change, since the average score in the sample is about 24 with a standard deviation less than three—affects *colGPA* by less than one-tenth of a point. This is a small effect, and it suggests that, once high school GPA is accounted for, the ACT score is not a strong predictor of college GPA. (Naturally, there are many other factors that contribute to GPA, but here we focus on statistics available for high school students.) Later, after we discuss statistical inference, we will show that not only is the coefficient on ACT practically small, it is also statistically insignificant.

If we focus on a simple regression analysis relating *colGPA* to *ACT* only, we obtain

$$\widehat{colGPA} = 2.40 + .0271\ ACT;$$

thus, the coefficient on *ACT* is almost three times as large as the estimate in (3.15). But this equation does *not* allow us to compare two people with the same high school GPA; it corresponds to a different experiment. We say more about the differences between multiple and simple regression later.

The case with more than two independent variables is similar. The OLS regression line is

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \ldots + \hat{\beta}_k x_k. \tag{3.16}$$

Written in terms of changes,

$$\Delta\hat{y} = \hat{\beta}_1 \Delta x_1 + \hat{\beta}_2 \Delta x_2 + \ldots + \hat{\beta}_k \Delta x_k. \tag{3.17}$$

The coefficient on x_1 measures the change in $\hat{y}$ due to a one-unit increase in x_1, holding all other independent variables fixed. That is,

$$\Delta\hat{y} = \hat{\beta}_1 \Delta x_1, \tag{3.18}$$

holding $x_2, x_3, \ldots, x_k$ fixed. Thus, we have controlled the variables $x_2, x_3, \ldots, x_k$ when estimating the effect of x_1 on y. The other coefficients have a similar interpretation.

The following is an example with three independent variables.

EXAMPLE 3.2

(Hourly Wage Equation)

Using the 526 observations on workers in WAGE1.RAW, we include *educ* (years of education), *exper* (years of labor market experience), and *tenure* (years with the current employer) in an equation explaining log(*wage*). The estimated equation is

$$\widehat{\log(wage)} = .284 + .092\ educ + .0041\ exper + .022\ tenure. \tag{3.19}$$

As in the simple regression case, the coefficients have a percentage interpretation. The only difference here is that they also have a ceteris paribus interpretation. The coefficient .092

means that, holding *exper* and *tenure* fixed, another year of education is predicted to increase log(*wage*) by .092, which translates into an approximate 9.2 percent [100(.092)] increase in *wage*. Alternatively, if we take two people with the same levels of experience and job tenure, the coefficient on *educ* is the proportionate difference in predicted wage when their education levels differ by one year. This measure of the return to education at least keeps two important productivity factors fixed; whether it is a good estimate of the ceteris paribus return to another year of education requires us to study the statistical properties of OLS (see Section 3.3).

On the Meaning of "Holding Other Factors Fixed" in Multiple Regression

The partial effect interpretation of slope coefficients in multiple regression analysis can cause some confusion, so we attempt to prevent that problem now.

In Example 3.1, we observed that the coefficient on *ACT* measures the predicted difference in *colGPA*, holding *hsGPA* fixed. The power of multiple regression analysis is that it provides this ceteris paribus interpretation even though the data have *not* been collected in a ceteris paribus fashion. In giving the coefficient on *ACT* a partial effect interpretation, it may seem that we actually went out and sampled people with the same high school GPA but possibly with different ACT scores. This is not the case. The data are a random sample from a large university: there were no restrictions placed on the sample values of *hsGPA* or *ACT* in obtaining the data. Rarely do we have the luxury of holding certain variables fixed in obtaining our sample. *If* we could collect a sample of individuals with the same high school GPA, then we could perform a simple regression analysis relating *colGPA* to *ACT*. Multiple regression effectively allows us to mimic this situation without restricting the values of any independent variables.

The power of multiple regression analysis is that it allows us to do in nonexperimental environments what natural scientists are able to do in a controlled laboratory setting: keep other factors fixed.

Changing More than One Independent Variable Simultaneously

Sometimes, we want to change more than one independent variable at the same time to find the resulting effect on the dependent variable. This is easily done using equation (3.17). For example, in equation (3.19), we can obtain the estimated effect on *wage* when an individual stays at the same firm for another year: *exper* (general workforce experience) and *tenure* both increase by one year. The total effect (holding *educ* fixed) is

$$\Delta \widehat{\log(wage)} = .0041\ \Delta exper + .022\ \Delta tenure = .0041 + .022 = .0261,$$

or about 2.6 percent. Since *exper* and *tenure* each increase by one year, we just add the coefficients on *exper* and *tenure* and multiply by 100 to turn the effect into a percent.

OLS Fitted Values and Residuals

After obtaining the OLS regression line (3.11), we can obtain a *fitted* or *predicted value* for each observation. For observation i, the fitted value is simply

EXAMPLE 3.4

(Determinants of College GPA)

From the grade point average regression that we did earlier, the equation with R^2 is

$$\widehat{colGPA} = 1.29 + .453\ hsGPA + .0094\ ACT$$
$$n = 141,\ R^2 = .176.$$

This means that *hsGPA* and *ACT* together explain about 17.6 percent of the variation in college GPA for this sample of students. This may not seem like a high percentage, but we must remember that there are many other factors—including family background, personality, quality of high school education, affinity for college—that contribute to a student's college performance. If *hsGPA* and *ACT* explained almost all of the variation in *colGPA*, then performance in college would be preordained by high school performance!

EXAMPLE 3.5

(Explaining Arrest Records)

CRIME1.RAW contains data on arrests during the year 1986 and other information on 2,725 men born in either 1960 or 1961 in California. Each man in the sample was arrested at least once prior to 1986. The variable *narr86* is the number of times the man was arrested during 1986: it is zero for most men in the sample (72.29 percent), and it varies from 0 to 12. (The percentage of men arrested once during 1986 was 20.51.) The variable *pcnv* is the proportion (not percentage) of arrests prior to 1986 that led to conviction, *avgsen* is average sentence length served for prior convictions (zero for most people), *ptime86* is months spent in prison in 1986, and *qemp86* is the number of quarters during which the man was employed in 1986 (from zero to four).

A linear model explaining arrests is

$$narr86 = \beta_0 + \beta_1 pcnv + \beta_2 avgsen + \beta_3 ptime86 + \beta_4 qemp86 + u,$$

where *pcnv* is a proxy for the likelihood for being convicted of a crime and *avgsen* is a measure of expected severity of punishment, if convicted. The variable *ptime86* captures the incarcerative effects of crime: if an individual is in prison, he cannot be arrested for a crime outside of prison. Labor market opportunities are crudely captured by *qemp86*.

First, we estimate the model without the variable *avgsen*. We obtain

$$\widehat{narr86} = .712 - .150\ pcnv - .034\ ptime86 - .104\ qemp86$$
$$n = 2{,}725,\ R^2 = .0413.$$

This equation says that, as a group, the three variables *pcnv*, *ptime86*, and *qemp86* explain about 4.1 percent of the variation in *narr86*.

Each of the OLS slope coefficients has the anticipated sign. An increase in the proportion of convictions lowers the predicted number of arrests. If we increase *pcnv* by .50 (a large increase in the probability of conviction), then, holding the other factors fixed, $\Delta\widehat{narr86} = -.150(.50) = -.075$. This may seem unusual because an arrest cannot change by a fraction. But we can use this value to obtain the predicted change in expected arrests for a large group of men. For example, among 100 men, the predicted fall in arrests when *pcnv* increases by .50 is −7.5.

Similarly, a longer prison term leads to a lower predicted number of arrests. In fact, if *ptime86* increases from 0 to 12, predicted arrests for a particular man fall by .034(12) = .408. Another quarter in which legal employment is reported lowers predicted arrests by .104, which would be 10.4 arrests among 100 men.

If *avgsen* is added to the model, we know that R^2 will increase. The estimated equation is

$$\widehat{narr86} = .707 - .151\ pcnv + .0074\ avgsen - .037\ ptime86 - .103\ qemp86$$

$$n = 2{,}725,\ R^2 = .0422.$$

Thus, adding the average sentence variable increases R^2 from .0413 to .0422, a practically small effect. The sign of the coefficient on *avgsen* is also unexpected: it says that a longer average sentence length increases criminal activity.

Example 3.5 deserves a final word of caution. The fact that the four explanatory variables included in the second regression explain only about 4.2 percent of the variation in *narr86* does not necessarily mean that the equation is useless. Even though these variables collectively do not explain much of the variation in arrests, it is still possible that the OLS estimates are reliable estimates of the ceteris paribus effects of each independent variable on *narr86*. As we will see, whether this is the case does not directly depend on the size of R^2. Generally, a low R^2 indicates that it is hard to predict individual outcomes on y with much accuracy, something we study in more detail in Chapter 6. In the arrest example, the small R^2 reflects what we already suspect in the social sciences: it is generally very difficult to predict individual behavior.

Regression Through the Origin

Sometimes, an economic theory or common sense suggests that β_0 should be zero, and so we should briefly mention OLS estimation when the intercept is zero. Specifically, we now seek an equation of the form

$$\tilde{y} = \tilde{\beta}_1 x_1 + \tilde{\beta}_2 x_2 + \ldots + \tilde{\beta}_k x_k, \qquad (3.30)$$

where the symbol "~" over the estimates is used to distinguish them from the OLS estimates obtained along with the intercept [as in (3.11)]. In (3.30), when $x_1 = 0$, $x_2 = 0$, …, $x_k = 0$, the predicted value is zero. In this case, $\tilde{\beta}_1, \ldots, \tilde{\beta}_k$ are said to be the OLS estimates from the regression of y on $x_1, x_2, \ldots, x_k$ *through the origin*.

The OLS estimates in (3.30), as always, minimize the sum of squared residuals, but with the intercept set at zero. You should be warned that the properties of OLS that we derived earlier no longer hold for regression through the origin. In particular, the OLS residuals no longer have a zero sample average. Further, if R^2 is defined as $1 - \text{SSR/SST}$, where SST is given in (3.24) and SSR is now $\sum_{i=1}^{n} (y_i - \tilde{\beta}_1 x_{i1} - \ldots - \tilde{\beta}_k x_{ik})^2$, then R^2 can actually be negative. This means that the sample average, $\bar{y}$, "explains" more of the variation in the y_i than the explanatory variables. Either we

W

PROBLEMS

3.1 Using the data in GPA2.RAW on 4,137 college students, the following equation was estimated by OLS:

$$\widehat{colgpa} = 1.392 - .0135\, hsperc + .00148\, sat$$

$$n = 4{,}137,\ R^2 = .273,$$

where *colgpa* is measured on a four-point scale, *hsperc* is the percentile in the high school graduating class (defined so that, for example, *hsperc* = 5 means the *top* five percent of the class), and *sat* is the combined math and verbal scores on the student achievement test.

(i) Why does it make sense for the coefficient on *hsperc* to be negative?

(ii) What is the predicted college GPA when *hsperc* = 20 and *sat* = 1050?

(iii) Suppose that two high school graduates, A and B, graduated in the same percentile from high school, but Student A's SAT score was 140 points higher (about one standard deviation in the sample). What is the predicted difference in college GPA for these two students? Is the difference large?

(iv) Holding *hsperc* fixed, what difference in SAT scores leads to a predicted *colgpa* difference of .50, or one-half of a grade point? Comment on your answer.

3.2 The data in WAGE2.RAW on working men was used to estimate the following equation:

$$\widehat{educ} = 10.36 - .094\, sibs + .131\, meduc + .210\, feduc$$

$$n = 722,\ R^2 = .214,$$

where *educ* is years of schooling, *sibs* is number of siblings, *meduc* is mother's years of schooling, and *feduc* is father's years of schooling.

(i) Does *sibs* have the expected effect? Explain. Holding *meduc* and *feduc* fixed, by how much does *sibs* have to increase to reduce predicted years of education by one year? (A noninteger answer is acceptable here.)

(ii) Discuss the interpretation of the coefficient on *meduc*.

(iii) Suppose that Man A has no siblings, and his mother and father each have 12 years of education. Man B has no siblings, and his mother and father each have 16 years of education. What is the predicted difference in years of education between B and A?

3.3 The following model is a simplified version of the multiple regression model used by Biddle and Hamermesh (1990) to study the tradeoff between time spent sleeping and working and to look at other factors affecting sleep:

$$sleep = \beta_0 + \beta_1 totwrk + \beta_2 educ + \beta_3 age + u,$$

where *sleep* and *totwrk* (total work) are measured in minutes per week and *educ* and *age* are measured in years. (See also Problem 2.12.)

W

(i) If adults trade off sleep for work, what is the sign of β_1?
(ii) What signs do you think β_2 and β_3 will have?
(iii) Using the data in SLEEP75.RAW, the estimated equation is

$$\widehat{sleep} = 3638.25 - .148\ totwrk - 11.13\ educ + 2.20\ age$$
$$n = 706,\ R^2 = .113.$$

If someone works five more hours per week, by how many minutes is *sleep* predicted to fall? Is this a large tradeoff?
(iv) Discuss the sign and magnitude of the estimated coefficient on *educ*.
(v) Would you say *totwrk*, *educ*, and *age* explain much of the variation in *sleep*? What other factors might affect the time spent sleeping? Are these likely to be correlated with *totwrk*?

3.4 The median starting salary for new law school graduates is determined by

$$\log(salary) = \beta_0 + \beta_1 LSAT + \beta_2 GPA + \beta_3 \log(libvol) + \beta_4 \log(cost) + \beta_5 rank + u,$$

where *LSAT* is the median LSAT score for the graduating class, *GPA* is the median college GPA for the class, *libvol* is the number of volumes in the law school library, *cost* is the annual cost of attending law school, and *rank* is a law school ranking (with *rank* = 1 being the best).

(i) Explain why we expect $\beta_5 \leq 0$.
(ii) What signs do you expect for the other slope parameters? Justify your answers.
(iii) Using the data in LAWSCH85.RAW, the estimated equation is

$$\log(\widehat{salary}) = 8.34 + .0047\ LSAT + .248\ GPA + .095\ \log(libvol) + .038\ \log(cost) - .0033\ rank$$
$$n = 136,\ R^2 = .842.$$

What is the predicted ceteris paribus difference in salary for schools with a median GPA different by one point? (Report your answer as a percent.)
(iv) Interpret the coefficient on the variable log(*libvol*).
(v) Would you say it is better to attend a higher ranked law school? How much is a difference in ranking of 20 worth in terms of predicted starting salary?

3.5 In a study relating college grade point average to time spent in various activities, you distribute a survey to several students. The students are asked how many hours they spend each week in four activities: studying, sleeping, working, and leisure. Any activity is put into one of the four categories, so that for each student, the sum of hours in the four activities must be 168.

(i) In the model

$$GPA = \beta_0 + \beta_1 study + \beta_2 sleep + \beta_3 work + \beta_4 leisure + u,$$

does it make sense to hold *sleep*, *work*, and *leisure* fixed, while changing *study*?

(ii) Explain why this model violates Assumption MLR.4.
(iii) How could you reformulate the model so that its parameters have a useful interpretation and it satisfies Assumption MLR.4?

3.6 Consider the multiple regression model containing three independent variables, under Assumptions MLR.1 through MLR.4:

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

You are interested in estimating the sum of the parameters on x_1 and x_2; call this $\theta_1 = \beta_1 + \beta_2$. Show that $\hat{\theta}_1 = \hat{\beta}_1 + \hat{\beta}_2$ is an unbiased estimator of θ_1.

3.7 Which of the following can cause OLS estimators to be biased?
(i) Heteroskedasticity.
(ii) Omitting an important variable.
(iii) A sample correlation coefficient of .95 between two independent variables both included in the model.

3.8 Suppose that average worker productivity at manufacturing firms (*avgprod*) depends on two factors, average hours of training (*avgtrain*) and average worker ability (*avgabil*):

$$avgprod = \beta_0 + \beta_1 avgtrain + \beta_2 avgabil + u.$$

Assume that this equation satisfies the Gauss-Markov assumptions. If grants have been given to firms whose workers have less than average ability, so that *avgtrain* and *avgabil* are negatively correlated, what is the likely bias in $\tilde{\beta}_1$ obtained from the simple regression of *avgprod* on *avgtrain*?

3.9 The following equation describes the median housing price in a community in terms of amount of pollution (*nox* for nitrous oxide) and the average number of rooms in houses in the community (*rooms*):

$$\log(price) = \beta_0 + \beta_1 \log(nox) + \beta_2 rooms + u.$$

(i) What are the probable signs of β_1 and β_2? What is the interpretation of β_1? Explain.
(ii) Why might *nox* [or more precisely, log(*nox*)] and *rooms* be negatively correlated? If this is the case, does the simple regression of log(*price*) on log(*nox*) produce an upward or downward biased estimator of β_1?
(iii) Using the data in HPRICE2.RAW, the following equations were estimated:

$$\log(\widehat{price}) = 11.71 - 1.043 \log(nox),\ n = 506,\ R^2 = .264.$$

$$\log(\widehat{price}) = 9.23 - .718 \log(nox) + .306\ rooms,\ n = 506,\ R^2 = .514.$$

Is the relationship between the simple and multiple regression estimates of the elasticity of *price* with respect to *nox* what you would have predicted, given your answer in part (ii)? Does this mean that $-.718$ is definitely closer to the true elasticity than -1.043?

3.10 Suppose that the population model determining y is

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

and this model satisifies the Gauss-Markov assumptions. However, we estimate the model that omits x_3. Let $\tilde{\beta}_0$, $\tilde{\beta}_1$, and $\tilde{\beta}_2$ be the OLS estimators from the regression of y on x_1 and x_2. Show that the expected value of $\tilde{\beta}_1$ (given the values of the independent variables in the sample) is

$$\mathrm{E}(\tilde{\beta}_1) = \beta_1 + \beta_3 \frac{\sum_{i=1}^{n} \hat{r}_{i1} x_{i3}}{\sum_{i=1}^{n} \hat{r}_{i1}^2},$$

where the $\hat{r}_{i1}$ are the OLS residuals from the regression of x_1 on x_2. [*Hint:* The formula for $\tilde{\beta}_1$ comes from equation (3.22). Plug $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + u_i$ into this equation. After some algebra, take the expectation treating x_{i3} and $\hat{r}_{i1}$ as nonrandom.]

3.11 The following equation represents the effects of tax revenue mix on subsequent employment growth for the population of counties in the United States:

$$growth = \beta_0 + \beta_1 share_P + \beta_2 share_I + \beta_3 share_S + other\ factors,$$

where *growth* is the percentage change in employment from 1980 to 1990, $share_P$ is the share of property taxes in total tax revenue, $share_I$ is the share of income tax revenues, and $share_S$ is the share of sales tax revenues. All of these variables are measured in 1980. The omitted share, $share_F$, includes fees and miscellaneous taxes. By definition, the four shares add up to one. Other factors would include expenditures on education, infrastructure, and so on (all measured in 1980).

(i) Why must we omit one of the tax share variables from the equation?

(ii) Give a careful interpretation of β_1.

3.12 (i) Consider the simple regression model $y = \beta_0 + \beta_1 x + u$ under the first four Gauss-Markov assumptions. For some function $g(x)$, for example $g(x) = x^2$ or $g(x) = \log(1 + x^2)$, define $z_i = g(x_i)$. Define a slope estimator as

$$\tilde{\beta}_1 = \left(\sum_{i=1}^{n} (z_i - \bar{z}) y_i\right) \Big/ \left(\sum_{i=1}^{n} (z_i - \bar{z}) x_i\right).$$

Show that $\tilde{\beta}_1$ is linear and unbiased. Remember, because $\mathrm{E}(u|x) = 0$, you can treat both x_i and z_i as nonrandom in your derivation.

(ii) Add the homoskedasticity assumption, MLR.5. Show that

$$\mathrm{Var}(\tilde{\beta}_1) = \sigma^2 \left(\sum_{i=1}^{n} (z_i - \bar{z})^2\right) \Big/ \left(\sum_{i=1}^{n} (z_i - \bar{z}) x_i\right)^2.$$

(iii) Show directly that, under the Gauss-Markov assumptions, $\mathrm{Var}(\hat{\beta}_1) \le \mathrm{Var}(\tilde{\beta}_1)$, where $\hat{\beta}_1$ is the OLS estimator. [*Hint:* The Cauchy-Schwartz inequality in Appendix B implies that

$$\left(n^{-1} \sum_{i=1}^{n} (z_i - \bar{z})(x_i - \bar{x})\right)^2 \le \left(n^{-1} \sum_{i=1}^{n} (z_i - \bar{z})^2\right)\left(n^{-1} \sum_{i=1}^{n} (x_i - \bar{x})^2\right);$$

notice that we can drop $\bar{x}$ from the sample covariance.]

W

COMPUTER EXERCISES

3.13 A problem of interest to health officials (and others) is to determine the effects of smoking during pregnancy on infant health. One measure of infant health is birth weight; a birth weight that is too low can put an infant at risk for contracting various illnesses. Since factors other than cigarette smoking that affect birth weight are likely to be correlated with smoking, we should take those factors into account. For example, higher income generally results in access to better prenatal care, as well as better nutrition for the mother. An equation that recognizes this is

$$bwght = \beta_0 + \beta_1 cigs + \beta_2 faminc + u.$$

(i) What is the most likely sign for β_2?

(ii) Do you think *cigs* and *faminc* are likely to be correlated? Explain why the correlation might be positive or negative.

(iii) Now, estimate the equation with and without *faminc*, using the data in BWGHT.RAW. Report the results in equation form, including the sample size and R-squared. Discuss your results, focusing on whether adding *faminc* substantially changes the estimated effect of *cigs* on *bwght*.

3.14 Use the data in HPRICE1.RAW to estimate the model

$$price = \beta_0 + \beta_1 sqrft + \beta_2 bdrms + u,$$

where *price* is the house price measured in thousands of dollars.

(i) Write out the results in equation form.

(ii) What is the estimated increase in price for a house with one more bedroom, holding square footage constant?

(iii) What is the estimated increase in price for a house with an additional bedroom that is 140 square feet in size? Compare this to your answer in part (ii).

(iv) What percentage of the variation in price is explained by square footage and number of bedrooms?

(v) The first house in the sample has $sqrft$ = 2,438 and $bdrms$ = 4. Find the predicted selling price for this house from the OLS regression line.

(vi) The actual selling price of the first house in the sample was $300,000 (so $price$ = 300). Find the residual for this house. Does it suggest that the buyer underpaid or overpaid for the house?

3.15 The file CEOSAL2.RAW contains data on 177 chief executive officers, which can be used to examine the effects of firm performance on CEO salary.

(i) Estimate a model relating annual salary to firm sales and market value. Make the model of the constant elasticity variety for both independent variables. Write the results out in equation form.

(ii) Add *profits* to the model from part (i). Why can this variable not be included in logarithmic form? Would you say that these firm performance variables explain most of the variation in CEO salaries?

(iii) Add the variable *ceoten* to the model in part (ii). What is the estimated percentage return for another year of CEO tenure, holding other factors fixed?

(iv) Find the sample correlation coefficient between the variables log(*mktval*) and *profits*. Are these variables highly correlated? What does this say about the OLS estimators?

3.16 Use the data in ATTEND.RAW for this exercise.

(i) Obtain the minimum, maximum, and average values for the variables *atndrte*, *priGPA*, and *ACT*.

(ii) Estimate the model

$$atndrte = \beta_0 + \beta_1 priGPA + \beta_2 ACT + u,$$

and write the results in equation form. Interpret the intercept. Does it have a useful meaning?

(iii) Discuss the estimated slope coefficients. Are there any surprises?

(iv) What is the predicted *atndrte* if $priGPA = 3.65$ and $ACT = 20$? What do you make of this result? Are there any students in the sample with these values of the explanatory variables?

(v) If Student A has $priGPA = 3.1$ and $ACT = 21$ and Student B has $priGPA = 2.1$ and $ACT = 26$, what is the predicted difference in their attendance rates?

3.17 Confirm the partialling out interpretation of the OLS estimates by explicitly doing the partialling out for Example 3.2. This first requires regressing *educ* on *exper* and *tenure* and saving the residuals, $\hat{r}_1$. Then, regress log(*wage*) on $\hat{r}_1$. Compare the coefficient on $\hat{r}_1$ with the coefficient on *educ* in the regression of log(*wage*) on *educ*, *exper*, and *tenure*.

3.18 Use the data set in WAGE2.RAW for this problem. As usual, be sure all of the following regressions contain an intercept.

(i) Run a simple regression of *IQ* on *educ* to obtain the slope coefficient, say $\tilde{\delta}_1$.

(ii) Run the simple regression of log (*wage*) on *educ*, and obtain the slope coefficient, $\tilde{\beta}_1$.

(iii) Run the multiple regression of log (*wage*) on *educ* and *IQ*, and obtain the slope coefficients, $\hat{\beta}_1 + \hat{\beta}_2$, respectively.

(iv) Verify that $\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_2\tilde{\delta}_1$

APPENDIX 3A

3A.1 Derivation of the First Order Conditions in Equation (3.13)

The analysis is very similar to the simple regression case. We must characterize the solutions to the problem

$$\min_{b_0, b_1, \ldots, b_k} \sum_{i=1}^{n} (y_i - b_0 - b_1x_{i1} - \ldots - b_kx_{ik})^2.$$

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 on the preceding page.)

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

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

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

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

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.

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.

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?

EXAMPLE 4.2

(Student Performance and School Size)

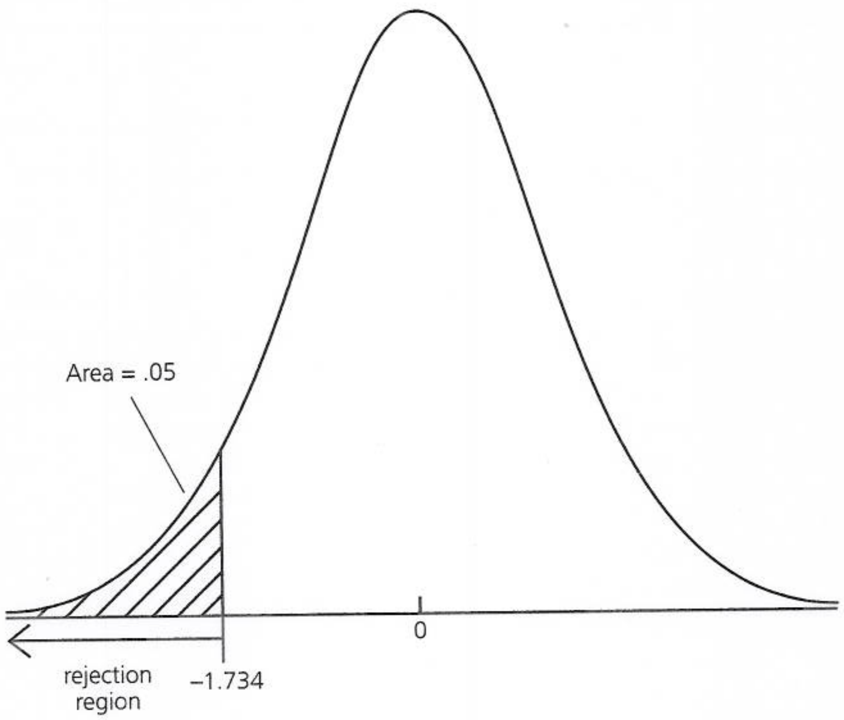
There is much interest in the effect of school size on student performance. (See, for example, *The New York Times Magazine*, 5/28/95.) One claim is that, everything else being equal, students at smaller schools fare better than those at larger schools. This hypothesis is assumed to be true even after 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 (*math10*). 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.

W

Figure 4.3

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



The estimated equation, with standard errors in parentheses, is

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

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.

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

$$\widehat{colGPA} = \underset{(0.33)}{1.39} + \underset{(.094)}{.412}\ hsGPA + \underset{(.011)}{.015}\ ACT - \underset{(.026)}{.083}\ skipped$$
$$n = 141,\ R^2 = .234.$$

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 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): $t = (\hat{\beta}_j + 1)/\text{se}(\hat{\beta}_j)$ (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 $t = (\hat{\beta}_1 + 1)/\text{se}(\hat{\beta}_1)$.

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

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

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.

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

Since 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

W

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

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, while firm size does affect the participation rate, the effect is not practically very large.

of the sample → t ↑ ($\hat{\beta}$/se($\hat{\beta}$))↑

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 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 Grants on Firm Scrap Rates)

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

We can use the scrap rate to measure the effect of worker training on productivity. For a sample of Michigan manufacturing firms in 1987, the following equation is estimated:

$$\log(\widehat{scrap}) = \underset{(4.91)}{13.72} - \underset{(.019)}{.028}\ hrsemp - \underset{(0.41)}{1.21}\ \log(sales) + \underset{(0.43)}{1.48}\ \log(employ)$$
$$n = 30,\ R^2 = .431.$$

(This regression uses a subset of the data in JTRAIN.RAW.) The variable *hrsemp* is annual hours of training per employee, *sales* is annual firm sales (in dollars), and *employ* is number of firm employees. The average scrap rate in the sample is about 3.5, and the average *hrsemp* is about 7.3.

The main variable of interest is *hrsemp*. One more hour of training per employee lowers log(*scrap*) by .028, which means the scrap rate is about 2.8% 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(2.8) = 14%. 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 $-.028/.019 = -1.47$, 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 $30 - 4 = 26$ 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 significant against the one-sided alternative at the 10% level. The p-value is easily computed as $P(T_{26} < -1.47) = .077$. 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 certain independent variables from the model. 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.

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

(Hedonic Price Model for Houses)

A model that explains the price of a good in terms of the good's characteristics is called an *hedonic price model*. The following equation is an hedonic price model for housing prices; the characteristics are square footage (*sqrft*), number of bedrooms (*bdrms*), and number of bathrooms (*bthrms*). Often, *price* appears in logarithmic form, as do some of the explanatory variables. Using $n = 19$ observations on houses that were sold in Waltham, Massachusetts, in 1990, the estimated equation (with standard errors in parentheses below the coefficient estimates) is

$$\begin{aligned}\log(\widehat{price}) = \underset{(1.15)}{7.46} + \underset{(.184)}{.634}\ \log(sqrft) - \underset{(.059)}{.066}\ bdrms + \underset{(.075)}{.158}\ bthrms\\ n = 19, R^2 = .806.\end{aligned}$$

Since *price* and *sqrft* both appear in logarithmic form, the price elasticity with respect to square footage is .634, so that, holding number of bedrooms and bathrooms fixed, a 1% increase in square footage increases the predicted housing price by about .634%. We can construct a 95% confidence interval for the population elasticity using the fact that the estimated model has $n - k - 1 = 19 - 3 - 1 = 15$ degrees of freedom. From Table G.2, we find the 97.5th percentile in the t_{15} distribution: $c = 2.131$. Thus, the 95% confidence interval for $\beta_{\log(sqrft)}$ is $.634 \pm 2.131(.184)$, or $(.242, 1.026)$. Since zero is excluded from this confidence interval, we reject H_0: $\beta_{\log(sqrft)} = 0$ against the two-sided alternative at the 5% level.

The coefficient on *bdrms* is negative, which seems counterintuitive. However, it is important to remember the ceteris paribus nature of this coefficient: it measures the effect of another bedroom, holding size of the house and number of bathrooms fixed. If two houses are the same size but one has more bedrooms, then the house with more bedrooms has smaller bedrooms; more bedrooms that are smaller is not necessarily a good thing. In any case, we can see that the 95% confidence interval for β_{bdrms} is fairly wide, and it contains the value zero: $-.066 \pm 2.131(.059)$ or $(-.192, .060)$. Thus, *bdrms* does not have a statistically significant ceteris paribus effect on housing price.

Given size and number of bedrooms, one more bathroom is predicted to increase housing price by about 15.8%. (Remember that we must multiply the coefficient on *bthrms* by 100 to turn the effect into a percent.) The 95% confidence interval for β_{bthrms} is $(-.002, .318)$. In this case, zero is barely in the confidence interval, so technically speaking $\hat{\beta}_{bthrms}$ is not statistically significant at the 5% level against a two-sided alternative. Since it is very close to being significant, we would probably conclude that number of bathrooms has an effect on log(*price*).

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(*price*) 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 and *univ* is number of years at a four-year college. 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

$$\mathrm{H}_0\colon \beta_1 = \beta_2. \tag{4.18}$$

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

$$\mathrm{H}_1: \beta_1 < \beta_2. \tag{4.19}$$

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 peformed the OLS regression. Using the data in TWOYEAR.RAW, which comes from Kane and Rouse (1995), we estimate equation (4.17):

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

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

w

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

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}.

We suggest another route that is much simpler to compute, less likely to lead to an error, and readily applied to a variety of problems. 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. Since $\theta_1 = \beta_1 - \beta_2$, 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, $\text{se}(\hat{\theta}_1)$ 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

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

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:

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

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

$$\begin{aligned}
\log(\widehat{salary}) = &\ \underset{(0.29)}{11.10} + \underset{(.0121)}{.0689}\, years + \underset{(.0026)}{.0126}\, gamesyr \\
&+ \underset{(.00110)}{.00098}\, bavg + \underset{(.0161)}{.0144}\, hrunsyr + \underset{(.0072)}{.0108}\, rbisyr \\
n = 353,\ &\text{SSR} = 183.186,\ R^2 = .6278,
\end{aligned} \tag{4.31}$$

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, while *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

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

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.

Since 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

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

for the OLS estimators that are derived using calculus. The formulas for the OLS estimators in the multiple regression model are similar to those in Key Concept 4.2 for the single-regressor model. These formulas are incorporated into modern statistical software. In the multiple regression model, the formulas are best expressed and discussed using matrix notation, so their presentation is deferred to Section 16.1.

The definitions and terminology of OLS in multiple regression are summarized in Key Concept 5.3.

Application to Test Scores and the Student-Teacher Ratio

In Section 4.2, we used OLS to estimate the intercept and slope coefficient of the regression relating test scores (*TestScore*) to the student-teacher ratio (*STR*), using our 420 observations for California school districts; the estimated OLS regression line, reported in Equation (4.7), is

$$\widehat{TestScore} = 698.9 - 2.28 \times STR. \tag{5.9}$$

Our concern has been that this relationship is misleading because the student-teacher ratio might be picking up the effect of having many English learners in districts with large classes. That is, it is possible that the OLS estimator is subject to omitted variable bias.

We are now in a position to address this concern by using OLS to estimate a multiple regression in which the dependent variable is the test score (Y_i) and there are two regressors: the student-teacher ratio (X_{1i}) and the percentage of English learners in the school district (X_{2i}) for our 420 districts ($i = 1, \dots, 420$). The estimated OLS regression line for this multiple regression is

$$\widehat{TestScore} = 686.0 - 1.10 \times STR - 0.65 \times PctEL, \tag{5.10}$$

where *PctEL* is the percentage of students in the district who are English learners. The OLS estimate of the intercept ($\hat{\beta}_0$) is 686.0, the OLS estimate of the coefficient on the student-teacher ratio ($\hat{\beta}_1$) is −1.10, and the OLS estimate of the coefficient on the percentage English learners ($\hat{\beta}_2$) is −0.65.

The estimated effect on test scores of a change in the student-teacher ratio in the multiple regression is approximately half as large as when the student-teacher ratio is the only regressor: in the single-regressor equation (Equation (5.9)), a unit decrease in the *STR* is estimated to increase test scores by 2.28 points, but in the multiple regression equation (Equation (5.10)), it is estimated

STW

Testing the Hypothesis $\beta_j = \beta_{j,0}$ Against the Alternative $\beta_j \neq \beta_{j,0}$

Key Concept 5.6

1. Compute the standard error of $\hat{\beta}_j$, $SE(\hat{\beta}_j)$.
2. Compute the t-statistic,

$$t = \frac{\hat{\beta}_j - \beta_{j,0}}{SE(\hat{\beta}_j)}. \tag{5.14}$$

3. Compute the p-value,

$$p\text{-value} = 2\Phi(-|t^{act}|), \tag{5.15}$$

where t^{act} is the value of the t-statistic actually computed. Reject the hypothesis at the 5% significance level if the p-value is less than 0.05 or, equivalently, if $|t^{act}| > 1.96$.

The standard error and (typically) the t-statistic and p-value testing $\beta_j = 0$ are computed automatically by regression software.

420 observations

Application to Test Scores and the Student-Teacher Ratio

Can we reject the null hypothesis that a change in the student-teacher ratio has no effect on test scores, once we control for the percentage of English learners in the district? What is a 95% confidence interval for the effect on test scores of a change in the student-teacher ratio, controlling for the percentage of English learners? We are now able to find out. The regression of test scores against STR and $PctEL$, estimated by OLS, was given in Equation (5.10) and is restated here with standard errors in parentheses below the coefficients:

$$\widehat{TestScore} = \underset{(8.7)}{686.0} - \underset{(0.43)}{1.10} \times STR - \underset{(0.031)}{0.650} \times PctEL. \tag{5.16}$$

To test the hypothesis that the true coefficient on STR is 0, we first need to compute the t-statistic in Equation (5.14). Because the null hypothesis says that the true value of this coefficient is zero, the t-statistic is $t = (-1.10 - 0)/0.43 = -2.54$. The associated p-value is $2\Phi(-2.54) = 1.1\%$; that is, the smallest signifi-

Key Concept 5.7

Confidence Intervals for a Single Coefficient in Multiple Regression

A 95% two-sided confidence interval for the coefficient β_j is an interval that contains the true value of β_j with a 95% probability; that is, it contains the true value of β_j in 95% of all possible randomly drawn samples. Equivalently, it is also the set of values of β_j that cannot be rejected by a 5% two-sided hypothesis test. When the sample size is large, the 95% confidence interval is:

$$95\% \text{ confidence interval for } \beta_j = (\hat{\beta}_j - 1.96SE(\hat{\beta}_j), \hat{\beta}_j + 1.96SE(\hat{\beta}_j)). \quad (5.17)$$

A 90% confidence interval is obtained by replacing 1.96 in Equation (5.17) with 1.645.

cance level at which we can reject the null hypothesis is 1.1%. Because the p-value is less than 5%, the null hypothesis can be rejected at the 5% significance level (but not quite at the 1% significance level).

A 95% confidence interval for the population coefficient on STR is $-1.10 \pm 1.96 \times 0.43 = (-1.95, -0.26)$; that is, we can be 95% confident that the true value of the coefficient is between -1.95 and -0.26. Interpreted in the context of the superintendent's interest in decreasing the student-teacher ratio by 2, the 95% confidence interval for the effect on test scores of this reduction is $(-1.95 \times 2, -0.26 \times 2) = (-3.90, -0.52)$.

Adding expenditures per pupil to the equation Your analysis of the multiple regression in Equation (5.16) has persuaded the superintendent that, based on the evidence so far, reducing class size will help test scores in her district. Now, however, she moves on to a more nuanced question. If she is to hire more teachers, she can pay for those teachers either through cuts elsewhere in the budget (no new computers, reduced maintenance, etc.), or by asking for an increase in her budget, which taxpayers do not favor. What, she asks, is the effect on test scores of reducing the student-teacher ratio, holding expenditures per pupil (and the percentage of English learners) constant?

This question can be addressed by estimating a regression of test scores on the student-teacher ratio, total spending per pupil, and the percentage of English learners. The OLS regression line is

JTW

$$\widehat{TestScore} = \underset{(15.5)}{649.6} - \underset{(0.48)}{0.29} \times STR + \underset{(1.59)}{3.87} \times Expn - \underset{(0.032)}{0.656} \times PctEL, \tag{5.18}$$

where *Expn* is total annual expenditures per pupil in the district in thousands of dollars.

The result is striking. Holding expenditures per pupil and the percentage of English learners constant, changing the student-teacher ratio is estimated to have a very small effect on test scores: the estimated coefficient on *STR* is −1.10 in Equation (5.16) but, after adding *Expn* as a regressor in Equation (5.18), it is only −0.29. Moreover, the t-statistic for testing that the true value of the coefficient is zero is now $t = (-0.29 - 0)/0.48 = -0.60$, so the hypothesis that the population value of this coefficient is indeed zero cannot be rejected even at the 10% significance level ($|-0.60| < 1.645$). Thus Equation (5.18) provides no evidence that hiring more teachers improves test scores if overall expenditures per pupil are held constant.

Note that the standard error on *STR* increased when *Expn* was added, from 0.43 in Equation (5.16) to 0.48 in Equation (5.18). This illustrates the general point that a correlation between regressors (the correlation between *STR* and *Expn* is −0.62) can make the OLS estimators less precise (see Appendix 5.2 for further discussion).

What about our angry taxpayer? He asserts that the population values of *both* the coefficient on the student-teacher ratio (β_1) *and* the coefficient on spending per pupil (β_2) are zero, that is, he hypothesizes that both $\beta_1 = 0$ and $\beta_2 = 0$. Although it might seem that we can reject this hypothesis because the t-statistic testing $\beta_2 = 0$ in Equation (5.18) is $t = 3.87/1.59 = 2.43$, this reasoning is flawed. The taxpayer's hypothesis is a joint hypothesis, and to test it we need a new tool, the F-statistic.

5.7 Tests of Joint Hypotheses

This section describes how to formulate joint hypotheses on multiple regression coefficients and how to test them using an F-statistic.

Testing Hypotheses on Two or More Coefficients

Joint null hypotheses. Consider the regression in Equation (5.18) of the test score against the student-teacher ratio, expenditures per pupil, and the percentage

5.5 Provide an example of a regression that arguably would have a high value of R^2 but would produce biased and inconsistent estimators of the regression coefficient(s). Explain why the R^2 is likely to be high. Explain why the OLS estimators would be biased and inconsistent.

Exercises

The first seven exercises refer to the table of estimated regressions below, computed using data for 1998 from the CPS. The data set consists of information on 4,000 full-time full-year workers. The highest educational achievement for

Results of Regressions of Average Hourly Earnings on Gender and Education Binary Variables and other characteristics using 1998 data from the Current Population Survey

Dependent variable: Average Hourly Earnings (AHE).

Regressor	(1)	(2)	(3)
College (X_1)	5.46 (0.21)	5.48 (0.21)	5.44 (0.21)
Female (X_2)	–2.64 (0.20)	–2.62 (0.20)	–2.62 (0.20)
Age (X_3)		0.29 (0.04)	0.29 (0.04)
Northeast (X_4)			0.69 (0.30)
Midwest (X_5)			0.60 (0.28)
South (X_6)			–0.27 (0.26)
Intercept	12.69 (0.14)	4.40 (1.05)	3.75 (1.06)
Summary Statistics and Joint Tests			
F-Statistic for regional effects = 0			6.10
SER	6.27	6.22	6.21
R^2	0.176	0.190	0.194
$\overline{R}^2$			
n	4000.0	4000.0	4000.0

5tw

each worker was either a high school diploma or a bachelor's degree. The worker's ages ranged from 25–34 years. The data set also contained information on the region of the country where the person lived, marital status, and number of children. For the purposes of this exercises let

AHE = average hourly earnings (in 1998 dollars)

College = binary variable (1 if college, 0 if high school)

Female = binary variable (1 if female, 0 if male)

Age = age (in years)

Ntheast = binary variable (1 if Region = Northeast, 0 otherwise)

Midwest = binary variable (1 if Region = Midwest, 0 otherwise)

South = binary variable (1 if Region = South, 0 otherwise)

West = binary variable (1 if Region = West, 0 otherwise)

5.1 Add "*" (5%) and "**" (1%) to the table to indicate statistical significance of the coefficients.

5.2 Compute $\overline{R}^2$ for each of the regressions.

5.3 Using the regression results in column (1):

*__a.__ Do workers with college degrees earn more, on average, than workers with only high school degrees? How much more? Is the earnings difference estimated from this regression statistically significant at the 5% level?

b. Do men earn more than women on average? How much more? Is the earnings difference estimated from this regression statistically significant at the 5% level?

5.4 Using the regression results in column (2):

a. Is age an important determinant of earnings? Explain.

b. Sally is 29-year-old female college graduate. Betsy is a 34-year-old female college graduate. Predict Sally's and Betsy's earnings and construct a 95% confidence interval for the expected difference between their earnings.

5.5 Using the regression results in column (3):

*__a.__ Do there appear to be important regional differences?

b. Why is the regressor *West* omitted from the regression? What would happen if it was included?

***c.** Juanita is a 28-year-old female college graduate from the South. Molly is a 28-year-old female college graduate from the West. Jennifer is a 28-year-old female college graduate from the Midwest.

ci. Construct a 95% confidence interval for the difference in expected earnings between Juanita and Molly.

cii. Calculate the expected difference in earnings between Juanita and Jennifer.

ciii. Explain how you would construct a 95% confidence interval for the difference in expected earnings between Juanita and Jennifer. (*Hint:* What would happen if you included *West* and excluded *Midwest* from the regression?)

5.6 The regression shown in column (2) was estimated again, this time using data from 1992. (4,000 observations selected at random from the March 1993 CPS, converted into 1998 dollars using the consumer price index.) The results are

$$\widehat{AHE} = \underset{(0.98)}{0.77} + \underset{(0.20)}{5.29}College - \underset{(0.18)}{2.59}Female + \underset{(0.03)}{0.40}Age,\ SER = 5.85,\ \overline{R}^2 = 0.21$$

Comparing this regression to the regression for 1998 shown in column (2), was there a statistically significant change in the coefficient on *College*?

***5.7** Evaluate the following statement: "In all of the regressions, the coefficient on *Female* is negative, large, and statistically significant. This provides strong statistical evidence of gender discrimination in the U.S. labor market."

5.8 Consider the regression model $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$. Use "Approach #2" from Section 5.8 to transform the regression so that you can use a *t*-statistic to test

a. $\beta_1 = \beta_2$;

b. $\beta_1 + a\beta_2 = 0$, where a is a constant;

c. $\beta_1 + \beta_2 = 1$. (*Hint:* You must redefine the dependent variable in the regression.)

5.9 Appendix 5.3 shows two formulas for the rule-of-thumb *F*-statistic, Equations (5.38) and (5.39). Show that the two formulas are equivalent.

5tw

sample, the median district income is 13.7 (that is, \$13,700 per person), and it ranges from 5.3 (\$5,300 per person) to 55.3 (\$55,300 per person).

Figure 6.2 shows a scatterplot of fifth-grade test scores against district income for the California data set, along with the OLS regression line relating these two variables. Test scores and average income are strongly positively correlated, with a correlation coefficient of 0.71; students from affluent districts do better on the tests than students from poor districts. But this scatterplot has a peculiarity: most of the points are below the OLS line when income is very low (under \$10,000) or very high (over \$40,000), but are above the line when income is between \$15,000 and \$30,000. There seems to be some curvature in the relationship between test scores and income that is not captured by the linear regression.

In short, it seems that the relationship between district income and test scores is not a straight line. Rather, it is nonlinear. A nonlinear function is a function with a slope that is not constant: the function $f(X)$ is linear if the slope of $f(X)$ is the same for all values of X, but if the slope depends on the value of X, then $f(X)$ is nonlinear.

If a straight line is not an adequate description of the relationship between district income and test scores, what is? Imagine drawing a curve that fits the points in Figure 6.2. This curve would be steep for low values of district income, then would flatten out as district income gets higher. One way to approximate such a curve mathematically is to model the relationship as a quadratic function. That is, we could model test scores as a function of income *and* the square of income.

A quadratic population regression model relating test scores and income is written mathematically as

$$TestScore_i = \beta_0 + \beta_1 Income_i + \beta_2 Income_i^2 + u_i, \tag{6.1}$$

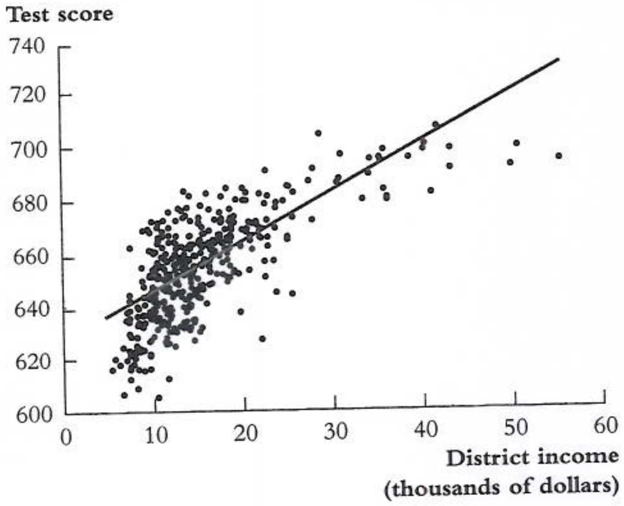
where β_0, β_1, and β_2 are coefficients, $Income_i$ is the income in the i^{th} district, $Income_i^2$, is the square of income in the i^{th} district, and u_i is an error term that, as usual, represents all the other factors that determine test scores. Equation (6.1) is called the **quadratic regression model** because the population regression function, $E(TestScore_i \mid Income_i) = \beta_0 + \beta_1 Income_i + \beta_2 Income_i^2$, is a quadratic function of the independent variable, *Income*.

If you knew the population coefficients β_0, β_1, and β_2 in Equation (6.1), you could predict the test score of a district based on its average income. But these population coefficients are unknown and therefore must be estimated using a sample of data.

At first, it might seem difficult to find the coefficients of the quadratic function that best fits the data in Figure 6.2. If you compare Equation (6.1) with the

FIGURE 6.2 Scatterplot of Test Score vs. District Income with a Linear OLS Regression Function

There is a positive correlation between test scores and district income (correlation = 0.71), but the linear OLS regression line does not adequately describe the relationship between these variables.



multiple regression model in Key Concept 5.2, however, you will see that Equation (6.1) is in fact a version of the multiple regression model with two regressors: the first regressor is *Income*, and the second regressor is $Income^2$. Thus, after defining the regressors as *Income* and $Income^2$, the nonlinear model in Equation (6.1) is simply a multiple regression model with two regressors!

Because the quadratic regression model is a variant of multiple regression, its unknown population coefficients can be estimated and tested using the OLS methods described in Chapter 5. Estimating the coefficients of Equation (6.1) using OLS for the 420 observations in Figure 6.2 yields

$$\widehat{TestScore} = \underset{(2.9)}{607.3} + \underset{(0.27)}{3.85}Income - \underset{(0.0048)}{0.0423}Income^2,\ \overline{R}^2 = 0.554, \qquad (6.2)$$

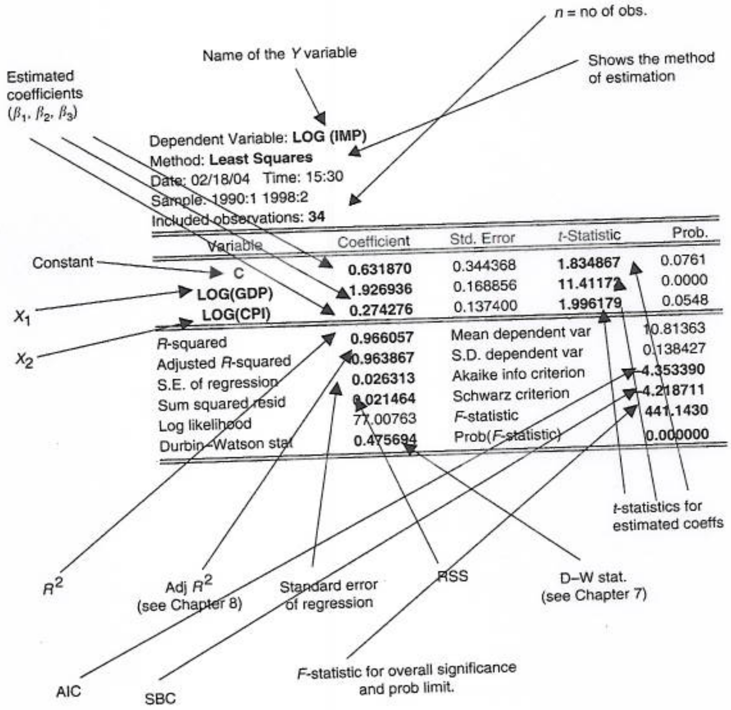
where (as usual) standard errors of the estimated coefficients are given in parentheses. The estimated regression function (6.2) is plotted in Figure 6.3, superimposed over the scatterplot of the data. The quadratic function captures the curvature in the scatterplot: it is steep for low values of district income but flattens out when district income is high. In short, the quadratic regression function seems to fit the data better than the linear one.

We can go one step beyond this visual comparison and formally test the hypothesis that the relationship between income and test scores is linear, against

Below we show an example of a regression result output from EViews (the case of Microfit is similar).

Reading the EViews multiple regression results output

IMP: imports



Hypothesis testing

Testing individual coefficients

As in simple regression analysis, in multiple regression a single test of hypothesis on a regression coefficient is carried out as a normal t test. We can again have one-tail tests (if there is some prior belief/theory for the sign of the coefficient) or two-tail tests, carried out in the usual way ($(\hat{\beta} - \beta)/s_{\hat{\beta}}$ follows t_{n-k}), and we can immediately make a decision about the significance or not of the $\hat{\beta}$s using the criterion $|t\text{-}stat| > |t\text{-}crit|$ having the t-statistic provided immediately by either Microfit or EViews (note that especially for large samples we can use the 'rule of thumb' $|t\text{-}stat| > 2$).

Testing linear restrictions

Sometimes in economics we need to test whether there are particular relationships between the estimated coefficients. Take for example a production function of the

First we need to construct/generate the dependent variable. In order to do that we have to type the following command in the EViews command line:

```
genr lnwage = log(wage)
```

Then, in order to estimate the multiple regression model, we have to select from the EViews toolbar **Quick/Estimate Equation** and type into the **Equation Specification** box the required model as:

```
lnwage c educ exper tenure
```

The results from this equation are shown in Table 5.1.

We can also save the equation (named **unrestrict01**) and save the regression results (by clicking on the '**freeze**' button) at an output table (named **Table01** in the file). As may be seen from the equation, all three variables have positive coefficients. These are all above the 'rule of thumb' critical t-value of 2, hence all are significant. So, it may be said that wages will increase as education, experience and tenure increases. Despite the significance of these three variables, the adjusted R^2 is quite low (0.145) as there are probably other variables that affect wages.

A Wald test of coefficient restrictions

Let's now assume that we want to test whether the effect of the *tenure* variable is the same with that of experience (*exper* variable). Referring to the estimation equation, we can see that the coefficient of *exper* is $C(3)$ and the coefficient of *tenure* is $C(4)$.

In order to test the hypothesis that the two effects are equal we need to conduct a Wald test in EViews. This can be done by clicking on **View/Coefficient Tests/Wald-Coefficient Restrictions**, in the regression results output and then by typing the restriction as:

$$C(3) = C(4) \tag{5.77}$$

Table 5.1 Results from the wage equation

Dependent Variable: LNWAGE
Method: Least Squares
Date: 02/02/04 Time: 11:10
Sample: 1 900
Included observations: 900

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.528329	0.112795	49.01237	0.0000
EDUC	0.073117	0.006636	11.01871	0.0000
EXPER	0.015358	0.003425	4.483631	0.0000
TENURE	0.012964	0.002631	4.927939	0.0000

R-squared	0.148647	Mean dependent var	6.786164
Adjusted R-squared	0.145797	S.D. dependent var	0.420312
S.E. of regression	0.388465	Akaike info criterion	0.951208
Sum squared resid	135.2110	Schwarz criterion	0.972552
Log likelihood	−424.0434	F-statistic	52.14758
Durbin–Watson stat	1.750376	Prob(F-statistic)	0.000000

(c) Estimate a regression that includes only *cig* and comment on your results.

(d) Present all three regressions summarized in a table and comment on your results, especially by comparing the changes in the estimated effects and the R^2 of the three different models. What does the F statistic suggest about the joint significance of the explanatory variables in the multiple regression case?

(e) Test the hypothesis that the effect of *cig* is two times bigger than the respective effect of *fam_inc* using the Wald test.

Exercise 5.2

Use the data from the file wage.wf1 and estimate an equation which includes as determinants of the logarithm of the wage rate the variables, *educ*, *exper* and *tenure*.

(a) Comment on your results.

(b) Conduct a test of whether another year of general workforce experience (captured by exper) has the same effect on log(wage) as another year of education (captured by educ). State clearly your null and alternative hypotheses and your restricted and unrestricted models. Use the Wald test to check for that hypothesis.

(c) Conduct a redundant variable test for the explanatory variable exper. Comment on your results.

(d) Estimate a model with exper and educ only and then conduct an omitted variable test for tenure in the model. Comment on your results.

Exercise 5.3

Use the data in the file money_uk.wf1 to estimate the parameters α, β and γ, in the equation below:

$$\ln(M/P)_t = a + \beta \ln Y_t + \gamma \ln R_t + u_t$$

(a) Briefly outline the theory behind the aggregate demand for money. Relate your discussion to the specification of the equation given above. In particular explain first the meaning of the dependent variable and then the interpretation of β and γ.

(b) Perform appropriate tests of significance on the estimated parameters in order to investigate each of the following propositions: (i) that the demand for money increases with the level of real income, (ii) the demand for money is income-elastic, and (iii) the demand for money is inversely related to the rate of interest.

Exercise 5.4

The file living.xls contains data for a variety of economic and social measures for a sample of 20 different countries, where:

Y = GNP per capita, 1984 \$US;
X2 = average % annual inflation rate (1973–84);

Atn

X3 = % of labour force in agriculture;
X4 = life expectancy at birth, 1984 (years);
X5 = number enrolled in secondary education as % of age group.

(a) Insert the data in EViews or Microfit.

(b) Estimate the regression coefficients in each of the following equations:

$$Y_t = \beta_1 + \beta_2 X_{2t} + u_t$$
$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$$
$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + u_t$$
$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \beta_5 X_{5t} + u_t$$

(c) How robust are the estimated coefficients? By this we mean, to what extent do the estimated values of each β_i change as further explanatory variables are added to the right-hand side of the equation?

(d) Assuming Y to be an index of economic development, carry out tests of significance on all slope coefficients in the final regression equation model. State clearly the null and alternative hypotheses for each case and give reasons for setting them like that.

Exercise 5.5

The file Cobb_Douglas_us.wf1 contains data for output (Y), labour (L) and stock of capital (K) for the United States. Estimate a Cobb–Douglas type regression equation and check for constant returns to scale using the Wald test.

A+M

imp-rtb

Table 6.8 Second model regression results (including both *CPI* and *PPI*)

Dependent Variable: LOG(IMP)
Method: Least Squares
Date: 02/17/04 Time: 02:19
Sample: 1990:1 1998:2
Included observations: 34

Variable	*Coefficient*	*Std. Error*	*t-Statistic*	*Prob.*
C	0.213906	0.358425	0.596795	0.5551
LOG(GDP)	1.969713	0.156800	12.56198	0.0000
LOG(CPI)	1.025473	0.323427	3.170645	0.0035
LOG(PPI)	−0.770644	0.305218	−2.524894	0.0171

R-squared	0.972006	Mean dependent var	10.81363
Adjusted *R*-squared	0.969206	S.D. dependent var	0.138427
S.E. of regression	0.024291	Akaike info criterion	−4.487253
Sum squared resid	0.017702	Schwarz criterion	−4.307682
Log likelihood	80.28331	F-statistic	347.2135
Durbin – Watson stat	0.608648	Prob(F-statistic)	0.000000

Table 6.9 Third model regression results (including only *PPI*)

Dependent Variable: LOG(IMP)
Method: Least Squares
Date: 02/17/04 Time: 02:22
Sample: 1990:1 1998:2
Included observations: 34

Variable	*Coefficient*	*Std. Error*	*t-Statistic*	*Prob.*
C	0.685704	0.370644	1.850031	0.0739
LOG(GDP)	2.093849	0.172585	12.13228	0.0000
LOG(PPI)	0.119566	0.136062	0.878764	0.3863

R-squared	0.962625	Mean dependent var	10.81363
Adjusted *R*-squared	0.960213	S.D. dependent var	0.138427
S.E. of regression	0.027612	Akaike info criterion	−4.257071
Sum squared resid	0.023634	Schwarz criterion	−4.122392
Log likelihood	75.37021	*F*-statistic	399.2113
Durbin–Watson stat	0.448237	Prob(*F*-statistic)	0.000000

Estimating the equation this time without log(*CPI*) but with log(*PPI*) we get the results shown in Table 6.9, which shows that log(*PPI*) is positive and insignificant! So, it is clear that the significance of log(*PPI*) in the specification above was due to the linear relationship that connects the two price variables.

So, the conclusions from this analysis are similar to the case of the collinear data set in Example 1 above, and can be summarized as follows:

1 The correlation among the explanatory variables was very high.
2 Standard errors or *t*-ratios of the estimated coefficients changed from estimation to estimation.