Introduction to Econometrics

THIRD EDITION UPDATE

James H. Stock
Harvard University

Mark W. Watson
Princeton University

2015

PEARSON

Boston Columbus Indianapolis New York San Francisco Hoboken
Amsterdam Cape Town Dubai London Madrid Milan Munich Paris Montréal Toronto
Delhi Mexico City São Paulo Sydney Hong Kong Seoul Singapore Taipei Tokyo

Equation (4.5) is the **linear regression model with a single regressor**, in which Y is the **dependent variable** and X is the **independent variable** or the **regressor**.

The first part of Equation (4.5), $\beta_0 + \beta_1 X_i$, is the **population regression line** or the **population regression function**. This is the relationship that holds between Y and X on average over the population. Thus, if you knew the value of X, according to this population regression line you would predict that the value of the dependent variable, Y, is $\beta_0 + \beta_1 X$.

The **intercept** β_0 and the **slope** β_1 are the **coefficients** of the population regression line, also known as the **parameters** of the population regression line. The slope β_1 is the change in Y associated with a unit change in X. The intercept is the value of the population regression line when $X = 0$; it is the point at which the population regression line intersects the Y axis. In some econometric applications, the intercept has a meaningful economic interpretation. In other applications, the intercept has no real-world meaning; for example, when X is the class size, strictly speaking the intercept is the predicted value of test scores when there are no students in the class! When the real-world meaning of the intercept is nonsensical, it is best to think of it mathematically as the coefficient that determines the level of the regression line.

The term u_i in Equation (4.5) is the **error term**. The error term incorporates all of the factors responsible for the difference between the i^{th} district's average test score and the value predicted by the population regression line. This error term contains all the other factors besides X that determine the value of the dependent variable, Y, for a specific observation, i. In the class size example, these other factors include all the unique features of the i^{th} district that affect the performance of its students on the test, including teacher quality, student economic background, luck, and even any mistakes in grading the test.

The linear regression model and its terminology are summarized in Key Concept 4.1.

Figure 4.1 summarizes the linear regression model with a single regressor for seven hypothetical observations on test scores (Y) and class size (X). The population regression line is the straight line $\beta_0 + \beta_1 X$. The population regression line slopes down ($\beta_1 < 0$), which means that districts with lower student–teacher ratios (smaller classes) tend to have higher test scores. The intercept β_0 has a mathematical meaning as the value of the Y axis intersected by the population regression line, but, as mentioned earlier, it has no real-world meaning in this example.

Because of the other factors that determine test performance, the hypothetical observations in Figure 4.1 do not fall exactly on the population regression line. For example, the value of Y for district #1, Y_1, is above the population regression line. This means that test scores in district #1 were better than predicted by the

Terminology for the Linear Regression Model with a Single Regressor

KEY CONCEPT 4.1

The linear regression model is

$$Y_i = \beta_0 + \beta_1 X_i + u_i,$$

where

the subscript i runs over observations, $i = 1, \ldots, n$;

Y_i is the *dependent variable*, the *regressand*, or simply the *left-hand variable*;

X_i is the *independent variable*, the *regressor*, or simply the *right-hand variable*;

$\beta_0 + \beta_1 X$ is the *population regression line* or the *population regression function*;

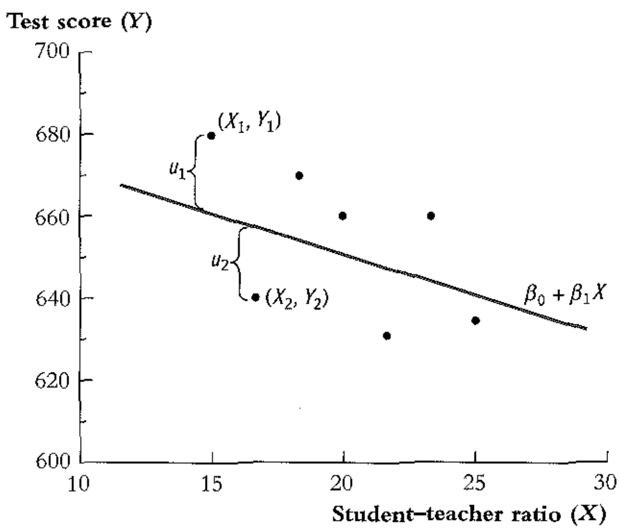
β_0 is the *intercept* of the population regression line;

β_1 is the *slope* of the population regression line; and

u_i is the *error term*.

FIGURE 4.1 Scatterplot of Test Score vs. Student–Teacher Ratio (Hypothetical Data)

The scatterplot shows hypothetical observations for seven school districts. The population regression line is $\beta_0 + \beta_1 X$. The vertical distance from the i^{th} point to the population regression line is $Y_i - (\beta_0 + \beta_1 X_i)$, which is the population error term u_i for the i^{th} observation.



population regression line, so the error term for that district, u_1, is positive. In contrast, Y_2 is below the population regression line, so test scores for that district were worse than predicted, and $u_2 < 0$.

Now return to your problem as advisor to the superintendent: What is the expected effect on test scores of reducing the student–teacher ratio by two students per teacher? The answer is easy: The expected change is $(-2) \times \beta_{ClassSize}$. But what is the value of $\beta_{ClassSize}$?

4.2 Estimating the Coefficients of the Linear Regression Model

In a practical situation such as the application to class size and test scores, the intercept β_0 and slope β_1 of the population regression line are unknown. Therefore, we must use data to estimate the unknown slope and intercept of the population regression line.

This estimation problem is similar to others you have faced in statistics. For example, suppose you want to compare the mean earnings of men and women who recently graduated from college. Although the population mean earnings are unknown, we can estimate the population means using a random sample of male and female college graduates. Then the natural estimator of the unknown population mean earnings for women, for example, is the average earnings of the female college graduates in the sample.

The same idea extends to the linear regression model. We do not know the population value of $\beta_{ClassSize}$, the slope of the unknown population regression line relating X (class size) and Y (test scores). But just as it was possible to learn about the population mean using a sample of data drawn from that population, so is it possible to learn about the population slope $\beta_{ClassSize}$ using a sample of data.

The data we analyze here consist of test scores and class sizes in 1999 in 420 California school districts that serve kindergarten through eighth grade. The test score is the districtwide average of reading and math scores for fifth graders. Class size can be measured in various ways. The measure used here is one of the broadest, which is the number of students in the district divided by the number of teachers—that is, the districtwide student–teacher ratio. These data are described in more detail in Appendix 4.1.

Table 4.1 summarizes the distributions of test scores and class sizes for this sample. The average student–teacher ratio is 19.6 students per teacher, and the standard deviation is 1.9 students per teacher. The 10th percentile of the distribution of the

TABLE 4.1 Summary of the Distribution of Student–Teacher Ratios and Fifth-Grade Test Scores for 420 K–8 Districts in California in 1999

			Percentile						
	Average	Standard Deviation	10%	25%	40%	50% (median)	60%	75%	90%
Student–teacher ratio	19.6	1.9	17.3	18.6	19.3	19.7	20.1	20.9	21.9
Test score	654.2	19.1	630.4	640.0	649.1	654.5	659.4	666.7	679.1

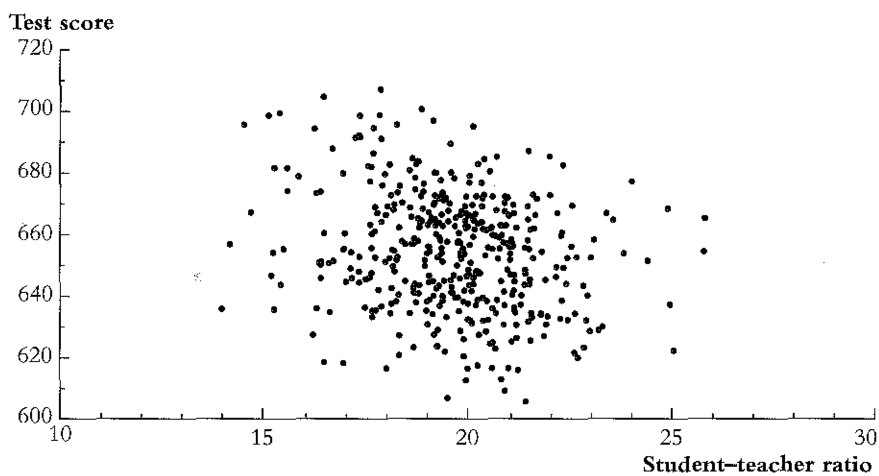
student–teacher ratio is 17.3 (that is, only 10% of districts have student–teacher ratios below 17.3), while the district at the 90th percentile has a student–teacher ratio of 21.9.

A scatterplot of these 420 observations on test scores and the student–teacher ratio is shown in Figure 4.2. The sample correlation is -0.23, indicating a weak negative relationship between the two variables. Although larger classes in this sample tend to have lower test scores, there are other determinants of test scores that keep the observations from falling perfectly along a straight line.

Despite this low correlation, if one could somehow draw a straight line through these data, then the slope of this line would be an estimate of $\beta_{ClassSize}$

FIGURE 4.2 Scatterplot of Test Score vs. Student–Teacher Ratio (California School District Data)

Data from 420 California school districts. There is a weak negative relationship between the student–teacher ratio and test scores: The sample correlation is -0.23.



based on these data. One way to draw the line would be to take out a pencil and a ruler and to "eyeball" the best line you could. While this method is easy, it is very unscientific, and different people will create different estimated lines.

How, then, should you choose among the many possible lines? By far the most common way is to choose the line that produces the "least squares" fit to these data—that is, to use the ordinary least squares (OLS) estimator.

The Ordinary Least Squares Estimator

The OLS estimator chooses the regression coefficients so that the estimated regression line is as close as possible to the observed data, where closeness is measured by the sum of the squared mistakes made in predicting Y given X.

As discussed in Section 3.1, the sample average, $\overline{Y}$, is the least squares estimator of the population mean, $E(Y)$; that is, $\overline{Y}$ minimizes the total squared estimation mistakes $\sum_{i=1}^{n}(Y_i - m)^2$ among all possible estimators m [see Expression (3.2)].

The OLS estimator extends this idea to the linear regression model. Let b_0 and b_1 be some estimators of β_0 and β_1. The regression line based on these estimators is $b_0 + b_1X$, so the value of Y_i predicted using this line is $b_0 + b_1X_i$. Thus the mistake made in predicting the i^{th} observation is $Y_i - (b_0 + b_1X_i) = Y_i - b_0 - b_1X_i$. The sum of these squared prediction mistakes over all n observations is

$$\sum_{i=1}^{n}(Y_i - b_0 - b_1X_i)^2. \tag{4.6}$$

The sum of the squared mistakes for the linear regression model in Expression (4.6) is the extension of the sum of the squared mistakes for the problem of estimating the mean in Expression (3.2). In fact, if there is no regressor, then b_1 does not enter Expression (4.6) and the two problems are identical except for the different notation [m in Expression (3.2), b_0 in Expression (4.6)]. Just as there is a unique estimator, $\overline{Y}$, that minimizes the Expression (3.2), so is there a unique pair of estimators of β_0 and β_1 that minimize Expression (4.6).

The estimators of the intercept and slope that minimize the sum of squared mistakes in Expression (4.6) are called the **ordinary least squares (OLS) estimators** of β_0 and β_1.

OLS has its own special notation and terminology. The OLS estimator of β_0 is denoted $\hat{\beta}_0$, and the OLS estimator of β_1 is denoted $\hat{\beta}_1$. The **OLS regression line**, also called the **sample regression line** or **sample regression function**, is the straight line constructed using the OLS estimators: $\hat{\beta}_0 + \hat{\beta}_1X$. The **predicted value** of Y_i

The OLS Estimator, Predicted Values, and Residuals

KEY CONCEPT 4.2

The OLS estimators of the slope β_1 and the intercept β_0 are

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} = \frac{s_{XY}}{s_X^2} \tag{4.7}$$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1\overline{X}. \tag{4.8}$$

The OLS predicted values $\hat{Y}_i$ and residuals $\hat{u}_i$ are

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i, \quad i = 1, \dots, n \tag{4.9}$$

$$\hat{u}_i = Y_i - \hat{Y}_i, \quad i = 1, \dots, n. \tag{4.10}$$

The estimated intercept ($\hat{\beta}_0$), slope ($\hat{\beta}_1$), and residual ($\hat{u}_i$) are computed from a sample of n observations of X_i and Y_i, $i = 1, \dots, n$. These are estimates of the unknown true population intercept (β_0), slope (β_1), and error term (u_i).

given X_i, based on the OLS regression line, is $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$. The **residual** for the i^{th} observation is the difference between Y_i and its predicted value: $\hat{u}_i = Y_i - \hat{Y}_i$.

The OLS estimators, $\hat{\beta}_0$ and $\hat{\beta}_1$, are sample counterparts of the population coefficients, β_0 and β_1. Similarly, the OLS regression line $\hat{\beta}_0 + \hat{\beta}_1 X$ is the sample counterpart of the population regression line $\beta_0 + \beta_1 X$, and the OLS residuals $\hat{u}_i$ are sample counterparts of the population errors u_i.

You could compute the OLS estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ by trying different values of b_0 and b_1 repeatedly until you find those that minimize the total squared mistakes in Expression (4.6); they are the least squares estimates. This method would be quite tedious, however. Fortunately, there are formulas, derived by minimizing Expression (4.6) using calculus, that streamline the calculation of the OLS estimators.

The OLS formulas and terminology are collected in Key Concept 4.2. These formulas are implemented in virtually all statistical and spreadsheet programs. These formulas are derived in Appendix 4.2.

OLS Estimates of the Relationship Between Test Scores and the Student–Teacher Ratio

When OLS is used to estimate a line relating the student–teacher ratio to test scores using the 420 observations in Figure 4.2, the estimated slope is −2.28 and the estimated intercept is 698.9. Accordingly, the OLS regression line for these 420 observations is

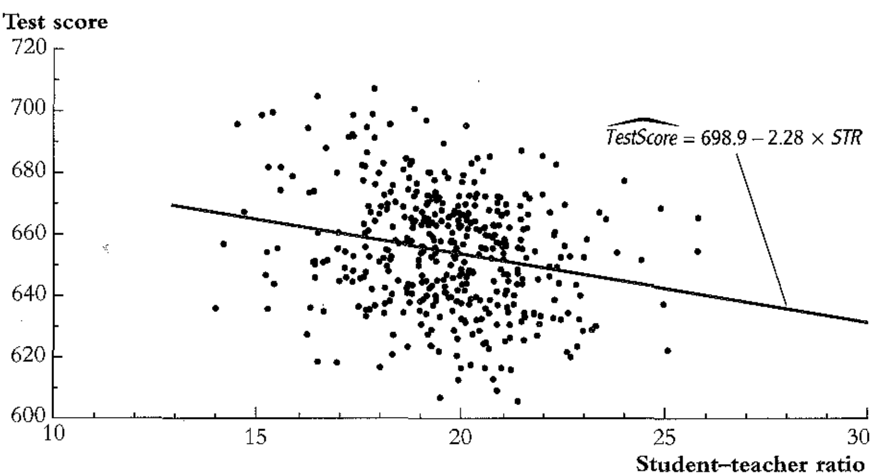
$$\widehat{TestScore} = 698.9 - 2.28 \times STR, \tag{4.11}$$

where *TestScore* is the average test score in the district and *STR* is the student–teacher ratio. The "^" over *TestScore* in Equation (4.11) indicates that it is the predicted value based on the OLS regression line. Figure 4.3 plots this OLS regression line superimposed over the scatterplot of the data previously shown in Figure 4.2.

The slope of −2.28 means that an increase in the student–teacher ratio by one student per class is, on average, associated with a decline in districtwide test scores by 2.28 points on the test. A decrease in the student–teacher ratio by two students per class is, on average, associated with an increase in test scores of 4.56 points $[= -2 \times (-2.28)]$. The negative slope indicates that more students per teacher (larger classes) is associated with poorer performance on the test.

FIGURE 4.3 The Estimated Regression Line for the California Data

The estimated regression line shows a negative relationship between test scores and the student–teacher ratio. If class sizes fall by one student, the estimated regression predicts that test scores will increase by 2.28 points.



It is now possible to predict the districtwide test score given a value of the student–teacher ratio. For example, for a district with 20 students per teacher, the predicted test score is $698.9 - 2.28 \times 20 = 653.3$. Of course, this prediction will not be exactly right because of the other factors that determine a district's performance. But the regression line does give a prediction (the OLS prediction) of what test scores would be for that district, based on their student–teacher ratio, absent those other factors.

Is this estimate of the slope large or small? To answer this, we return to the superintendent's problem. Recall that she is contemplating hiring enough teachers to reduce the student–teacher ratio by 2. Suppose her district is at the median of the California districts. From Table 4.1, the median student–teacher ratio is 19.7 and the median test score is 654.5. A reduction of two students per class, from 19.7 to 17.7, would move her student–teacher ratio from the 50th percentile to very near the 10th percentile. This is a big change, and she would need to hire many new teachers. How would it affect test scores?

According to Equation (4.11), cutting the student–teacher ratio by 2 is predicted to increase test scores by approximately 4.6 points; if her district's test scores are at the median, 654.5, they are predicted to increase to 659.1. Is this improvement large or small? According to Table 4.1, this improvement would move her district from the median to just short of the 60th percentile. Thus a decrease in class size that would place her district close to the 10% with the smallest classes would move her test scores from the 50th to the 60th percentile. According to these estimates, at least, cutting the student–teacher ratio by a large amount (two students per teacher) would help and might be worth doing depending on her budgetary situation, but it would not be a panacea.

What if the superintendent were contemplating a far more radical change, such as reducing the student–teacher ratio from 20 students per teacher to 5? Unfortunately, the estimates in Equation (4.11) would not be very useful to her. This regression was estimated using the data in Figure 4.2, and, as the figure shows, the smallest student–teacher ratio in these data is 14. These data contain no information on how districts with extremely small classes perform, so these data alone are not a reliable basis for predicting the effect of a radical move to such an extremely low student–teacher ratio.

Why Use the OLS Estimator?

There are both practical and theoretical reasons to use the OLS estimators $\hat{\beta}_0$ and $\hat{\beta}_1$. Because OLS is the dominant method used in practice, it has become the common language for regression analysis throughout economics, finance (see "The 'Beta' of a Stock" box), and the social sciences more generally. Presenting results

The "Beta" of a Stock

A fundamental idea of modern finance is that an investor needs a financial incentive to take a risk. Said differently, the expected return[1] on a risky investment, R, must exceed the return on a safe, or risk-free, investment, R_f. Thus the expected excess return, $R - R_f$, on a risky investment, like owning stock in a company, should be positive.

At first it might seem like the risk of a stock should be measured by its variance. Much of that risk, however, can be reduced by holding other stocks in a "portfolio"—in other words, by diversifying your financial holdings. This means that the right way to measure the risk of a stock is not by its *variance* but rather by its *covariance* with the market.

The capital asset pricing model (CAPM) formalizes this idea. According to the CAPM, the expected excess return on an asset is proportional to the expected excess return on a portfolio of all available assets (the "market portfolio"). That is, the CAPM says that

$$R - R_f = \beta(R_m - R_f), \tag{4.12}$$

where R_m is the expected return on the market portfolio and β is the coefficient in the population regression of $R - R_f$ on $R_m - R_f$. In practice, the risk-free return is often taken to be the rate of interest on short-term U.S. government debt. According to the CAPM, a stock with a $\beta < 1$ has less risk than the market portfolio and therefore has a lower expected excess return than the market portfolio. In contrast, a stock with a $\beta > 1$ is riskier than the market portfolio and thus commands a higher expected excess return.

The "beta" of a stock has become a workhorse of the investment industry, and you can obtain estimated betas for hundreds of stocks on investment firm websites. Those betas typically are estimated by OLS regression of the actual excess return on the stock against the actual excess return on a broad market index.

The table below gives estimated betas for seven U.S. stocks. Low-risk producers of consumer staples like Kellogg have stocks with low betas; riskier stocks have high betas.

Company	Estimated β
Verizon (telecommunications)	0.0
Wal-Mart (discount retailer)	0.3
Kellogg (breakfast cereal)	0.5
Waste Management (waste disposal)	0.6
Google (information technology)	1.0
Ford Motor Company (auto producer)	1.3
Bank of America (bank)	2.2

Source: finance.yahoo.com.

[1]The return on an investment is the change in its price plus any payout (dividend) from the investment as a percentage of its initial price. For example, a stock bought on January 1 for \$100, which then paid a \$2.50 dividend during the year and sold on December 31 for \$105, would have a return of $R = [(\$105 - \$100) + \$2.50]/\$100 = 7.5\%$.

using OLS (or its variants discussed later in this book) means that you are "speaking the same language" as other economists and statisticians. The OLS formulas are built into virtually all spreadsheet and statistical software packages, making OLS easy to use.