CHAPTER 3

Simple Linear Regression

3.1 Introduction

In this chapter, we study extensively the estimation of a linear relationship between two variables, Y_i and X_i, of the form:

$$Y_i = \alpha + \beta X_i + u_i \quad i = 1, 2, \ldots, n \tag{3.1}$$

where Y_i denotes the i-th observation on the dependent variable Y which could be consumption, investment or output, and X_i denotes the i-th observation on the independent variable X which could be disposable income, the interest rate or an input. These observations could be collected on firms or households at a given point in time, in which case we call the data a cross-section. Alternatively, these observations may be collected over time for a specific industry or country in which case we call the data a time-series. n is the number of observations, which could be the number of firms or households in a cross-section, or the number of years if the observations are collected annually. α and β are the intercept and slope of this simple linear relationship between Y and X. They are assumed to be unknown parameters to be estimated from the data. A plot of the data, i.e., Y versus X would be very illustrative showing what type of relationship exists empirically between these two variables. For example, if Y is consumption and X is disposable income then we would expect a positive relationship between these variables and the data may look like Figure 3.1 when plotted for a random sample of households. If α and β were known, one could draw the straight line $(\alpha + \beta X)$ as shown in Figure 3.1. It is clear that not all the observations (X_i, Y_i) lie on the straight line $(\alpha + \beta X)$. In fact, equation (3.1) states that the difference between each Y_i and the corresponding $(\alpha + \beta X_i)$ is due to a random error u_i. This error may be due to (i) the omission of relevant factors that could influence consumption, other than disposable income, like real wealth or varying tastes, or unforseen events that induce households to consume more or less, (ii) measurement error, which could be the result of households not reporting their consumption or income accurately, or (iii) wrong choice of a linear relationship between consumption and income, when the true relationship may be nonlinear. These different causes of the error term will have different effects on the distribution of this error. In what follows, we consider only disturbances that satisfy some restrictive assumptions. In later chapters we relax these assumptions to account for more general kinds of error terms.

In real life, α and β are not known, and have to be estimated from the observed data $\{(X_i, Y_i)$ for $i = 1, 2, \ldots, n\}$. This also means that the true line $(\alpha + \beta X)$ as well as the true disturbances (the u_i's) are unobservable. In this case, α and β could be estimated by the best fitting line through the data. Different researchers may draw different lines through the same data. What makes one line better than another? One measure of misfit is the amount of error from the observed Y_i to the guessed line, let us call the latter $\widehat{Y}_i = \widehat{\alpha} + \widehat{\beta} X_i$, where the hat ($\hat{}$) denotes a guess on the appropriate parameter or variable. Each observation (X_i, Y_i) will have a corresponding observable error attached to it, which we will call $e_i = Y_i - \widehat{Y}_i$, see Figure 3.2. In other words, we obtain the guessed $Y_i, (\widehat{Y}_i)$ corresponding to each X_i from the guessed line,

B.H. Baltagi, *Econometrics*, Springer Texts in Business and Economics, DOI 10.1007/978-3-642-20059-5_3,

$\widehat{\alpha} + \widehat{\beta} X_i$. Next, we find our error in guessing that Y_i, by subtracting the actual Y_i from the guessed $\widehat{Y}_i$. The only difference between Figure 3.1 and Figure 3.2 is the fact that Figure 3.1 draws the true consumption line which is unknown to the researcher, whereas Figure 3.2 is a guessed consumption line drawn through the data. Therefore, while the u_i's are unobservable, the e_i's are observable. Note that there will be n errors for each line, one error corresponding to every observation.

Similarly, there will be another set of n errors for another guessed line drawn through the data. For each guessed line, we can summarize its corresponding errors by one number, the sum of squares of these errors, which seems to be a natural criterion for penalizing a wrong guess. Note that a simple sum of these errors is not a good choice for a measure of misfit since positive errors end up canceling negative errors when both should be counted in our measure. However, this does not mean that the sum of squared error is the only single measure of misfit. Other measures include the sum of absolute errors, but this latter measure is mathematically more difficult to handle. Once the measure of misfit is chosen, α and β could then be estimated by minimizing this measure. In fact, this is the idea behind least squares estimation.

3.2 Least Squares Estimation and the Classical Assumptions

Least squares minimizes the residual sum of squares where the residuals are given by

$$e_i = Y_i - \widehat{\alpha} - \widehat{\beta} X_i \quad i = 1, 2, \ldots, n$$

and $\widehat{\alpha}$ and $\widehat{\beta}$ denote guesses on the regression parameters α and β, respectively. The residual sum of squares denoted by $RSS = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y_i - \widehat{\alpha} - \widehat{\beta} X_i)^2$ is minimized by the two first-order conditions:

$$\partial(\textstyle\sum_{i=1}^n e_i^2)/\partial\alpha = -2\sum_{i=1}^n e_i = 0; \text{ or } \sum_{i=1}^n Y_i - n\widehat{\alpha} - \widehat{\beta}\sum_{i=1}^n X_i = 0 \tag{3.2}$$

$$\partial(\textstyle\sum_{i=1}^n e_i^2)/\partial\beta = -2\sum_{i=1}^n e_i X_i = 0; \text{ or } \sum_{i=1}^n Y_i X_i - \widehat{\alpha}\displaystyle\sum_{i=1}^n X_i - \widehat{\beta}\sum_{i=1}^n X_i^2 = 0 \tag{3.3}$$

Solving the least squares normal equations given in (3.2) and (3.3) for α and β one gets

$$\widehat{\alpha}_{OLS} = \bar{Y} - \widehat{\beta}_{OLS}\bar{X} \text{ and } \widehat{\beta}_{OLS} = \textstyle\sum_{i=1}^n x_i y_i / \sum_{i=1}^n x_i^2 \tag{3.4}$$

where $\bar{Y} = \sum_{i=1}^n Y_i/n$, $\bar{X} = \sum_{i=1}^n X_i/n$, $y_i = Y_i - \bar{Y}$, $x_i = X_i - \bar{X}$, $\sum_{i=1}^n x_i^2 = \sum_{i=1}^n X_i^2 - n\bar{X}^2$, $\sum_{i=1}^n y_i^2 = \sum_{i=1}^n Y_i^2 - n\bar{Y}^2$ and $\sum_{i=1}^n x_i y_i = \sum_{i=1}^n X_i Y_i - n\bar{X}\bar{Y}$.

These estimators are subscripted by OLS denoting the ordinary least squares estimators. The OLS residuals $e_i = Y_i - \widehat{\alpha}_{OLS} - \widehat{\beta}_{OLS} X_i$ automatically satisfy the two numerical relationships given by (3.2) and (3.3). The first relationship states that (i) $\sum_{i=1}^n e_i = 0$, the residuals sum to zero. This is true as long as there is a constant in the regression. This numerical property of the least squares residuals also implies that the estimated regression line passes through the sample means $(\bar{X}, \bar{Y})$. To see this, average the residuals, or equation (3.2), this gives immediately $\bar{Y} = \widehat{\alpha}_{OLS} + \widehat{\beta}_{OLS}\bar{X}$. The second relationship states that (ii) $\sum_{i=1}^n e_i X_i = 0$, the residuals and the explanatory variable are uncorrelated. Other *numerical properties* that the OLS estimators satisfy are the following: (iii) $\sum_{i=1}^n \widehat{Y}_i = \sum_{i=1}^n Y_i$ and (iv) $\sum_{i=1}^n e_i \widehat{Y}_i = 0$. Property (iii) states that the sum of the estimated Y_i's or the predicted Y_i's from the sample is equal to the sum of the

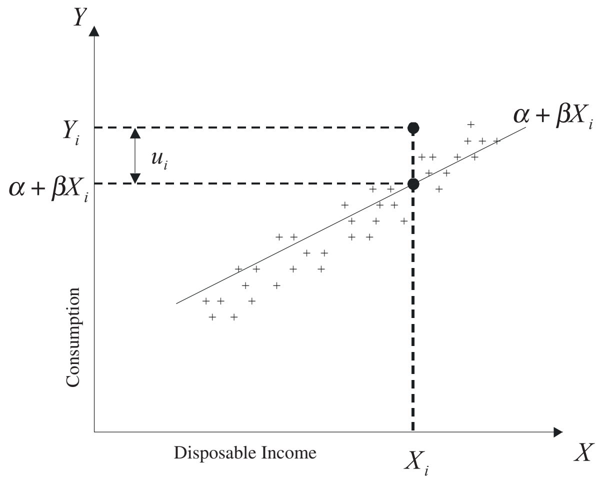

Figure 3.1 'True' Consumption Function

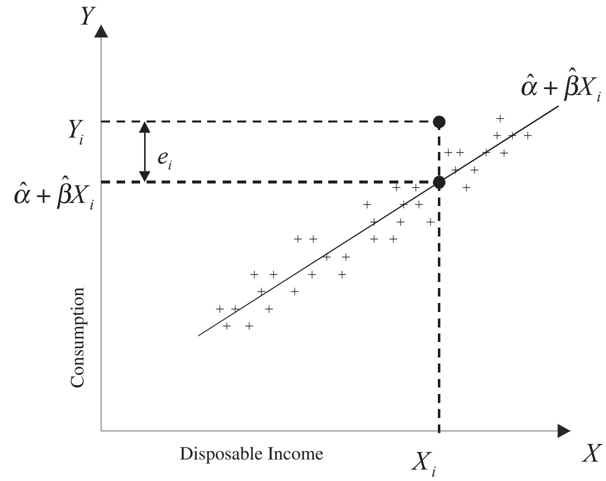

Figure 3.2 Estimated Consumption Function

actual Y_i's. Property (iv) states that the OLS residuals and the predicted Y_i's are uncorrelated. The proof of (iii) and (iv) *follow from (i) and (ii)* see problem 1. Of course, underlying our estimation of (3.1) is the assumption that (3.1) is the *true* model generating the data. In this case, (3.1) is *linear* in the parameters α and β, and contains only *one* explanatory variable X_i besides the constant. The inclusion of other explanatory variables in the model will be considered in Chapter 4, and the relaxation of the linearity assumption will be considered in Chapters 8 and 13. In order to study the statistical properties of the OLS estimators of α and β, we need to impose some statistical assumptions on the model generating the data.

Assumption 1: The disturbances have zero mean, i.e., $E(u_i) = 0$ for every $i = 1, 2, \ldots, n$. This assumption is needed to insure that on the average we are on the true line.

To see what happens if $E(u_i) \neq 0$, consider the case where households consistently under-report their consumption by a constant amount of δ dollars, while their income is measured accurately, say by cross-referencing it with their IRS tax forms. In this case,

$$(Observed\ Consumption) = (True\ Consumption) - \delta$$

and our regression equation is really

$$(True\ Consumption)_i = \alpha + \beta(Income)_i + u_i$$

But we observe,

$$(Observed\ Consumption)_i = \alpha + \beta(Income)_i + u_i - \delta$$

This can be thought of as the old regression equation with a new disturbance term $u_i^* = u_i - \delta$. Using the fact that $\delta > 0$ and $E(u_i) = 0$, one gets $E(u_i^*) = -\delta < 0$. This says that for all households with the same income, say $\$20,000$, their observed consumption will be on the average below that predicted from the true line $[\alpha + \beta(\$20,000)]$ by an amount δ. Fortunately, one can deal with this problem of constant but non-zero mean of the disturbances by reparametizing the model as

$$(Observed\ Consumption)_i = \alpha^* + \beta(Income)_i + u_i$$

where $\alpha^* = \alpha - \delta$. In this case, $E(u_i) = 0$ and α^* and β can be estimated from the regression. Note that while α^* is estimable, α and δ are non-estimable. Also note that for all \$20,000 income households, their average consumption is $[(\alpha - \delta) + \beta(\$20,000)]$.

Assumption 2: The disturbances have a constant variance, i.e., $\text{var}(u_i) = \sigma^2$ for every $i = 1, 2, \ldots, n$. This insures that every observation is equally reliable.

To see what this assumption means, consider the case where $\text{var}(u_i) = \sigma_i^2$, for $i = 1, 2, \ldots, n$. In this case, each observation has a different variance. An observation with a large variance is less reliable than one with a smaller variance. But, how can this differing variance happen? In the case of consumption, households with large disposable income (a large X_i, say \$100,000) may be able to save more (or borrow more to spend more) than households with smaller income (a small X_i, say \$10,000). In this case, the variation in consumption for the \$100,000 income household will be much larger than that for the \$10,000 income household. Therefore, the corresponding variance for the \$100,000 observation will be larger than that for the \$10,000 observation. Consequences of different variances for different observations will be studied more rigorously in Chapter 5.

Assumption 3: The disturbances are not correlated, i.e., $E(u_i u_j) = 0$ for $i \neq j, i, j = 1, 2, \ldots, n$. Knowing the i-th disturbance does not tell us anything about the j-th disturbance, for $i \neq j$.

For the consumption example, the unforseen disturbance which caused the i-th household to consume more, (like a visit of a relative), has nothing to do with the unforseen disturbances of any other household. This is likely to hold for a random sample of households. However, it is less likely to hold for a time series study of consumption for the aggregate economy, where a disturbance in 1945, a war year, is likely to affect consumption for several years after that. In this case, we say that the disturbance in 1945 is related to the disturbances in 1946, 1947, and so on. Consequences of correlated disturbances will be studied in Chapter 5.

Assumption 4: The explanatory variable X is nonstochastic, i.e., fixed in repeated samples, and hence, not correlated with the disturbances. Also, $\sum_{i=1}^n x_i^2/n \neq 0$ and has a finite limit as n tends to infinity.

This assumption states that we have at least two distinct values for X. This makes sense, since we need at least two distinct points to draw a straight line. Otherwise $\bar{X} = X$, the common value, and $x = X - \bar{X} = 0$, which violates $\sum_{i=1}^n x_i^2 \neq 0$. In practice, one always has several distinct values of X. More importantly, this assumption implies that X is not a random variable and hence is not correlated with the disturbances.

In section 5.3, we will relax the assumption of a non-stochastic X. Basically, X becomes a random variable and our assumptions have to be recast *conditional* on the set of X's that are observed. This is the more realistic case with economic data. The zero mean assumption becomes $E(u_i/X) = 0$, the constant variance assumption becomes $\text{var}(u_i/X) = \sigma^2$, the no serial correlation assumption becomes $E(u_i u_j/X) = 0$ for $i \neq j$. The conditional expectation here is with respect to *every* observation on X_i from $i = 1, 2, \ldots n$. Of course, one can show that if $E(u_i/X) = 0$ for all i, then X_i and u_i are not correlated. The reverse is not necessarily true, see

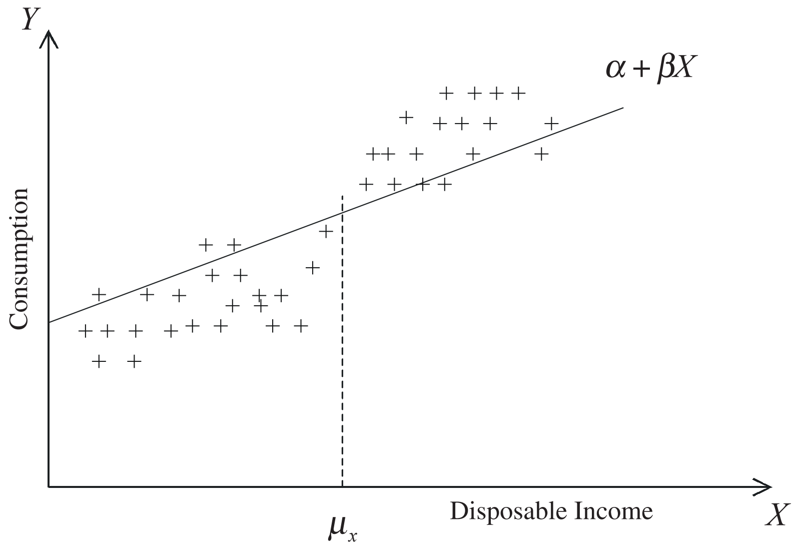

Figure 3.3 Consumption Function with $\text{Cov}(X, u) > 0$

problem 3 of Chapter 2. That problem shows that two random variables, say u_i and X_i could be uncorrelated, i.e., not linearly related when in fact they are nonlinearly related with $u_i = X_i^2$. Hence, $E(u_i/X_i) = 0$ is a stronger assumption than u_i and X_i are not correlated. By the law of iterated expectations given in the Appendix of Chapter 2, $E(u_i/X) = 0$ implies that $E(u_i) = 0$. It also implies that u_i is uncorrelated with *any* function of X_i. This is a stronger assumption than u_i is uncorrelated with X_i. Therefore, conditional on X_i, the mean of the disturbances is zero and does not depend on X_i. In this case, $E(Y_i/X_i) = \alpha + \beta X_i$ is linear in α and β and is assumed to be the *true* conditional mean of Y given X.

To see what a violation of assumption 4 means, suppose that X is a random variable and that X and u are positively correlated, then in the consumption example, households with income above the average income will be associated with disturbances above their mean of zero, and hence positive disturbances. Similarly, households with income below the average income will be associated with disturbances below their mean of zero, and hence negative disturbances. This means that the disturbances are systematically affected by values of the explanatory variable and the scatter of the data will look like Figure 3.3. Note that if we now erase the true line $(\alpha + \beta X)$, and estimate this line from the data, the least squares line drawn through the data is going to have a smaller intercept and a larger slope than those of the true line. The scatter should look like Figure 3.4 where the disturbances are random variables, not correlated with the X_i's, drawn from a distribution with zero mean and constant variance. Assumptions 1 and 4 insure that $E(Y_i/X_i) = \alpha + \beta X_i$, i.e., on the average we are on the true line. Several economic models will be studied where X and u are correlated. The consequences of this correlation will be studied in Chapters 5 and 11.

We now generate a data set which satisfies all four classical assumptions. Let α and β take the arbitrary values, say 10 and 0.5 respectively, and consider a set of 20 fixed X's, say income classes from \$10 to \$105 (in thousands of dollars), in increments of \$5, i.e., \$10, \$15, \$20, \$25, ..., \$105. Our consumption variable Y_i is constructed as $(10 + 0.5X_i + u_i)$ where u_i is a

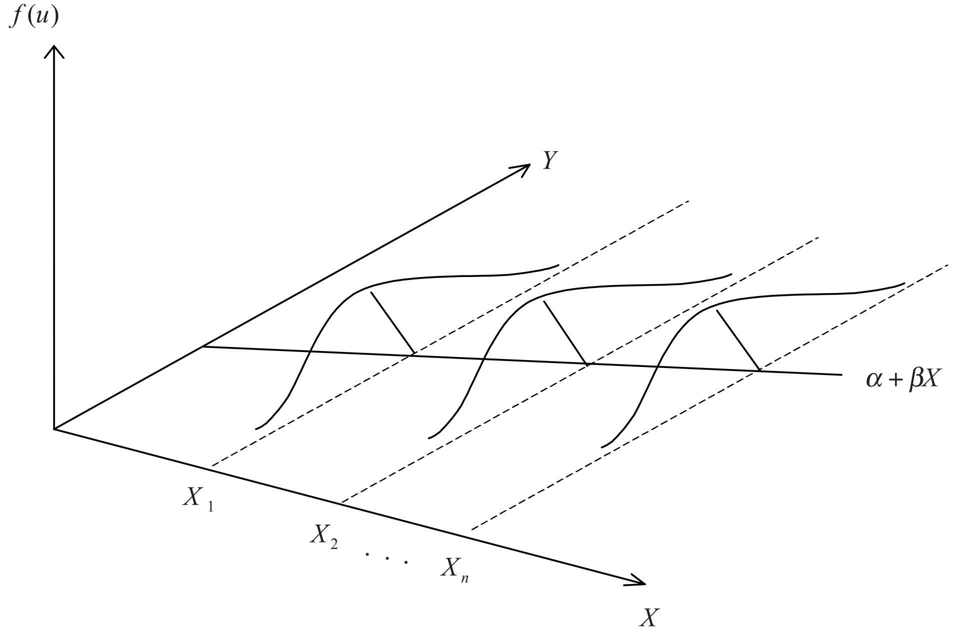

Figure 3.4 Random Disturbances Around the Regression

disturbance which is a random draw from a distribution with zero mean and constant variance, say $\sigma^2 = 9$. Computers generate random numbers with various distributions.

In this case, Figure 3.4 would depict our data, with the true line being $(10 + 0.5X)$ and u_i being random draws from the computer which are by construction independent and identically distributed with mean zero and variance 9. For every set of 20 u_i's randomly generated, given the fixed X_i's, we obtain a corresponding set of 20 Y_i's from our linear regression model. This is what we mean in assumption 4 when we say that the X's are fixed in repeated samples. Monte Carlo experiments generate a large number of samples, say a 1000, in the fashion described above. For each data set generated, least squares can be performed and the properties of the resulting estimators which are derived analytically in the remainder of this chapter, can be verified. For example, the average of the 1000 estimates of α and β can be compared to their true values to see whether these least squares estimates are unbiased. Note what will happen to Figure 3.4 if $E(u_i) = -\delta$ where $\delta > 0$, or $\text{var}(u_i) = \sigma_i^2$ for $i = 1, 2, \ldots, n$. In the first case, the mean of $f(u)$, the probability density function of u, will shift off the true line $(10+0.5X)$ by $-\delta$. In other words, we can think of the distributions of the u_i's, shown in Figure 3.4 , being centered on a new imaginary line parallel to the true line but lower by a distance δ. This means that one is more likely to draw negative disturbances than positive disturbances, and the observed Y_i's are more likely to be below the true line than above it. In the second case, each $f(u_i)$ will have a different variance, hence the spread of this probability density function will vary with each observation. In this case, Figure 3.4 will have a distribution for the u_i's which has a different spread for each observation. In other words, if the u_i's are say normally distributed, then u_1 is drawn from a $N(0, \sigma_1^2)$ distribution, whereas u_2 is drawn from a $N(0, \sigma_2^2)$ distribution, and so on. Violation of the classical assumptions can also be studied using Monte Carlo experiments, see Chapter 5.

3.3 Statistical Properties of Least Squares

(i) Unbiasedness

Given assumptions 1–4, it is easy to show that $\widehat{\beta}_{OLS}$ is unbiased for β. In fact, using equation (3.4) one can write

$$\widehat{\beta}_{OLS} = \sum_{i=1}^n x_i y_i / \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i Y_i / \sum_{i=1}^n x_i^2 = \beta + \sum_{i=1}^n x_i u_i / \sum_{i=1}^n x_i^2 \tag{3.5}$$

where the second equality follows from the fact that $y_i = Y_i - \bar{Y}$ and $\sum_{i=1}^n x_i \bar{Y} = \bar{Y} \sum_{i=1}^n x_i = 0$. The third equality follows from substituting Y_i from (3.1) and using the fact that $\sum_{i=1}^n x_i = 0$. Taking expectations of both sides of (3.5) and using assumptions 1 and 4, one can show that $E(\widehat{\beta}_{OLS}) = \beta$. Furthermore, one can derive the variance of $\widehat{\beta}_{OLS}$ from (3.5) since

$$\begin{aligned} \text{var}(\widehat{\beta}_{OLS}) &= E(\widehat{\beta}_{OLS} - \beta)^2 = E(\textstyle\sum_{i=1}^n x_i u_i / \sum_{i=1}^n x_i^2)^2 \\ &= \text{var}(\textstyle\sum_{i=1}^n x_i u_i / \sum_{i=1}^n x_i^2) = \sigma^2 / \sum_{i=1}^n x_i^2 \end{aligned} \tag{3.6}$$

where the last equality uses assumptions 2 and 3, i.e., that the u_i's are not correlated with each other and that their variance is constant, see problem 4. Note that the variance of the OLS estimator of β depends upon σ^2, the variance of the disturbances in the true model, and on the variation in X. The larger the variation in X the larger is $\sum_{i=1}^n x_i^2$ and the smaller is the variance of $\widehat{\beta}_{OLS}$.

(ii) Consistency

Next, we show that $\widehat{\beta}_{OLS}$ is consistent for β. A sufficient condition for consistency is that $\widehat{\beta}_{OLS}$ is unbiased and its variance tends to zero as n tends to infinity. We have already shown $\widehat{\beta}_{OLS}$ to be unbiased, it remains to show that its variance tends to zero as n tends to infinity.

$$\lim_{n\to\infty} \text{var}(\widehat{\beta}_{OLS}) = \lim_{n\to\infty} [(\sigma^2/n)/(\textstyle\sum_{i=1}^n x_i^2/n)] = 0$$

where the second equality follows from the fact that $(\sigma^2/n) \to 0$ and $(\sum_{i=1}^n x_i^2/n) \neq 0$ and has a finite limit, see assumption 4. Hence, plim $\widehat{\beta}_{OLS} = \beta$ and $\widehat{\beta}_{OLS}$ is consistent for β. Similarly one can show that $\widehat{\alpha}_{OLS}$ is unbiased and consistent for α with variance $\sigma^2 \sum_{i=1}^n X_i^2 / n \sum_{i=1}^n x_i^2$, and $\text{cov}(\widehat{\alpha}_{OLS}, \widehat{\beta}_{OLS}) = -\bar{X}\sigma^2 / \sum_{i=1}^n x_i^2$, see problem 5.

(iii) Best Linear Unbiased

Using (3.5) one can write $\widehat{\beta}_{OLS}$ as $\sum_{i=1}^n w_i Y_i$ where $w_i = x_i / \sum_{i=1}^n x_i^2$. This proves that $\widehat{\beta}_{OLS}$ is a linear combination of the Y_i's, with weights w_i satisfying the following properties:

$$\textstyle\sum_{i=1}^n w_i = 0; \sum_{i=1}^n w_i X_i = 1; \sum_{i=1}^n w_i^2 = 1/\sum_{i=1}^n x_i^2 \tag{3.7}$$

The next theorem shows that among all linear unbiased estimators of β, it is $\widehat{\beta}_{OLS}$ which has the smallest variance. This is known as the Gauss-Markov Theorem.

Theorem 1: Consider any arbitrary linear estimator $\widetilde{\beta} = \sum_{i=1}^n a_i Y_i$ for β, where the a_i's denote arbitrary constants. If $\widetilde{\beta}$ is unbiased for β, and assumptions 1 to 4 are satisfied, then $\text{var}(\widetilde{\beta}) \geq \text{var}(\widehat{\beta}_{OLS})$.

Proof: Substituting Y_i from (3.1) into $\widetilde{\beta}$, one gets $\widetilde{\beta} = \alpha\sum_{i=1}^n a_i + \beta\sum_{i=1}^n a_iX_i + \sum_{i=1}^n a_iu_i$. For $\widetilde{\beta}$ to be unbiased for β it must follow that $E(\widetilde{\beta}) = \alpha\sum_{i=1}^n a_i + \beta\sum_{i=1}^n a_iX_i = \beta$ for all observations $i = 1, 2, \ldots, n$. This means that $\sum_{i=1}^n a_i = 0$ and $\sum_{i=1}^n a_iX_i = 1$ for all $i = 1, 2, \ldots, n$. Hence, $\widetilde{\beta} = \beta + \sum_{i=1}^n a_iu_i$ with $\text{var}(\widetilde{\beta}) = \text{var}(\sum_{i=1}^n a_iu_i) = \sigma^2\sum_{i=1}^n a_i^2$ where the last equality follows from assumptions 2 and 3. But the a_i's are constants which differ from the w_i's, the weights of the OLS estimator, by some other constants, say d_i's, i.e., $a_i = w_i + d_i$ for $i = 1, 2, \ldots, n$. Using the properties of the a_i's and w_i one can deduce similar properties on the d_i's i.e., $\sum_{i=1}^n d_i = 0$ and $\sum_{i=1}^n d_iX_i = 0$. In fact,

$$\sum_{i=1}^n a_i^2 = \sum_{i=1}^n d_i^2 + \sum_{i=1}^n w_i^2 + 2\sum_{i=1}^n w_id_i$$

where $\sum_{i=1}^n w_id_i = \sum_{i=1}^n x_id_i / \sum_{i=1}^n x_i^2 = 0$. This follows from the definition of w_i and the fact that $\sum_{i=1}^n d_i = \sum_{i=1}^n d_iX_i = 0$. Hence,

$$\text{var}(\widetilde{\beta}) = \sigma^2\sum_{i=1}^n a_i^2 = \sigma^2\sum_{i=1}^n d_i^2 + \sigma^2\sum_{i=1}^n w_i^2 = \text{var}(\widehat{\beta}_{OLS}) + \sigma^2\sum_{i=1}^n d_i^2$$

Since $\sigma^2\sum_{i=1}^n d_i^2$ is non-negative, this proves that $\text{var}(\widetilde{\beta}) \geq \text{var}(\widehat{\beta}_{OLS})$ with the equality holding only if $d_i = 0$ for all $i = 1, 2, \ldots, n$, i.e., only if $a_i = w_i$, in which case $\widetilde{\beta}$ reduces to $\widehat{\beta}_{OLS}$. Therefore, any linear estimator of β, like $\widetilde{\beta}$ that is unbiased for β has variance at least as large as $\text{var}(\widehat{\beta}_{OLS})$. This proves that $\widehat{\beta}_{OLS}$ is BLUE, Best among all Linear Unbiased Estimators of β.

Similarly, one can show that $\widehat{\alpha}_{OLS}$ is linear in Y_i and has the smallest variance among all linear unbiased estimators of α, if assumptions 1 to 4 are satisfied, see problem 6. This result implies that the OLS estimator of α is also BLUE.

3.4 Estimation of σ^2

The variance of the regression disturbances σ^2 is unknown and has to be estimated. In fact, both the variance of $\widehat{\beta}_{OLS}$ and that of $\widehat{\alpha}_{OLS}$ depend upon σ^2, see (3.6) and problem 5. An unbiased estimator for σ^2 is $s^2 = \sum_{i=1}^n e_i^2/(n-2)$. To prove this, we need the fact that

$$e_i = Y_i - \widehat{\alpha}_{OLS} - \widehat{\beta}_{OLS}X_i = y_i - \widehat{\beta}_{OLS}x_i = (\beta - \widehat{\beta}_{OLS})x_i + (u_i - \bar{u})$$

where $\bar{u} = \sum_{i=1}^n u_i/n$. The second equality substitutes $\widehat{\alpha}_{OLS} = \bar{Y} - \widehat{\beta}_{OLS}\bar{X}$ and the third equality substitutes $y_i = \beta x_i + (u_i - \bar{u})$. Hence,

$$\sum_{i=1}^n e_i^2 = (\widehat{\beta}_{OLS} - \beta)^2\sum_{i=1}^n x_i^2 + \sum_{i=1}^n(u_i - \bar{u})^2 - 2(\widehat{\beta}_{OLS} - \beta)\sum_{i=1}^n x_i(u_i - \bar{u}),$$

and

$$\begin{aligned} E(\textstyle\sum_{i=1}^n e_i^2) &= \textstyle\sum_{i=1}^n x_i^2\text{var}(\widehat{\beta}_{OLS}) + (n-1)\sigma^2 - 2E(\sum_{i=1}^n x_iu_i)^2/\sum_{i=1}^n x_i^2 \\ &= \sigma^2 + (n-1)\sigma^2 - 2\sigma^2 = (n-2)\sigma^2 \end{aligned}$$

where the first equality uses the fact that $E(\sum_{i=1}^n(u_i - \bar{u})^2) = (n-1)\sigma^2$ and $\widehat{\beta}_{OLS} - \beta = \sum_{i=1}^n x_iu_i/\sum_{i=1}^n x_i^2$. The second equality uses the fact that $\text{var}(\widehat{\beta}_{OLS}) = \sigma^2/\sum_{i=1}^n x_i^2$ and

$$E(\sum_{i=1}^n x_iu_i)^2 = \sigma^2\sum_{i=1}^n x_i^2.$$

Therefore, $E(s^2) = E(\sum_{i=1}^n e_i^2/(n-2)) = \sigma^2$.

Intuitively, the estimator of σ^2 could be obtained from $\sum_{i=1}^n(u_i - \bar{u})^2/(n-1)$ if the true disturbances were known. Since the u's are not known, consistent estimates of them are used. These are the e_i's. Since $\sum_{i=1}^n e_i = 0$, our estimator of σ^2 becomes $\sum_{i=1}^n e_i^2/(n-1)$. Taking expectations we find that the correct divisor ought to be $(n-2)$ and not $(n-1)$ for this estimator to be unbiased for σ^2. This is plausible, since we have estimated two parameters α and β in obtaining the e_i's, and there are only $n-2$ independent pieces of information left in the data. To prove this fact, consider the OLS normal equations given in (3.2) and (3.3). These equations represent two relationships involving the e_i's. Therefore, knowing $(n-2)$ of the e_i's we can deduce the remaining two e_i's from (3.2) and (3.3).

3.5 Maximum Likelihood Estimation

Assumption 5: The u_i's are independent and identically distributed $N(0, \sigma^2)$.

This assumption allows us to derive distributions of estimators and other test statistics. In fact using (3.5) one can easily see that $\widehat{\beta}_{OLS}$ is a linear combination of the u_i's. But, a linear combination of normal random variables is itself a normal random variable, see Chapter 2, problem 15. Hence, $\widehat{\beta}_{OLS}$ is $N(\beta, \sigma^2/\sum_{i=1}^n x_i^2)$. Similarly $\widehat{\alpha}_{OLS}$ is $N(\alpha, \sigma^2\sum_{i=1}^n X_i^2/n\sum_{i=1}^n x_i^2)$, and Y_i is $N(\alpha+\beta X_i, \sigma^2)$. Moreover, we can write the joint probability density function of the u's as $f(u_1, u_2, \ldots, u_n; \alpha, \beta, \sigma^2) = (1/2\pi\sigma^2)^{n/2}\exp(-\sum_{i=1}^n u_i^2/2\sigma^2)$. To get the likelihood function we make the transformation $u_i = Y_i - \alpha - \beta X_i$ and note that the Jacobian of the transformation is 1. Therefore,

$$f(Y_1, Y_2, \ldots, Y_n; \alpha, \beta, \sigma^2) = (1/2\pi\sigma^2)^{n/2}\exp\{-\textstyle\sum_{i=1}^n(Y_i - \alpha - \beta X_i)^2/2\sigma^2\} \quad (3.8)$$

Taking the log of this likelihood, we get

$$\log L(\alpha, \beta, \sigma^2) = -(n/2)\log(2\pi\sigma^2) - \textstyle\sum_{i=1}^n(Y_i - \alpha - \beta X_i)^2/2\sigma^2 \quad (3.9)$$

Maximizing this likelihood with respect to α, β and σ^2 one gets the maximum likelihood estimators (MLE). However, only the second term in the log likelihood contains α and β and that term (without the negative sign) has already been minimized with respect to α and β in (3.2) and (3.3) giving us the OLS estimators. Hence, $\widehat{\alpha}_{MLE} = \widehat{\alpha}_{OLS}$ and $\widehat{\beta}_{MLE} = \widehat{\beta}_{OLS}$. Similarly, by differentiating $\log L$ with respect to σ^2 and setting this derivative equal to zero one gets $\widehat{\sigma}^2_{MLE} = \sum_{i=1}^n e_i^2/n$, see problem 7. Note that this differs from s^2 only in the divisor. In fact, $E(\widehat{\sigma}^2_{MLE}) = (n-2)\sigma^2/n \neq \sigma^2$. Hence, $\widehat{\sigma}^2_{MLE}$ is biased but note that it is still asymptotically unbiased.

So far, the gains from imposing assumption 5 are the following: The likelihood can be formed, maximum likelihood estimators can be derived, and distributions can be obtained for these estimators. One can also derive the Cramér-Rao lower bound for unbiased estimators of the parameters and show that the $\widehat{\alpha}_{OLS}$ and $\widehat{\beta}_{OLS}$ attain this bound whereas s^2 does not. This derivation is postponed until Chapter 7. In fact, one can show following the theory of complete sufficient statistics that $\widehat{\alpha}_{OLS}$, $\widehat{\beta}_{OLS}$ and s^2 are *minimum variance unbiased* estimators for α, β and σ^2, see Chapter 2. This is a stronger result (for $\widehat{\alpha}_{OLS}$ and $\widehat{\beta}_{OLS}$) than that obtained using the Gauss-Markov Theorem. It says, that among all unbiased estimators of α and β, the OLS estimators are the best. In other words, our set of estimators include now *all* unbiased estimators and not just *linear* unbiased estimators. This stronger result is obtained at the expense of a stronger distributional assumption, i.e., normality. If the distribution of the disturbances is not normal, then OLS is no longer MLE. In this case, MLE will be more efficient than OLS as long as the distribution of the disturbances is correctly specified. Some of the advantages and disadvantages of MLE were discussed in Chapter 2.

We found the distributions of $\widehat{\alpha}_{OLS}$, $\widehat{\beta}_{OLS}$, now we give that of s^2. In Chapter 7, it is shown that $\sum_{i=1}^n e_i^2/\sigma^2$ is a chi-squared with $(n-2)$ degrees of freedom. Also, s^2 is independent of $\widehat{\alpha}_{OLS}$ and $\widehat{\beta}_{OLS}$. This is useful for test of hypotheses. In fact, the major gain from assumption 5 is that we can perform test of hypotheses.

Standardizing the normal random variable $\widehat{\beta}_{OLS}$, one gets $z = (\widehat{\beta}_{OLS} - \beta)/(\sigma^2/\sum_{i=1}^n x_i^2)^{\frac{1}{2}} \sim N(0,1)$. Also, $(n-2)s^2/\sigma^2$ is distributed as χ^2_{n-2}. Hence, one can divide z, a $N(0,1)$ random variable, by the square root of $(n-2)s^2/\sigma^2$ divided by its degrees of freedom $(n-2)$ to get a t-statistic with $(n-2)$ degrees of freedom. The resulting statistic is $t_{obs} = (\widehat{\beta}_{OLS} - \beta)/(s^2/\sum_{i=1}^n x_i^2)^{\frac{1}{2}} \sim t_{n-2}$, see problem 8. This statistic can be used to test $H_0; \beta = \beta_0$, versus $H_1; \beta \neq \beta_0$, where β_0 is a known constant. Under H_0, t_{obs} can be calculated and its value can be compared to a critical value from a t-distribution with $(n-2)$ degrees of freedom, at a specified critical value of $\alpha\%$. Of specific interest is the hypothesis $H_0; \beta = 0$, which states that there is no linear relationship between Y_i and X_i. Under H_0,

$$t_{obs} = \widehat{\beta}_{OLS}/(s^2/\textstyle\sum_{i=1}^n x_i^2)^{\frac{1}{2}} = \widehat{\beta}_{OLS}/\widehat{se}(\widehat{\beta}_{OLS})$$

where $\widehat{se}(\widehat{\beta}_{OLS}) = (s^2/\sum_{i=1}^n x_i^2)^{\frac{1}{2}}$. If $|t_{obs}| > t_{\alpha/2;n-2}$, then H_0 is rejected at the $\alpha\%$ significance level. $\mathrm{t}_{\alpha/2;n-2}$ represents a critical value obtained from a t-distribution with $n-2$ degrees of freedom. It is determined such that the area to its right under a t_{n-2} distribution is equal to $\alpha/2$.

Similarly one can get a confidence interval for β by using the fact that, $\Pr[-t_{\alpha/2;n-2} < t_{obs} < t_{\alpha/2;n-2}] = 1 - \alpha$ and substituting for t_{obs} its value derived above as $(\widehat{\beta}_{OLS} - \beta)/\widehat{se}(\widehat{\beta}_{OLS})$. Since the critical values are known, $\widehat{\beta}_{OLS}$ and $\widehat{se}(\widehat{\beta}_{OLS})$ can be calculated from the data, the following $(1-\alpha)\%$ confidence interval for β emerges

$$\widehat{\beta}_{OLS} \pm t_{\alpha/2;n-2}\widehat{se}(\widehat{\beta}_{OLS}).$$

Tests of hypotheses and confidence intervals on α and σ^2 can be similarly constructed using the normal distribution of $\widehat{\alpha}_{OLS}$ and the χ^2_{n-2} distribution of $(n-2)s^2/\sigma^2$.

3.6 A Measure of Fit

We have obtained the least squares estimates of α, β and σ^2 and found their distributions under normality of the disturbances. We have also learned how to test hypotheses regarding these parameters. Now we turn to a measure of fit for this estimated regression line. Recall, that $e_i = Y_i - \widehat{Y}_i$ where $\widehat{Y}_i$ denotes the predicted Y_i from the least squares regression line at the value X_i, i.e., $\widehat{\alpha}_{OLS} + \widehat{\beta}_{OLS}X_i$. Using the fact that $\sum_{i=1}^n e_i = 0$, we deduce that $\sum_{i=1}^n Y_i = \sum_{i=1}^n \widehat{Y}_i$, and therefore, $\bar{Y} = \bar{\widehat{Y}}$. The actual and predicted values of Y have the same sample mean, see numerical properties (i) and (iii) of the OLS estimators discussed in section 2. This is true as long as there is a constant in the regression. Adding and subtracting $\bar{Y}$ from e_i, we get $e_i = y_i - \widehat{y}_i$, or $y_i = e_i + \widehat{y}_i$. Squaring and summing both sides:

$$\sum_{i=1}^n y_i^2 = \sum_{i=1}^n e_i^2 + \sum_{i=1}^n \widehat{y}_i^2 + 2\sum_{i=1}^n e_i\widehat{y}_i = \sum_{i=1}^n e_i^2 + \sum_{i=1}^n \widehat{y}_i^2 \tag{3.10}$$

where the last equality follows from the fact that $\widehat{y}_i = \widehat{\beta}_{OLS}x_i$ and $\sum_{i=1}^n e_i x_i = 0$. In fact,

$$\sum_{i=1}^n e_i\widehat{y}_i = \sum_{i=1}^n e_i\widehat{Y}_i = 0$$

means that the OLS residuals are uncorrelated with the predicted values from the regression, see numerical properties (ii) and (iv) of the OLS estimates discussed in section 3.2. In other words, (3.10) says that the total variation in Y_i, around its sample mean $\bar{Y}$ i.e., $\sum_{i=1}^n y_i^2$, can be decomposed into two parts: the first is the regression sums of squares $\sum_{i=1}^n \widehat{y}_i^2 = \widehat{\beta}_{OLS}^2 \sum_{i=1}^n x_i^2$, and the second is the residual sums of squares $\sum_{i=1}^n e_i^2$. In fact, regressing Y on a constant yields $\widetilde{\alpha}_{OLS} = \bar{Y}$, see problem 2, and the unexplained residual sums of squares of this naive model is

$$\sum_{i=1}^n (Y_i - \widetilde{\alpha}_{OLS})^2 = \sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n y_i^2.$$

Therefore, $\sum_{i=1}^n \widehat{y}_i^2$ in (3.10) gives the explanatory power of X after the constant is fit.

Using this decomposition, one can define the explanatory power of the regression as the ratio of the regression sums of squares to the total sums of squares. In other words, define $R^2 = \sum_{i=1}^n \widehat{y}_i^2 / \sum_{i=1}^n y_i^2$ and this value is clearly between 0 and 1. In fact, dividing (3.10) by $\sum_{i=1}^n y_i^2$ one gets $R^2 = 1 - \sum_{i=1}^n e_i^2 / \sum_{i=1}^n y_i^2$. The $\sum_{i=1}^n e_i^2$ is a measure of misfit which was minimized by least squares. If $\sum_{i=1}^n e_i^2$ is large, this means that the regression is not explaining a lot of the variation in Y and hence, the R^2 value would be small. Alternatively, if the $\sum_{i=1}^n e_i^2$ is small, then the fit is good and R^2 is large. In fact, for a perfect fit, where all the observations lie on the fitted line, $Y_i = \widehat{Y}_i$ and $e_i = 0$, which means that $\sum_{i=1}^n e_i^2 = 0$ and $R^2 = 1$. The other extreme case is where the regression sums of squares $\sum_{i=1}^n \widehat{y}_i^2 = 0$. In other words, the linear regression explains nothing of the variation in Y_i. In this case, $\sum_{i=1}^n y_i^2 = \sum_{i=1}^n e_i^2$ and $R^2 = 0$. Note that since $\sum_{i=1}^n \widehat{y}_i^2 = 0$ implies $\widehat{y}_i = 0$ for every i, which in turn means that $\widehat{Y}_i = \bar{Y}$ for every i. The fitted regression line is a horizontal line drawn at $Y = \bar{Y}$, and the independent variable X does not have any explanatory power in a linear relationship with Y.

Note that R^2 has two alternative meanings: (i) It is the simple squared correlation coefficient between Y_i and $\widehat{Y}_i$, see problem 9. Also, for the simple regression case, (ii) it is the simple squared correlation between X and Y. This means that before one runs the regression of Y on X, one can compute r_{xy}^2 which in turn tells us the proportion of the variation in Y that will be explained by X. If this number is pretty low, we have a weak linear relationship between Y and X and we know that a poor fit will result if Y is regressed on X. It is worth emphasizing that R^2 is a measure of *linear* association between Y and X. There could exist, for example, a perfect quadratic relationship between X and Y, yet the estimated least squares line through the data is a flat line implying that $R^2 = 0$, see problem 3 of Chapter 2. One should also be suspicious of least squares regressions with R^2 that are too close to 1. In some cases, we may not want to include a constant in the regression. In such cases, one should use an *uncentered* R^2 as a measure fit. The appendix to this chapter defines both *centered* and *uncentered* R^2 and explains the difference between them.

3.7 Prediction

Let us now predict Y_0 given X_0. Usually this is done for a time series regression, where the researcher is interested in predicting the future, say one period ahead. This new observation Y_0 is generated by (3.1), i.e.,

$$Y_0 = \alpha + \beta X_0 + u_0 \tag{3.11}$$

What is the Best Linear Unbiased Predictor (BLUP) of $E(Y_0)$? From (3.11), $E(Y_0) = \alpha + \beta X_0$ is a linear combination of α and β. Using the Gauss-Markov result, $\widehat{Y}_0 = \widehat{\alpha}_{OLS} + \widehat{\beta}_{OLS} X_0$ is BLUE for $\alpha + \beta X_0$ and the variance of this predictor of $E(Y_0)$ is $\sigma^2[(1/n) + (X_0 - \bar{X})^2 / \sum_{i=1}^n x_i^2]$, see problem 10. But, what if we are interested in the BLUP for Y_0 itself ? Y_0 differs from E(Y_0) by u_0, and the best predictor of u_0 is zero, so the BLUP for Y_0 is still $\widehat{Y}_0$. The forecast error is

$$Y_0 - \widehat{Y}_0 = [Y_0 - E(Y_0)] + [E(Y_0) - \widehat{Y}_0] = u_0 + [E(Y_0) - \widehat{Y}_0]$$

where u_0 is the error committed even if the true regression line is known, and $E(Y_0) - \widehat{Y}_0$ is the difference between the sample and population regression lines. Hence, the variance of the forecast error becomes:

$$\text{var}(u_0) + \text{var}[E(Y_0) - \widehat{Y}_0] + 2\text{cov}[u_0, E(Y_0) - \widehat{Y}_0] = \sigma^2[1 + (1/n) + (X_0 - \bar{X})^2 / \textstyle\sum_{i=1}^n x_i^2]$$

This says that the variance of the forecast error is equal to the variance of the predictor of $E(Y_0)$ plus the var(u_0) plus twice the covariance of the predictor of E(Y_0) and u_0. But, this last covariance is zero, since u_0 is a new disturbance and is not correlated with the disturbances in the sample upon which $\widehat{Y}_i$ is based. Therefore, the predictor of the average consumption of a $\$20,000$ income household is the same as the predictor of consumption for a specific household whose income is $\$20,000$. The difference is not in the predictor itself but in the variance attached to it. The latter variance being larger only by σ^2, the variance of u_0. The variance of the predictor therefore, depends upon σ^2, the sample size, the variation in the X's, and how far X_0 is from the sample mean of the observed data. To summarize, the smaller σ^2 is, the larger n and $\sum_{i=1}^n x_i^2$ are, and the closer X_0 is to $\bar{X}$, the smaller is the variance of the predictor. One can construct 95% confidence intervals to these predictions for every value of X_0. In fact, this is $(\widehat{\alpha}_{OLS} + \widehat{\beta}_{OLS} X_0) \pm t_{.025;n-2}\{s[1 + (1/n) + (X_0 - \bar{X})^2 / \sum_{i=1}^n x_i^2]^{\frac{1}{2}}\}$ where s replaces σ, and $\text{t}_{.025;n-2}$ represents the 2.5% critical value obtained from a t-distribution with $n-2$ degrees of freedom. Figure 3.5 shows this confidence band around the estimated regression line. This is a hyperbola which is the narrowest at $\bar{X}$ as expected, and widens as we predict away from $\bar{X}$.

3.8 Residual Analysis

A plot of the residuals of the regression is very important. The residuals are consistent estimates of the true disturbances. But unlike the u_i's, these e_i's are not independent. In fact, the OLS normal equations (3.2)and (3.3) give us two relationships between these residuals. Therefore, knowing $(n-2)$ of these residuals the remaining two residuals can be deduced. If we had the true u_i's, and we plotted them, they should look like a random scatter around the horizontal axis with no specific pattern to them. A plot of the e_i's that shows a certain pattern like a set of positive residuals followed by a set of negative residuals as shown in Figure 3.6(a) may be

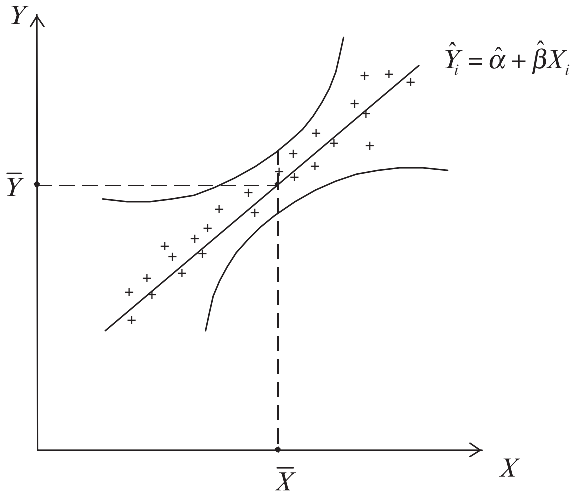

Figure 3.5 95% Confidence Bands

indicative of a violation of one of the 5 assumptions imposed on the model, or simply indicating a wrong functional form. For example, if assumption 3 is violated, so that the u_i's are say positively correlated, then it is likely to have a positive residual followed by a positive one, and a negative residual followed by a negative one, as observed in Figure 3.6(b). Alternatively, if we fit a linear regression line to a true quadratic relation between Y and X, then a scatter of residuals like that in Figure 3.6(c) will be generated. We will study how to deal with this violation and how to test for it in Chapter 5.

Large residuals are indicative of bad predictions in the sample. A large residual could be a typo, where the researcher entered this observation wrongly. Alternatively, it could be an influential observation, or an outlier which behaves differently from the other data points in the sample and therefore, is further away from the estimated regression line than the other data points. The fact that OLS minimizes the sum of squares of these residuals means that a large weight is put on this observation and hence it is influential. In other words, removing this observation from the sample may change the estimates and the regression line significantly. For more on the study of influential observations, see Belsely, Kuh and Welsch (1980). We will focus on this issue in Chapter 8 of this book.

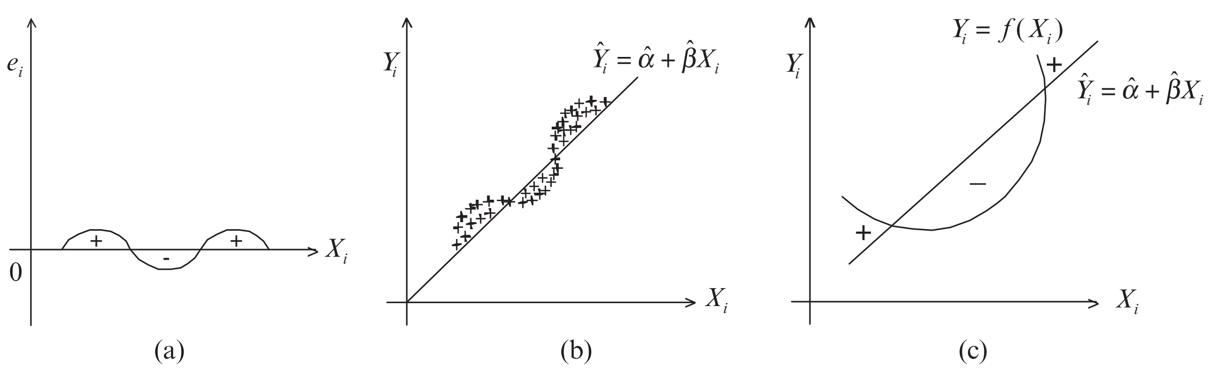

Figure 3.6 Positively Correlated Residuals

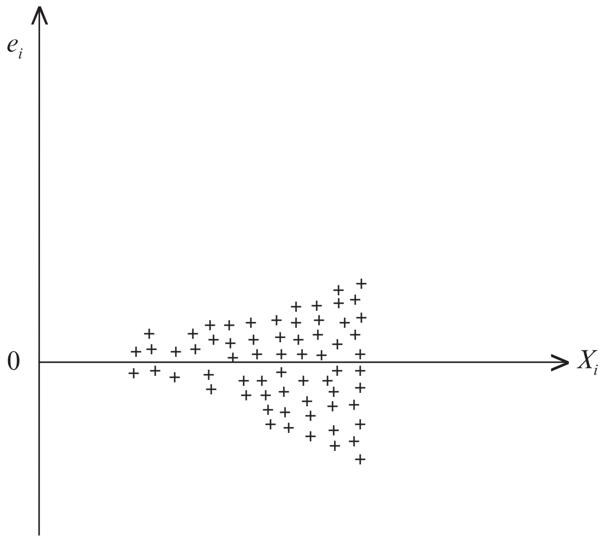

Figure 3.7 Residual Variation Growing with X

One can also plot the residuals versus the X_i's. If a pattern like Figure 3.7 emerges, this could be indicative of a violation of assumption 2 because the variation of the residuals is growing with X_i when it should be constant for all observations. Alternatively, it could imply a relationship between the X_i's and the true disturbances which is a violation of assumption 4.

In summary, one should always plot the residuals to check the data, identify influential observations, and check violations of the 5 assumptions underlying the regression model. In the next few chapters, we will study various tests of the violation of the classical assumptions. Most of these tests are based on the residuals of the model. These tests along with residual plots should help the researcher gauge the adequacy of his or her model.

Table 3.1 Simple Regression Computations

OBS	Consumption y_i	Income x_i	$y_i=Y_i-\bar{Y}$	$x_i=X_i$-$\bar{X}$	x_iy_i	x_i^2	$\hat{Y}_i$	e_i
1	4.6	5	–1.9	–2.5	4.75	6.25	4.476190	0.123810
2	3.6	4	–2.9	–3.5	10.15	12.25	3.666667	–0.066667
3	4.6	6	–1.9	–1.5	2.85	2.25	5.285714	–0.685714
4	6.6	8	0.1	0.5	0.05	0.25	6.904762	–0.304762
5	7.6	8	1.1	0.5	0.55	0.25	6.904762	0.695238
6	5.6	7	–0.9	–0.5	0.45	0.25	6.095238	–0.495238
7	5.6	6	–0.9	–1.5	1.35	2.25	5.285714	0.314286
8	8.6	9	2.1	1.5	3.15	2.25	7.714286	0.885714
9	8.6	10	2.1	2.5	5.25	6.25	8.523810	0.076190
10	9.6	12	3.1	4.5	13.95	20.25	10.142857	–0.542857
SUM	6.5	75	0	0	42.5	52.5	65	0
MEAN	6.5	7.5					6.5	

3.9 Numerical Example

Table 3.1 gives the annual consumption of 10 households each selected randomly from a group of households with a fixed personal disposable income. Both income and consumption are measured in \$10,000, so that the first household earns \$50,000 and consumes \$46,000 annually. It is worthwhile doing the computations necessary to obtain the least squares regression estimates of consumption on income in this simple case and to compare them with those obtained from a regression package. In order to do this, we first compute $\bar{Y} = 6.5$ and $\bar{X} = 7.5$ and form two new columns of data made up of $y_i = Y_i - \bar{Y}$ and $x_i = X_i - \bar{X}$. To get $\widehat{\beta}_{OLS}$ we need $\sum_{i=1}^{n} x_i y_i$, so we multiply these last two columns by each other and sum to get 42.5. The denominator of $\widehat{\beta}_{OLS}$ is given by $\sum_{i=1}^{n} x_i^2$. This is why we square the x_i column to get x_i^2 and sum to obtain 52.5. Our estimate of $\widehat{\beta}_{OLS} = 42.5/52.5 = 0.8095$ which is the estimated marginal propensity to consume. This is the extra consumption brought about by an extra dollar of disposable income.

$$\widehat{\alpha}_{OLS} = \bar{Y} - \widehat{\beta}_{OLS}\bar{X} = 6.5 - (0.8095)(7.5) = 0.4286$$

This is the estimated consumption at zero personal disposable income. The fitted values or predicted values from this regression are computed from $\widehat{Y}_i = \widehat{\alpha}_{OLS} + \widehat{\beta}_{OLS}X_i = 0.4286 + 0.8095X_i$ and are given in Table 3.1. Note that the mean of $\widehat{Y}_i$ is equal to the mean of Y_i confirming one of the numerical properties of least squares. The residuals are computed from $e_i = Y_i - \widehat{Y}_i$ and they satisfy $\sum_{i=1}^{n} e_i = 0$. It is left to the reader to verify that $\sum_{i=1}^{n} e_i X_i = 0$. The residual sum of squares is obtained by squaring the column of residuals and summing it. This gives us $\sum_{i=1}^{n} e_i^2 = 2.495238$. This means that $s^2 = \sum_{i=1}^{n} e_i^2/(n-2) = 0.311905$. Its square root is given by $s = 0.558$. This is known as the standard error of the regression. In this case, the estimated $\text{var}(\widehat{\beta}_{OLS})$ is $s^2/\sum_{i=1}^{n} x_i^2 = 0.311905/52.5 = 0.005941$ and the estimated

$$\text{var}(\widehat{\alpha}) = s^2\left[\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n} x_i^2}\right] = 0.311905\left[\frac{1}{10} + \frac{(7.5)^2}{52.5}\right] = 0.365374$$

Taking the square root of these estimated variances, we get the estimated standard errors of $\widehat{\alpha}_{OLS}$ and $\widehat{\beta}_{OLS}$ given by $\widehat{se}(\widehat{\alpha}_{OLS}) = 0.60446$ and $\widehat{se}(\widehat{\beta}_{OLS}) = 0.077078$.

Since the disturbances are normal, the OLS estimators are also the maximum likelihood estimators, and are normally distributed themselves. For the null hypothesis $H_0^a; \beta = 0$; the observed t-statistic is

$$t_{obs} = (\widehat{\beta}_{OLS} - 0)/\widehat{se}(\widehat{\beta}_{OLS}) = 0.809524/0.077078 = 10.50$$

and this is highly significant, since $\Pr[|t_8| > 10.5] < 0.0001$. This probability can be obtained using most regression packages. It is also known as the p-value or probability value. It shows that this t-value is highly unlikely and we reject H_0^a that $\beta = 0$. Similarly, the null hypothesis $H_0^b; \alpha = 0$, gives an observed t-statistic of $t_{obs} = (\widehat{\alpha}_{OLS} - 0)/\widehat{se}(\widehat{\alpha}_{OLS}) = 0.428571/0.604462 = 0.709$, which is not significant, since its p-value is $\Pr[|t_8| > 0.709] < 0.498$. Hence, we do not reject the null hypothesis H_0^b that $\alpha = 0$.

The total sum of squares is $\sum_{i=1}^{n} y_i^2 = \sum_{i=1}^{n}(Y_i - \bar{Y})^2$ which can be obtained by squaring the y_i column in Table 3.1 and summing. This yields $\sum_{i=1}^{n} y_i^2 = 36.9$. Also, the regression sum of squares $= \sum_{i=1}^{n} \widehat{y}_i^2 = \sum_{i=1}^{n}(\widehat{Y}_i - \bar{Y})^2$ which can be obtained by subtracting $\bar{Y} = \overline{\widehat{Y}} = 6.5$ from the $\widehat{Y}_i$ column, squaring that column and summing. This yields 34.404762. This could have also been obtained as

$$\sum_{i=1}^n \widehat{y}_i^2 = \widehat{\beta}_{OLS}^2 \sum_{i=1}^n x_i^2 = (0.809524)^2(52.5) = 34.404762.$$

A final check is that $\sum_{i=1}^n \widehat{y}_i^2 = \sum_{i=1}^n y_i^2 - \sum_{i=1}^n e_i^2 = 36.9 - 2.495238 = 34.404762$ as required.

Recall, that $R^2 = r_{xy}^2 = (\sum_{i=1}^n x_i y_i)^2/(\sum_{i=1}^n x_i^2)(\sum_{i=1}^n y_i^2) = (42.5)^2/(52.5)(36.9) = 0.9324$. This could have also been obtained as $R^2 = 1 - (\sum_{i=1}^n e_i^2 / \sum_{i=1}^n y_i^2) = 1 - (2.495238/36.9) = 0.9324$, or as

$$R^2 = r_{y\hat{y}}^2 = \sum_{i=1}^n \widehat{y}_i^2 / \sum_{i=1}^n y_i^2 = 34.404762/36.9 = 0.9324.$$

This means that personal disposable income explains 93.24% of the variation in consumption. A plot of the actual, predicted and residual values versus time is given in Figure 3.8. This was done using EViews.

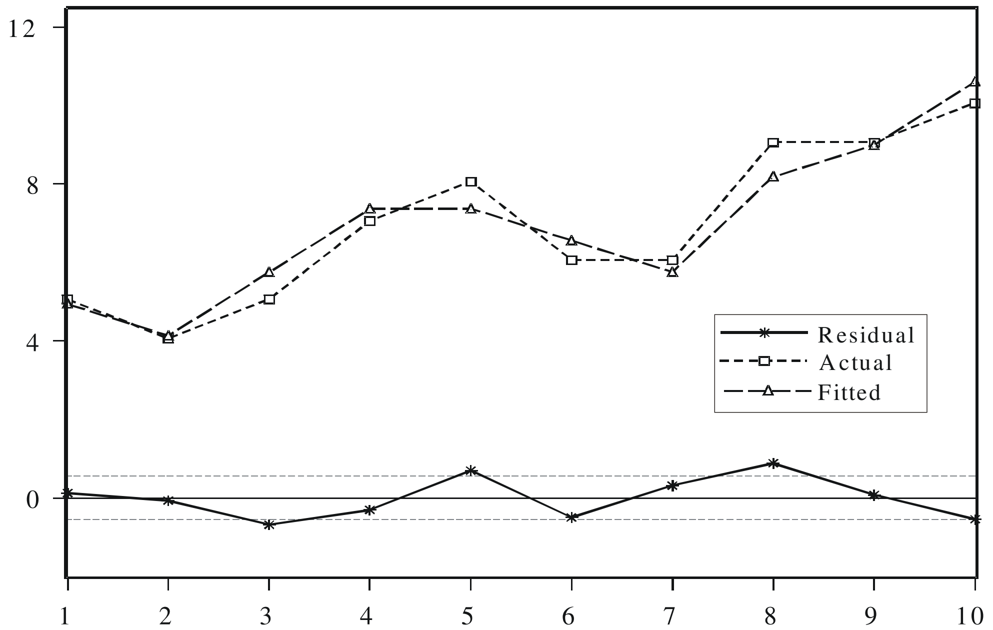

Figure 3.8 Residual Plot

3.10 Empirical Example

Table 3.2 gives (i) the logarithm of cigarette consumption (in packs) per person of smoking age (> 16 years) for 46 states in 1992, (ii) the logarithm of real price of cigarettes in each state, and (iii) the logarithm of real disposable income per capita in each state. This is drawn from Baltagi and Levin (1992) study on dynamic demand for cigarettes. It can be downloaded as Cigarett.dat from the Springer web site.

Table 3.2 Cigarette Consumption Data

LNC: log of consumption (in packs) per person of smoking age (>16)
LNP: log of real price (1983$/pack)
LNY: log of real disposable income per-capita (in thousand 1983$)

OBS	STATE	LNC	LNP	LNY
1	AL	4.96213	0.20487	4.64039
2	AZ	4.66312	0.16640	4.68389
3	AR	5.10709	0.23406	4.59435
4	CA	4.50449	0.36399	4.88147
5	CT	4.66983	0.32149	5.09472
6	DE	5.04705	0.21929	4.87087
7	DC	4.65637	0.28946	5.05960
8	FL	4.80081	0.28733	4.81155
9	GA	4.97974	0.12826	4.73299
10	ID	4.74902	0.17541	4.64307
11	IL	4.81445	0.24806	4.90387
12	IN	5.11129	0.08992	4.72916
13	IA	4.80857	0.24081	4.74211
14	KS	4.79263	0.21642	4.79613
15	KY	5.37906	–0.03260	4.64937
16	LA	4.98602	0.23856	4.61461
17	ME	4.98722	0.29106	4.75501
18	MD	4.77751	0.12575	4.94692
19	MA	4.73877	0.22613	4.99998
20	MI	4.94744	0.23067	4.80620
21	MN	4.69589	0.34297	4.81207
22	MS	4.93990	0.13638	4.52938
23	MO	5.06430	0.08731	4.78189
24	MT	4.73313	0.15303	4.70417
25	NE	4.77558	0.18907	4.79671
26	NV	4.96642	0.32304	4.83816
27	NH	5.10990	0.15852	5.00319
28	NJ	4.70633	0.30901	5.10268
29	NM	4.58107	0.16458	4.58202
30	NY	4.66496	0.34701	4.96075
31	ND	4.58237	0.18197	4.69163
32	OH	4.97952	0.12889	4.75875
33	OK	4.72720	0.19554	4.62730
34	PA	4.80363	0.22784	4.83516
35	RI	4.84693	0.30324	4.84670
36	SC	5.07801	0.07944	4.62549
37	SD	4.81545	0.13139	4.67747
38	TN	5.04939	0.15547	4.72525
39	TX	4.65398	0.28196	4.73437
40	UT	4.40859	0.19260	4.55586
41	VT	5.08799	0.18018	4.77578
42	VA	4.93065	0.11818	4.85490
43	WA	4.66134	0.35053	4.85645
44	WV	4.82454	0.12008	4.56859
45	WI	4.83026	0.22954	4.75826
46	WY	5.00087	0.10029	4.71169

Data: Cigarette Consumption of 46 States in 1992

Table 3.3 Cigarette Consumption Regression

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob > F
Model	1	0.48048	0.48048	18.084	0.0001
Error	44	1.16905	0.02657		
Root MSE		0.16300	R-square	0.2913	
Dep Mean		4.84784	Adj R-sq	0.2752	
C.V.		3.36234			

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	5.094108	0.06269897	81.247	0.0001
LNP	1	–1.198316	0.28178857	–4.253	0.0001

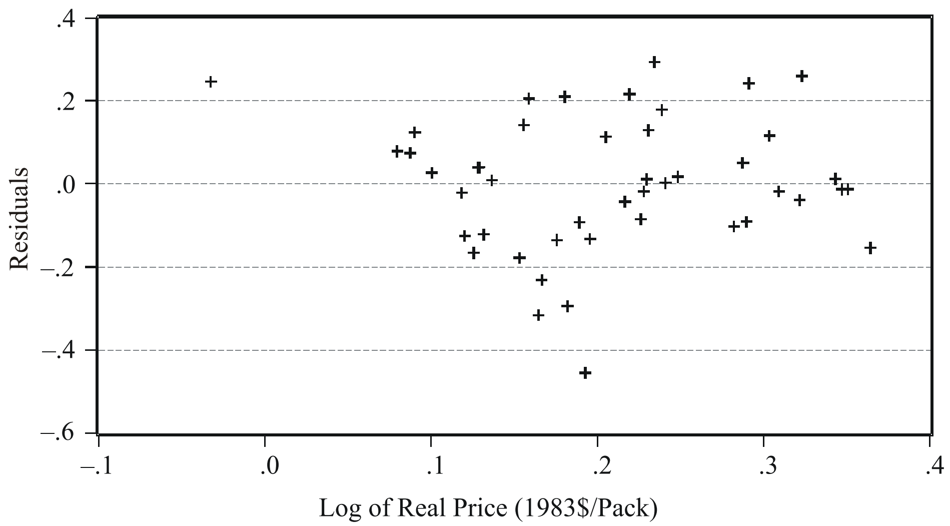

Figure 3.9 Residuals Versus LNP

Table 3.3 gives the SAS output for the regression of log$C$ on log$P$. The price elasticity of demand for cigarettes in this simple model is ($d$log$C$/log$P$) which is the slope coefficient. This is estimated to be -1.198 with a standard error of 0.282. This says that a 10% increase in real price of cigarettes has an estimated 12% drop in per capita consumption of cigarettes. The R^2 of this regression is 0.29, s^2 is given by the Mean Square Error of the regression which is 0.0266. Figure 3.9 plots the residuals of this regression versus the independent variable, while Figure 3.10 plots the predictions along with the 95% confidence interval band for these predictions. One observation clearly stands out as an influential observation given its distance from the rest of the data and that is the observation for Kentucky, a producer state with very low real price. This observation almost anchors the straight line fit through the data. More on influential observations in Chapter 8.

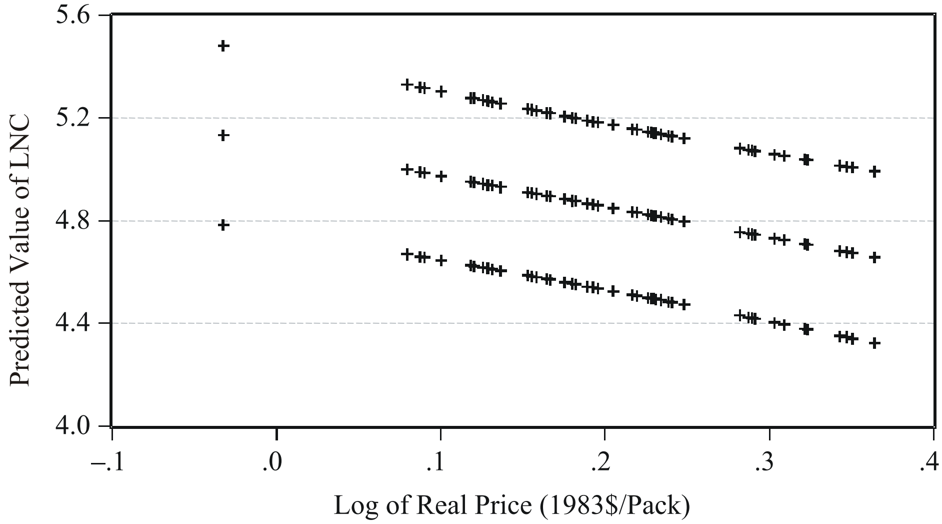

Figure 3.10 95% Confidence Band for Predicted Values

Problems

1. For the simple regression with a constant $Y_i = \alpha + \beta X_i + u_i$, given in equation (3.1) verify the following *numerical properties* of the OLS estimator:

$$\sum_{i=1}^{n} e_i = 0, \sum_{i=1}^{n} e_i X_i = 0, \sum_{i=1}^{n} e_i \widehat{Y}_i = 0, \sum_{i=1}^{n} \widehat{Y}_i = \sum_{i=1}^{n} Y_i$$

2. For the *regression with only a constant* $Y_i = \alpha + u_i$ with $u_i \sim \text{IID}(0, \sigma^2)$, show that the least squares estimate of $\widehat{\alpha}$ is $\widehat{\alpha}_{OLS} = \bar{Y}$, $\text{var}(\widehat{\alpha}_{OLS}) = \sigma^2/n$, and the residual sums of squares is $\sum_{i=1}^{n} y_i^2 = \sum_{i=1}^{n}(Y_i - \bar{Y})^2$.

3. For the simple regression *without* a constant $Y_i = \beta X_i + u_i$, with $u_i \sim \text{IID}(0, \sigma^2)$.

   (a) Derive the OLS estimator of β and find its variance.

   (b) What numerical properties of the OLS estimators described in problem 1 still hold for this model?

   (c) derive the maximum likelihood estimator of β and σ^2 under the assumption $u_i \sim \text{IIN}(0, \sigma^2)$.

   (d) Assume σ^2 is known. Derive the Wald, LM and LR tests for H_0; $\beta = 1$ versus H_1; $\beta \neq 1$.

4. Use the fact that $E(\sum_{i=1}^{n} x_i u_i)^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j E(u_i u_j)$; and assumptions 2 and 3 to prove equation (3.6).

5. Using the regression given in equation (3.1):

   (a) Show that $\widehat{\alpha}_{OLS} = \alpha + (\beta - \widehat{\beta}_{OLS})\bar{X} + \bar{u}$; and deduce that $E(\widehat{\alpha}_{OLS}) = \alpha$.

   (b) Using the fact that $\widehat{\beta}_{OLS} - \beta = \sum_{i=1}^{n} x_i u_i / \sum_{i=1}^{n} x_i^2$; use the results in part (a) to show that $\text{var}(\widehat{\alpha}_{OLS}) = \sigma^2[(1/n) + (\bar{X}^2/\sum_{i=1}^{n} x_i^2)] = \sigma^2 \sum_{i=1}^{n} X_i^2 / n \sum_{i=1}^{n} x_i^2$.

   (c) Show that $\widehat{\alpha}_{OLS}$ is consistent for α.

   (d) Show that $\text{cov}(\widehat{\alpha}_{OLS}, \widehat{\beta}_{OLS}) = -\bar{X}\text{var}(\widehat{\beta}_{OLS}) = -\sigma^2 \bar{X}/\sum_{i=1}^{n} x_i^2$. This means that the sign of the covariance is determined by the sign of $\bar{X}$. If $\bar{X}$ is positive, this covariance will be negative. This also means that if $\widehat{\alpha}_{OLS}$ is over-estimated, $\widehat{\beta}_{OLS}$ will be under-estimated.

6. Using the regression given in equation (3.1):
   (a) Prove that $\widehat{\alpha}_{OLS} = \sum_{i=1}^n \lambda_i Y_i$ where $\lambda_i = (1/n) - \bar{X}w_i$ and $w_i = x_i / \sum_{i=1}^n x_i^2$.
   (b) Show that $\sum_{i=1}^n \lambda_i = 1$ and $\sum_{i=1}^n \lambda_i X_i = 0$.
   (c) Prove that any other linear estimator of α, say $\widetilde{\alpha} = \sum_{i=1}^n b_i Y_i$ must satisfy $\sum_{i=1}^n b_i = 1$ and $\sum_{i=1}^n b_i X_i = 0$ for $\widetilde{\alpha}$ to be unbiased for α.
   (d) Let $b_i = \lambda_i + f_i$; show that $\sum_{i=1}^n f_i = 0$ and $\sum_{i=1}^n f_i X_i = 0$.
   (e) Prove that $\text{var}(\widetilde{\alpha}) = \sigma^2 \sum_{i=1}^n b_i^2 = \sigma^2 \sum_{i=1}^n \lambda_i^2 + \sigma^2 \sum_{i=1}^n f_i^2 = \text{var}(\widehat{\alpha}_{OLS}) + \sigma^2 \sum_{i=1}^n f_i^2$.

7. (a) Differentiate (3.9) with respect to α and β and show that $\widehat{\alpha}_{MLE} = \widehat{\alpha}_{OLS}$, $\widehat{\beta}_{MLE} = \widehat{\beta}_{OLS}$.
   (b) Differentiate (3.9) with respect to σ^2 and show that $\widehat{\sigma}^2_{MLE} = \sum_{i=1}^n e_i^2/n$.

8. *The t-Statistic in a Simple Regression.* It is well known that a standard normal random variable $N(0,1)$ divided by a square root of a chi-squared random variable divided by its degrees of freedom $(\chi^2_\nu/\nu)^{\frac{1}{2}}$ results in a random variable that is t-distributed with ν degrees of freedom, provided the $N(0,1)$ and the χ^2 variables are independent, see Chapter 2. Use this fact to show that $(\widehat{\beta}_{OLS} - \beta)/[s/(\sum_{i=1}^n x_i^2)^{\frac{1}{2}}] \sim t_{n-2}$.

9. *Relationship Between R^2 and r^2_{xy}.*
   (a) Using the fact that $R^2 = \sum_{i=1}^n \widehat{y}_i^2 / \sum_{i=1}^n y_i^2$; $\widehat{y}_i = \widehat{\beta}_{OLS} x_i$; and $\widehat{\beta}_{OLS} = \sum_{i=1}^n x_i y_i / \sum_{i=1}^n x_i^2$, show that $R^2 = r^2_{xy}$ where,
   $$r^2_{xy} = (\textstyle\sum_{i=1}^n x_i y_i)^2 / (\sum_{i=1}^n x_i^2)(\sum_{i=1}^n y_i^2).$$
   (b) Using the fact that $y_i = \widehat{y}_i + e_i$, show that $\sum_{i=1}^n \widehat{y}_i y_i = \sum_{i=1}^n \widehat{y}_i^2$, and hence, deduce that $r^2_{y\widehat{y}} = (\sum_{i=1}^n y_i \widehat{y}_i)^2 / (\sum_{i=1}^n y_i^2)(\sum_{i=1}^n \widehat{y}_i^2)$ is equal to R^2.

10. *Prediction.* Consider the problem of predicting Y_0 from (3.11). Given X_0,
   (a) Show that $E(Y_0) = \alpha + \beta X_0$.
   (b) Show that $\widehat{Y}_0$ is unbiased for $E(Y_0)$.
   (c) Show that $\text{var}(\widehat{Y}_0) = \text{var}(\widehat{\alpha}_{OLS}) + X_0^2 \text{var}(\widehat{\beta}_{OLS}) + 2X_0 \text{cov}(\widehat{\alpha}_{OLS}, \widehat{\beta}_{OLS})$. Deduce that $\text{var}(\widehat{Y}_0) = \sigma^2[(1/n) + (X_0 - \bar{X})^2 / \sum_{i=1}^n x_i^2]$.
   (d) Consider a linear predictor of $E(Y_0)$, say $\widetilde{Y}_0 = \sum_{i=1}^n a_i Y_i$, show that $\sum_{i=1}^n a_i = 1$ and $\sum_{i=1}^n a_i X_i = X_0$ for this predictor to be unbiased for $E(Y_0)$.
   (e) Show that the $\text{var}(\widetilde{Y}_0) = \sigma^2 \sum_{i=1}^n a_i^2$. Minimize $\sum_{i=1}^n a_i^2$ subject to the restrictions given in (d). Prove that the resulting predictor is $\widetilde{Y}_0 = \widehat{\alpha}_{OLS} + \widehat{\beta}_{OLS} X_0$ and that the minimum variance is $\sigma^2[(1/n) + (X_0 - \bar{X})^2 / \sum_{i=1}^n x_i^2]$.

11. *Optimal Weighting of Unbiased Estimators.* This is based on Baltagi (1995). For the simple regression without a constant $Y_i = \beta X_i + u_i, i = 1, 2, \ldots, N$; where β is a scalar and $u_i \sim \text{IID}(0, \sigma^2)$ independent of X_i. Consider the following three unbiased estimators of β:
   $$\widehat{\beta}_1 = \textstyle\sum_{i=1}^n X_i Y_i / \sum_{i=1}^n X_i^2, \ \widehat{\beta}_2 = \bar{Y}/\bar{X}$$
   and
   $$\widehat{\beta}_3 = \textstyle\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) / \sum_{i=1}^n (X_i - \bar{X})^2,$$
   where $\bar{X} = \sum_{i=1}^n X_i/n$ and $\bar{Y} = \sum_{i=1}^n Y_i/n$.
   (a) Show that $\text{cov}(\widehat{\beta}_1, \widehat{\beta}_2) = \text{var}(\widehat{\beta}_1) > 0$, and that $\rho_{12} =$ (the correlation coefficient of $\widehat{\beta}_1$ and $\widehat{\beta}_2$) $= [\text{var}(\widehat{\beta}_1)/\text{var}(\widehat{\beta}_2)]^{\frac{1}{2}}$ with $0 < \rho_{12} \leq 1$. Show that the optimal combination of $\widehat{\beta}_1$ and $\widehat{\beta}_2$, given by $\widehat{\beta} = \alpha\widehat{\beta}_1 + (1-\alpha)\widehat{\beta}_2$ where $-\infty < \alpha < \infty$ occurs at $\alpha^* = 1$. Optimality here refers to minimizing the variance. **Hint**: Read the paper by Samuel-Cahn (1994).

(b) Similarly, show that $\text{cov}(\widehat{\beta}_1, \widehat{\beta}_3) = \text{var}(\widehat{\beta}_1) > 0$, and that ρ_{13} = (the correlation coefficient of $\widehat{\beta}_1$ and $\widehat{\beta}_3$) $= [\text{var}(\widehat{\beta}_1)/\text{var}(\widehat{\beta}_3)]^{\frac{1}{2}} = (1-\rho_{12}^2)^{\frac{1}{2}}$ with $0 < \rho_{13} < 1$. Conclude that the optimal combination $\widehat{\beta}_1$ and $\widehat{\beta}_3$ is again $\alpha^* = 1$.

(c) Show that $\text{cov}(\widehat{\beta}_2, \widehat{\beta}_3) = 0$ and that optimal combination of $\widehat{\beta}_2$ and $\widehat{\beta}_3$ is $\widehat{\beta} = (1-\rho_{12}^2)\widehat{\beta}_3 + \rho_{12}^2\widehat{\beta}_2 = \widehat{\beta}_1$. This exercise demonstrates a more general result, namely that the BLUE of β in this case $\widehat{\beta}_1$, has a positive correlation with any other linear unbiased estimator of β, and that this correlation can be easily computed from the ratio of the variances of these two estimators.

12. *Efficiency as Correlation.* This is based on Oksanen (1993). Let $\widehat{\beta}$ denote the Best Linear Unbiased Estimator of β and let $\widetilde{\beta}$ denote any linear unbiased estimator of β. Show that the relative efficiency of $\widetilde{\beta}$ with respect to $\widehat{\beta}$ is the squared correlation coefficient between $\widehat{\beta}$ and $\widetilde{\beta}$. **Hint**: Compute the variance of $\widehat{\beta} + \lambda(\widehat{\beta} - \widetilde{\beta})$ for any λ. This variance is minimized at $\lambda = 0$ since $\widehat{\beta}$ is BLUE. This should give you the result that $E(\widehat{\beta}^2) = E(\widehat{\beta}\widetilde{\beta})$ which in turn proves the required result, see Zheng (1994).

13. For the numerical illustration given in section 3.9, what happens to the least squares regression coefficient estimates $(\widehat{\alpha}_{OLS}, \widehat{\beta}_{OLS})$, s^2, the estimated $se(\widehat{\alpha}_{OLS})$ and $se(\widehat{\beta}_{OLS})$, t-statistic for $\widehat{\alpha}_{OLS}$ and $\widehat{\beta}_{OLS}$ for $H_0^a; \alpha = 0$, and $H_0^b; \beta = 0$ and R^2 when:

(a) Y_i is regressed on $X_i + 5$ rather than X_i. In other words, we add a constant 5 to each observation of the explanatory variable X_i and rerun the regression. It is very instructive to see how the computations in Table 3.1 are affected by this simple transformation on X_i.

(b) $Y_i + 2$ is regressed on X_i. In other words, a constant 2 is added to Y_i.

(c) Y_i is regressed on $2X_i$. (A constant 2 is multiplied by X_i).

14. For the cigarette consumption data given in Table 3.2.

(a) Give the descriptive statistics for $\log C$, $\log P$ and $\log Y$. Plot their histogram. Also, plot $\log C$ versus $\log Y$ and $\log C$ versus $\log P$. Obtain the correlation matrix of these variables.

(b) Run the regression of $\log C$ on $\log Y$. What is the income elasticity estimate? What is its standard error? Test the null hypothesis that this elasticity is zero. What is the s and R^2 of this regression?

(c) Show that the square of the simple correlation coefficient between $\log C$ and $\log Y$ is equal to R^2. Show that the square of the correlation coefficient between the fitted and actual values of $\log C$ is also equal to R^2.

(d) Plot the residuals versus income. Also, plot the fitted values along with their 95% confidence band.

15. Consider the simple regression with no constant: $Y_i = \beta X_i + u_i \quad i = 1, 2, \ldots, n$

where $u_i \sim \text{IID}(0, \sigma^2)$ independent of X_i. Theil (1971) showed that among all *linear* estimators in Y_i, the *minimum mean square* estimator for β, i.e., that which minimizes $E(\widetilde{\beta} - \beta)^2$ is given by

$$\widetilde{\beta} = \beta^2 \textstyle\sum_{i=1}^n X_i Y_i / (\beta^2 \sum_{i=1}^n X_i^2 + \sigma^2).$$

(a) Show that $E(\widetilde{\beta}) = \beta/(1+c)$, where $c = \sigma^2/\beta^2 \sum_{i=1}^n X_i^2 > 0$.

(b) Conclude that the Bias $(\widetilde{\beta}) = E(\widetilde{\beta}) - \beta = -[c/(1+c)]\beta$. Note that this bias is positive (negative) when β is negative (positive). This also means that $\widetilde{\beta}$ is biased towards zero.

(c) Show that $\text{MSE}(\widetilde{\beta}) = E(\widetilde{\beta} - \beta)^2 = \sigma^2/[\sum_{i=1}^n X_i^2 + (\sigma^2/\beta^2)]$. Conclude that it is smaller than the $\text{MSE}(\widehat{\beta}_{OLS})$.

Table 3.4 Energy Data for 20 countries

Country	**RGDP (in 10^6 1975 U.S.$'s)**	**EN 10^6 Kilograms Coal Equivalents**
Malta	1251	456
Iceland	1331	1124
Cyprus	2003	1211
Ireland	11788	11053
Norway	27914	26086
Finland	28388	26405
Portugal	30642	12080
Denmark	34540	27049
Greece	38039	20119
Switzerland	42238	23234
Austria	45451	30633
Sweden	59350	45132
Belgium	62049	58894
Netherlands	82804	84416
Turkey	91946	32619
Spain	159602	88148
Italy	265863	192453
U.K.	279191	268056
France	358675	233907
W. Germany	428888	352.677

16. Table 3.4 gives cross-section Data for 1980 on real gross domestic product (RGDP) and aggregate energy consumption (EN) for 20 countries

    (a) Enter the data and provide descriptive statistics. Plot the histograms for RGDP and EN. Plot EN versus RGDP.

    (b) Estimate the regression:

    $$\log(En) = \alpha + \beta \log(RGDP) + u.$$

    Be sure to plot the residuals. What do they show?

    (c) Test $H_0; \beta = 1$.

    (d) One of your Energy data observations has a misplaced decimal. Multiply it by 1000. Now repeat parts (a), (b) and (c).

    (e) Was there any reason for ordering the data from the lowest to highest energy consumption? Explain.

    Lesson Learned: Always plot the residuals. Always check your data very carefully.

17. Using the Energy Data given in Table 3.4, corrected as in problem 16 part (d), is it legitimate to reverse the form of the equation?

    $$log(RDGP) = \gamma + \delta log(En) + \epsilon$$

    (a) Economically, does this change the interpretation of the equation? Explain.

    (b) Estimate this equation and compare R^2 of this equation with that of the previous problem. Also, check if $\widehat{\delta} = 1/\widehat{\beta}$. Why are they different?

(c) Statistically, by reversing the equation, which assumptions do we violate?

(d) Show that $\widehat{\delta}\widehat{\beta} = R^2$.

(e) *Effects of changing units in which variables are measured.* Suppose you measured energy in BTU's instead of kilograms of coal equivalents so that the original series was multiplied by 60. How does it change α and β in the following equations?

$$\log(En) = \alpha + \beta\log(RDGP) + u \qquad En = \alpha^* + \beta^* RGDP + \nu$$

Can you explain why $\widehat{\alpha}$ changed, but not $\widehat{\beta}$ for the log-log model, whereas both $\widehat{\alpha}^*$and $\widehat{\beta}^*$ changed for the linear model?

(f) For the log-log specification and the linear specification, compare the GDP elasticity for Malta and W. Germany. Are both equally plausible?

(g) Plot the residuals from both linear and log-log models. What do you observe?

(h) Can you compare the R^2 and standard errors from both models in part (g)? **Hint**: Retrieve $\log(En)$ and $\widehat{\log}(\mathrm{E}n)$ in the log-log equation, exponentiate, then compute the residuals and s. These are comparable to those obtained from the linear model.

18. For the model considered in problem 16: $\log(En) = \alpha + \beta\log(RGDP) + u$ and measuring energy in BTU's (like part (e) of problem 17).

(a) What is the 95% confidence prediction interval at the sample mean?

(b) What is the 95% confidence prediction interval for Malta?

(c) What is the 95% confidence prediction interval for West Germany?

References

Additional readings on the material covered in this chapter can be found in:

Baltagi, B.H. (1995), "Optimal Weighting of Unbiased Estimators," *Econometric Theory*, Problem 95.3.1, 11:637.

Baltagi, B.H. and D. Levin (1992), "Cigarette Taxation: Raising Revenues and Reducing Consumption," *Structural Change and Economic Dynamics*, 3: 321–335.

Belsley, D.A., E. Kuh and R.E. Welsch (1980), *Regression Diagnostics* (Wiley: New York).

Greene, W. (1993), *Econometric Analysis* (Macmillian: New York).

Gujarati, D. (1995), *Basic Econometrics* (McGraw-Hill: New York).

Johnston, J. (1984), *Econometric Methods* (McGraw-Hill: New York).

Kelejian, H. and W. Oates (1989), *Introduction to Econometrics* (Harper and Row: New York).

Kennedy, P. (1992), *A Guide to Econometrics* (MIT Press: Cambridge).

Kmenta, J. (1986), *Elements of Econometrics* (Macmillan: New York).

Maddala, G.S. (1992), *Introduction to Econometrics* (Macmillan: New York).

Oksanen, E.H. (1993), "Efficiency as Correlation," *Econometric Theory*, Problem 93.1.3, 9: 146.

Samuel-Cahn, E. (1994), "Combining Unbiased Estimators," *The American Statistician*, 48: 34–36.

Wallace, D. and L. Silver (1988), *Econometrics: An Introduction* (Addison Wesley: New York).

Zheng, J.X. (1994), "Efficiency as Correlation," *Econometric Theory*, Solution 93.1.3, 10: 228.

Appendix
Centered and Uncentered R^2

From the OLS regression on (3.1) we get

$$Y_i = \widehat{Y}_i + e_i \quad i = 1, 2, \ldots, n \tag{A.1}$$

where $\widehat{Y}_i = \widehat{\alpha}_{OLS} + X_i\widehat{\beta}_{OLS}$. Squaring and summing the above equation we get

$$\sum_{i=1}^n Y_i^2 = \sum_{i=1}^n \widehat{Y}_i^2 + \sum_{i=1}^n e_i^2 \tag{A.2}$$

since $\sum_{i=1}^n \widehat{Y}_i e_i = 0$. The *uncentered* R^2 is given by

$$\text{uncentered } R^2 = 1 - \sum_{i=1}^n e_i^2 / \sum_{i=1}^n Y_i^2 = \sum_{i=1}^n \widehat{Y}_i^2 / \sum_{i=1} Y_i^2 \tag{A.3}$$

Note that the total sum of squares for Y_i is *not* expressed in deviation from the sample mean $\bar{Y}$. In other words, the uncentered R^2 is the proportion of variation of $\sum_{i=1}^n Y_i^2$ that is explained by the regression on X. Regression packages usually report the *centered* R^2 which was defined in section 3.6 as $1 - (\sum_{i=1}^n e_i^2 / \sum_{i=1}^n y_i^2)$ where $y_i = Y_i - \bar{Y}$. The latter measure focuses on explaining the variation in Y_i *after* fitting the constant.

From section 3.6, we have seen that a naive model with only a constant in it gives $\bar{Y}$ as the estimate of the constant, see also problem 2. The variation in Y_i that is not explained by this naive model is $\sum_{i=1}^n y_i^2 = \sum_{i=1}^n (Y_i - \bar{Y})^2$. Subtracting $n\bar{Y}^2$ from both sides of (A.2) we get

$$\sum_{i=1}^n y_i^2 = \sum_{i=1}^n \widehat{Y}_i^2 - n\bar{Y}^2 + \sum_{i=1}^n e_i^2$$

and the centered R^2 is

$$\text{centered } R^2 = 1 - (\sum_{i=1}^n e_i^2 / \sum_{i=1}^n y_i^2) = (\sum_{i=1}^n \widehat{Y}_i^2 - n\bar{Y}^2) / \sum_{i=1}^n y_i^2 \tag{A.4}$$

If there is a constant in the model $\bar{Y} = \overline{\widehat{Y}}$, see section 3.6, and $\sum_{i=1}^n \widehat{y}_i^2 = \sum_{i=1}^n (\widehat{Y}_i - \overline{\widehat{Y}})^2 = \sum_{i=1}^n \widehat{Y}_i^2 - n\bar{Y}^2$. Therefore, the centered $R^2 = \sum_{i=1}^n \widehat{y}_i^2 / \sum_{i=1}^n y_i^2$ which is the R^2 reported by regression packages. If there is no constant in the model, some regression packages give you the option of (no constant) and the R^2 reported is usually the uncentered R^2. Check your regression package documentation to verify what you are getting. We will encounter uncentered R^2 again in constructing test statistics using regressions, see for example Chapter 11.