OLS Lecture 2

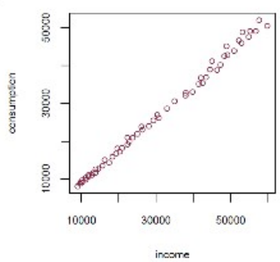
MPC Example

$$C = a + MPC \times Y \qquad (4.3)$$

This is another economic model represented by a "straight line".

- a – intercept
- b – slope

Figure 4.3: Income and consumption in the U.K. (Verbeek and Marno, 2008).



Demand for liquor

How much less alcohol will people consume if we raise the price? In first-year microeconomics you learned about the law of demand. The quantity demanded of a product should depend on its price (and other things):

$$Q_d = a + bP \qquad (4.1)$$

- a intercept
- b slope (should be negative)

Figure 4.1: A typical demand "curve". Note this is an "inverse" demand curve (quantity demanded is on the vertical axis, and price on the horizontal axis).

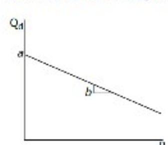
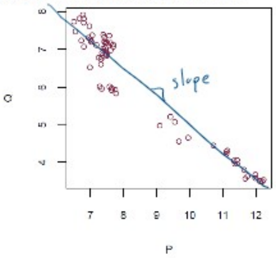


Figure 4.2: Per capita consumption, and price, of spirits. Choosing a line through the data necessarily chooses the slope of the line, b, which determines how much Q_d decreases for an increase in P.



Econometric model (population model):

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \tag{4.1}$$

Notation

- X is called the ~~*independent*~~ explanatory variable or *regressor*. It is the variable that is assumed to *cause the* Y *variable*. In the "Demand for Liquor" example, this variable was *price* (P). See equation 4.1. In the MPC example the regressor was *income*. See equation 4.3.
- Y is the *dependent* variable. This variable is assumed to be caused by X (it *depends* on X). In the demand example the dependent variable was *quantity demanded* (Q_d) and in the MPC example it was *consumption* (C).

- β_0 is the population intercept. It was labelled a in both examples. It is unobservable, but we can try to estimate it.
- β_1 is the population slope. When X increases by 1, Y increases by β_1. This is the primary object of interest, and is unobservable. We want to estimate β_1. β_1 is interpreted as the marginal effect in many economics models.
- ϵ is the regression *error* term. It consists of all the other factors or variables that determine Y, other than the X variable. All of these other variables causing Y are combined into ϵ. ϵ is considered to be a random variable since we can not observe it.
- $i = 1, \ldots, n$. The subscript i denotes the observation. n is the sample size. For example, Y_4 refers to the fourth Y observation in the data set.

Y is directly related to X

$Y = 1X$

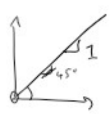
4.3.1 The importance of β_1

Note that in equation 4.4, the object of interest is β_1. It is the thing we are trying to estimate. It is the causal, or marginal effect, of X on Y. That is, a change in X of ΔX causes a β_1 change in Y:

$$\frac{\Delta Y}{\Delta X} = \beta_1$$

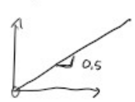
$$Y = \beta_0 + \beta_1 X + \epsilon$$

$$\frac{\partial Y}{\partial X} = 0 + \beta_1 + 0 = \beta_1$$



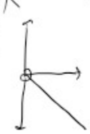
$Y = cX$

$\downarrow$ 0.5



$Y = dX$

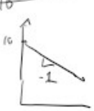
$\downarrow$ -1

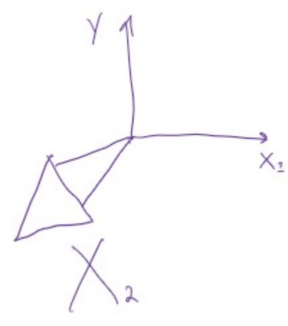


$Y = \beta_0 + \beta_1 X$

$Y = f + dX$

10





$\epsilon \longrightarrow Y$

random random

4.3.2 The importance of ϵ

ϵ (epsilon) is the random component of the model. Without ϵ, statistics/econometrics is not required. ϵ represents all of the other things that determine Y, other than X. They are all added up and lumped into this one random variable. Because we can not observe all of these other factors, we consider them to be random. The fact that ϵ is random makes Y random as well.

ϵ (epsilon) is the random component of the model. Without ϵ, statistics/econometrics is not required. ϵ represents all of the other things that determine Y, other than X. They are all added up and lumped into this one random variable. Because we can not observe all of these other factors, we consider them to be random. The fact that ϵ is random makes Y random as well.

Later, we will make some assumptions about the randomness of ϵ, that will ultimately determine the properties of the way that we choose to estimate β_1.

data generation process

4.3.3 Why it's called a population model

Equation 4.4 is called a "population" model because it represents the true, but unknown way in which the Y variable is "created" or "determined". β_0 and β_1 are unknown (and so is ϵ). We will observe a sample of Y and X, and use the sample to try to figure out the βs.

4.4 The estimated model

Our primary goal is to estimate β_1 (the marginal effect of X on Y), but to do so we'll also have to estimate β_0. This estimated intercept and slope will define a straight line. These estimates will be denoted b_0 and b_1, the OLS intercept and slope.

Let's start with a very simple example using data that I made up: $Y = \{1, 4, 5, 4\}$, $X = \{2, 4, 6, 8\}$. The data, and estimated OLS line, are shown in figure 4.4. The OLS estimated intercept is $b_0 = 1$, and the estimated slope is $b_1 = 0.5$.

Truth

$Y = \beta_0 + \beta_1 X + \epsilon$

unknown

Estimate

$Y = b_0 + b_1 X + e$

estimated

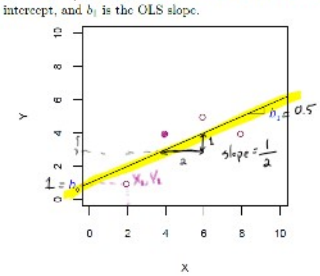
$\boxed{Y = \beta_0 + \beta_1 X + \epsilon}$

can be (-)

$Y = 10 + (-6)X + \epsilon$

$= 10 - 6X + \epsilon$

Figure 4.4: A simple data set with the estimated OLS line in blue. b_0 is the OLS intercept, and b_1 is the OLS slope.



4.4.1 OLS predicted values ($\hat{Y}_i$)

The OLS predicted (or fitted) values, are the values for Y that we get when we "plug" the X values back into the estimated OLS line. These predicted Y values are denoted by $\hat{Y}$. We can find each predicted value, $\hat{Y}_i$, by plugging each X_i into the estimated equation.

In general, the estimated equation (or line) is written as:

$X = \{2, 4, 6, 8\}$

$$\hat{Y}_i = b_0 + b_1 X_i. \tag{4.5}$$

For our simple example, equation 4.5 becomes $\hat{Y}_i = 1 + 0.5X_i$, and each OLS predicted values is:

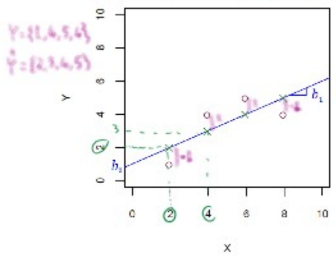
$\hat{Y}_1 = 1 + 0.5(2) = 2$

$\hat{Y}_2 = 1 + 0.5(4) = 3$

$\hat{Y}_3 = 1 + 0.5(6) = 4$

$\hat{Y}_4 = 1 + 0.5(8) = 5$

Figure 4.5: The OLS predicted values shown by ×.



4.4.2 OLS residuals (e_i)

An OLS predicted value tells us what the estimated model predicts for Y when given a particular value of X. When we plug in the sample values for X (as we did in the previous section), we see that the predicted values ($\hat{Y}_i$) don't quite line up with the actual Y_i values. The differences between the two are the OLS *residuals*. The OLS residuals are like prediction errors, and are determined by:

reality − prediction

$$e_i = Y_i - \hat{Y}_i \tag{4.6}$$

Using equation 4.6 for our simple example, each OLS residual is:

$Y = \{1, 4, 5, 4\}$

$\hat{Y} = \{2, 3, 4, 5\}$

$e_1 = 1 - 2 = -1$

$e_2 = 4 - 3 = 1$

$e_3 = 5 - 4 = 1$

$e_4 = 4 - 5 = -1$

$\sum e_i = 0$ × NO

want distance

$\sum |e_i|$ NOT the LS estimator

$\sum e_i^2 = 4$

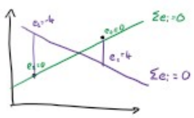
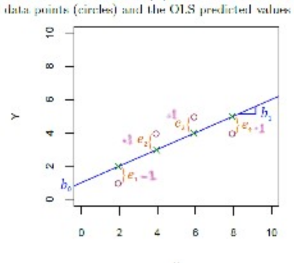


Figure 4.6: The OLS residuals (e_i) are the vertical distances between the actual data points (circles) and the OLS predicted values (×).



least squares

How to choose the OLS line

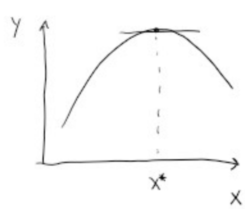
The OLS estimators are defined in the following way. They are the values for b_0 and b_1 that minimize the sum of squared vertical distances between the OLS line and the actual data points (Y_i). These vertical distances have already been defined as the OLS residuals (e_i). So the "objective" is to choose b_0 and b_1 so that $\sum_{i=1}^{n} e_i^2$ is minimized. This is an optimization problem from calculus. Formally stated, the OLS estimator is the solution to the minimization problem:

minimization

$$\min_{b_0, b_1} \sum_{i=1}^{n} e_i^2 \tag{4.8}$$

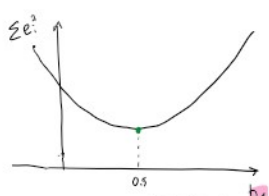
$\min_{b_0,b_1} \sum_{i=1}^{n} e_i^2$

$\sum e_i^2 = 4$ in the example



$\frac{\partial y}{\partial x} = 0$ FOC

↳ solve to get x^*



$\frac{\partial \sum e_i^2}{\partial b_1} = 0$

↳ solve for b_1

why not $\min \sum |e_i|$? → LAD

$\min \sum e_i^4$

$\min \sum e_i^6$

why not horizontal distances?

why use $\bar{Y}$ to get μ, instead of median/mode/

↳ unbiased

↳ efficient

↳ consistent

The solution:

$$\boxed{b_1 = \frac{\sum_{i=1}^{n}\left[\left(Y_i - \bar{Y}\right)\left(X_i - \bar{X}\right)\right]}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}} \tag{4.10}$$

$$\boxed{b_1 = \frac{\sum_{i=1}^{n}\left[\left(Y_i - \bar{Y}\right)\left(X_i - \bar{X}\right)\right]}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}} \quad (4.10)$$

$$\boxed{b_0 = \bar{Y} - b_1\bar{X}}$$

↘ lm(Y~X)

min Σe_i^4

min Σe_i^6

why not horizontal distances?

orthogonal ?

↳ unbiased

↳ efficient

↳ consistent

↳ also why we min Σe_i^2

↳ Least "squares"