Introductory Econometrics

Lecture 2: The Simple Regression Model

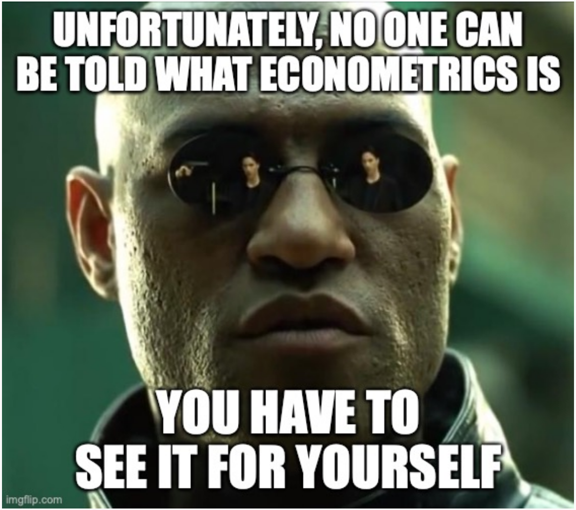
Alex Alekseev

October 23, 2022

University of Regensburg, Department of Economics

Previously on *Introductory Econometrics*...

- What is Econometrics?
- A Trade Example: What Determines Trade Flows?
- Economic Models and the Need for Econometrics
- Causation vs. Correlation
- Types of Economic Data


UNFORTUNATELY, NO ONE CAN
BE TOLD WHAT ECONOMETRICS IS
YOU HAVE TO
SEE IT FOR YOURSELF
imgflip.com

The Population Regression Model

Set-Up

- Suppose we have two variables, X and Y, and we want to study how Y varies with X or to predict Y for given values of X

Set-Up

- Suppose we have two variables, X and Y, and we want to study how Y varies with X or to predict Y for given values of X

Example 1

How much does the hourly wage (Y) change with one more year of schooling (X)

Set-Up

- Suppose we have two variables, X and Y, and we want to study how Y varies with X or to predict Y for given values of X

Example 2

What is the predicted import to Germany (Y) from an exporter with a GDP of $100bn ($X$)T

Set-Up

- Suppose we have two variables, X and Y, and we want to study how Y varies with X or to predict Y for given values of X
- If we had perfect knowledge, the we could express the relationship between X and Y as

$$Y = f(X, Z_1, \ldots, Z_s),$$

  where $Z_1, \ldots, Z_s$ are some other variables affecting Y, in addition to X

Practical Issues

- The relationship $Y = f(X, Z_1, \ldots, Z_s)$ might be too complicated to be useful
- There might not exist an exact relationship
- There exists an exact relationship but not all the variables can be observed
- We do not know $f(\cdot)$

Our solution: focus on a relationship that holds **on average**

Random Variables

- We will treat Y as a **random variable** (recall probability theory)
- It's values represent the results of a random choice from all the units in the population

Random Variables

- We will treat Y as a **random variable** (recall probability theory)
- It's values represent the results of a random choice from all the units in the population

Example 1

The population consists of all the apartments in Regensburg. A value of Y represents a rent of a single apartment randomly chosen from all the apartments.

Random Variables

- We will treat Y as a **random variable** (recall probability theory)
- It's values represent the results of a random choice from all the units in the population

Example 2

The population consists of all the possible values of imports to Germany from a specific country and period. A value of Y represents a single randomly chosen value of import.

Random Variables

- We will treat Y as a **random variable** (recall probability theory)
- It's values represent the results of a random choice from all the units in the population

Example 3

For a die, the population consists of all the numbers written on each side. A value of Y represents a single randomly generated outcome of a die roll.

Random Variables

- We will treat Y as a **random variable** (recall probability theory)
- It's values represent the results of a random choice from all the units in the population
- If Y is discrete it can be described by the set of **values** it takes and the **probabilities** of those values

Example: A Coin Flip

Outcome	Payoff	Probability
Heads	$+1$	0.5
Tails	-1	0.5

Expected Value

- We can describe the "average" value of a variable using its **expected value**, denoted as $\mathbb{E}[Y]$
- In the discrete case, the expected value of a variable (taking n values) is defined as

$$\mathbb{E}[Y] = y_1 \mathbb{P}(Y = y_1) + \ldots + y_n \mathbb{P}(Y = y_n)$$

Expected Value

- We can describe the "average" value of a variable using its **expected value**, denoted as $\mathbb{E}[Y]$
- In the discrete case, the expected value of a variable (taking n values) is defined as

$$\mathbb{E}[Y] = y_1\mathbb{P}(Y = y_1) + \ldots + y_n\mathbb{P}(Y = y_n)$$

Example: A Coin Flip

Outcome	Payoff	Probability
Heads	$+1$	0.5
Tails	-1	0.5

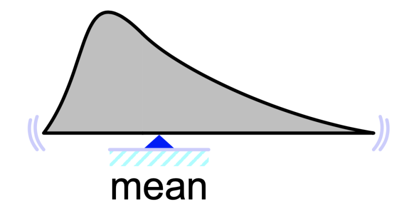
$$\mathbb{E}[Y] = 1 \times 0.5 + (-1) \times 0.5 = 0$$

Expected Value

- We can describe the "average" value of a variable using its **expected value**, denoted as $\mathbb{E}[Y]$
- In the discrete case, the expected value of a variable (taking n values) is defined as

$$\mathbb{E}[Y] = y_1 \mathbb{P}(Y = y_1) + \ldots + y_n \mathbb{P}(Y = y_n)$$

Expected Value: Intuition



Expected Value

- We can describe the "average" value of a variable using its **expected value**, denoted as $\mathbb{E}[Y]$
- In the discrete case, the expected value of a variable (taking n values) is defined as

$$\mathbb{E}[Y] = y_1\mathbb{P}(Y = y_1) + \ldots + y_n\mathbb{P}(Y = y_n)$$

Summation Notation

$$\sum_{i=1}^{n} y_i \equiv y_1 + y_2 + \ldots + y_n$$

$$\mathbb{E}[Y] = \sum_{i=1}^{n} y_i\mathbb{P}(Y = y_i)$$

Conditional Expectation

- For our analysis, we are interested in **conditional expectations**
- The conditional expectation of Y given $X = x$ is denoted as $\mathbb{E}[Y \mid X = x]$

Conditional Expectation

- For our analysis, we are interested in **conditional expectations**
- The conditional expectation of Y given $X = x$ is denoted as $\mathbb{E}[Y \mid X = x]$

Example

Suppose we only consider the apartments in Regensburg with a size of $x = 75m^2$. The expected rent for those apartments is $\mathbb{E}[Y \mid X = 75]$

Conditional Expectation

- For our analysis, we are interested in **conditional expectations**
- The conditional expectation of Y given $X = x$ is denoted as $\mathbb{E}[Y \mid X = x]$
- The conditional expectation is defined as

$$\mathbb{E}[Y \mid X = x] = \sum_{i=1}^{n} y_i \mathbb{P}(Y = y_i \mid X = x)$$

Conditional Expectation

- For our analysis, we are interested in **conditional expectations**
- The conditional expectation of Y given $X = x$ is denoted as $\mathbb{E}[Y \mid X = x]$
- The conditional expectation is defined as

$$\mathbb{E}[Y \mid X = x] = \sum_{i=1}^{n} y_i \mathbb{P}(Y = y_i \mid X = x)$$

Example: Conditional Probability of a Coin Flip

	Fair ($x = 1$)	Biased ($x = 2$)
$y = +1$	0.5	0.25
$y = -1$	0.5	0.75

$$\mathbb{E}[Y \mid X = 1] = 1 \times 0.5 + (-1) \times 0.5 = 0$$
$$\mathbb{E}[Y \mid X = 2] = 1 \times 0.25 + (-1) \times 0.75 = -0.5$$

Conditional Expectation Function

- We can treat the conditional expectation as a function of X

$$\mathbb{E}[Y \mid X] \equiv g(X)$$

Conditional Expectation Function

- We can treat the conditional expectation as a function of X

$$\mathbb{E}[Y \mid X] \equiv g(X)$$

Conditional Expectation Function (CEF)

Note that since X is a random variable, the conditional expectation function $\mathbb{E}[Y \mid X]$ is also a **random variable**. The value of $\mathbb{E}[Y \mid X]$ at a given $X = x$ is not a random variable.

Conditional Expectation Function

- We can treat the conditional expectation as a function of X

$$\mathbb{E}[Y \mid X] \equiv g(X)$$

- We can use the CEF to write the **population regression model**

$$Y = \mathbb{E}[Y \mid X] + u,$$

where $u \equiv Y - \mathbb{E}[Y \mid X = x]$ is the **error term**

The CEF Decomposition

Theorem: The CEF-Decomposition Property

Any random variable Y can be written as

$$Y = \mathbb{E}[Y \mid X] + u,$$

where u is mean-independent of X ($\mathbb{E}[u \mid X] = 0$) and u is uncorrelated with any function of X ($Cov(u, h(X)) = 0$)

The CEF Decomposition

Theorem: The CEF-Decomposition Property

Any random variable Y can be written as

$$Y = \mathbb{E}[Y \mid X] + u,$$

where u is mean-independent of X ($\mathbb{E}[u \mid X] = 0$) and u is uncorrelated with any function of X ($Cov(u, h(X)) = 0$)

- Mean-independence: $\mathbb{E}[u \mid X] = 0$

$$\mathbb{E}[u \mid X] = \mathbb{E}[Y - \mathbb{E}[Y \mid X] \mid X] = \mathbb{E}[Y \mid X] - \mathbb{E}[Y \mid X] = 0$$

The CEF Decomposition

Theorem: The CEF-Decomposition Property

Any random variable Y can be written as

$$Y = \mathbb{E}[Y \mid X] + u,$$

where u is mean-independent of X ($\mathbb{E}[u \mid X] = 0$) and u is uncorrelated with any function of X ($Cov(u, h(X)) = 0$)

The Law of Iterated Expectations

$$\mathbb{E}[\mathbb{E}[Y \mid X]] = \mathbb{E}[Y]$$

"Expectation kills the conditioning"

The CEF Decomposition

Theorem: The CEF-Decomposition Property

Any random variable Y can be written as

$$Y = \mathbb{E}[Y \mid X] + u,$$

where u is mean-independent of X ($\mathbb{E}[u \mid X] = 0$) and u is uncorrelated with any function of X ($Cov(u, h(X)) = 0$)

- Mean independence together with the Law of Iterated Expectations implies that

$$\mathbb{E}[u] = \mathbb{E}[\mathbb{E}[u \mid X]] = 0$$

The CEF Decomposition

Theorem: The CEF-Decomposition Property

Any random variable Y can be written as

$$Y = \mathbb{E}[Y \mid X] + u,$$

where u is mean-independent of X ($\mathbb{E}[u \mid X] = 0$) and u is uncorrelated with any function of X ($Cov(u, h(X)) = 0$)

Covariance

$$Cov(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

The CEF Decomposition

Theorem: The CEF-Decomposition Property

Any random variable Y can be written as

$$Y = \mathbb{E}[Y \mid X] + u,$$

where u is mean-independent of X ($\mathbb{E}[u \mid X] = 0$) and u is uncorrelated with any function of X ($Cov(u, h(X)) = 0$)

- u uncorrelated with any function of X: $Cov(u, h(X)) = 0$

$$\begin{aligned} Cov(u, h(X)) &= \mathbb{E}[uh(X)] - \mathbb{E}[u]\mathbb{E}[h(X)] = \mathbb{E}[\mathbb{E}[uh(X) \mid X]] \\ &= \mathbb{E}[h(X)\mathbb{E}[u \mid X]] = 0 \end{aligned}$$

The CEF Decomposition

Theorem: The CEF-Decomposition Property

Any random variable Y can be written as

$$Y = \mathbb{E}[Y \mid X] + u,$$

where u is mean-independent of X ($\mathbb{E}[u \mid X] = 0$) and u is uncorrelated with any function of X ($Cov(u, h(X)) = 0$)

Interpretation: Any random variable can be decomposed into a part that is "explained by X" (the CEF, the **systematic** part) and a part that is orthogonal to X (the error term, the **unsystematic** part)

Linear CEF Assumption

- How do we define the shape of the CEF?

Assumption

The CEF is linear

$$\mathbb{E}[Y \mid X] = \beta_0 + \beta_1 X$$

Linear CEF Assumption

- How do we define the shape of the CEF?

Assumption

The CEF is linear

$$\mathbb{E}[Y \mid X] = \beta_0 + \beta_1 X$$

- The assumption restricts the flexibility of the CEF
- The assumption is satisfied if there are other variables that affect Y linearly, as long as they are also linear in X

Other Variables

- Suppose that

$$\mathbb{E}[Y \mid X, Z] = \delta_0 + \delta_1 X + \delta_2 Z$$

and that

$$\mathbb{E}[Z \mid X] = \alpha_0 + \alpha_1 X$$

- Then the CEF is

$$\begin{aligned}
\mathbb{E}[Y \mid X] &= \sum_{i=1}^{k} \mathbb{E}[Y \mid X, Z = z_i]\mathbb{P}(Z = z_i \mid X) \\
&= \sum_{i=1}^{k} (\delta_0 + \delta_1 X + \delta_2 z_i)\mathbb{P}(Z = z_i \mid X) \\
&= \delta_0 + \delta_1 X + \delta_2 \mathbb{E}[Z \mid X] \\
&= \delta_0 + \delta_1 X + \delta_2(\alpha_0 + \alpha_1 X) \\
&= \underbrace{\gamma_0}_{\delta_0 + \delta_2 \alpha_0} + \underbrace{\gamma_1}_{\delta_1 + \delta_2 \alpha_1} X
\end{aligned}$$

Population Regression Model

- The assumption about the linearity of the CEF leads to following simple linear population regression model

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

Population Regression Model

- The assumption about the linearity of the CEF leads to following simple linear population regression model

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

Note

Recall that, by construction, the error term u is mean-independent of X, has zero expectation, and is uncorrelated with any function of X

Population Regression Model

- The assumption about the linearity of the CEF leads to following simple linear population regression model

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

- We typically call β_0 the **intercept** and β_1 the **slope**

More on Terminology

- The Y variable usually goes by the following names
  - dependent variable
  - outcome variable
  - response variable
  - regressand
- The X variable usually goes by the following names
  - independent variable
  - explanatory variable
  - regressor
  - covariate

Population Model and Simulations

- While in most cases we do not know the population model, we can specify our own model while doing **simulations**
- For example, suppose that X and u are two independent random draws from the set of numbers $\{-2.5, -1.5, -0.5, 0.5, 1, 1.5, 2.5\}$ (equally likely)
- Then we can specify our own population model (also called the **data-generating process** (DGP)), e.g., as

$$Y = 2 + 3X + u$$

- We can generate data using this model and then use these data to **estimate** the model

The Sample Regression Model

Estimation Sketch

A **sample**

X	Y
0.64	3.93
0.59	3.76
...	...
0.07	2.20
0.14	2.42
0.95	4.85

$(x_i, y_i)_{i=1}^n$

- random
- representative
- drawn from the population

Estimation Sketch

A **sample** $+$ An **estimator**

X	Y
0.64	3.93
0.59	3.76
...	...
0.07	2.20
0.14	2.42
0.95	4.85

$(x_i, y_i)_{i=1}^n$

- random
- representative
- drawn from the population

$f((x_i, y_i)_{i=1}^n)$

Estimation Sketch

A **sample** + An **estimator**

X	Y
0.64	3.93
0.59	3.76
...	...
0.07	2.20
0.14	2.42
0.95	4.85

$(x_i, y_i)_{i=1}^n$

- random
- representative
- drawn from the population

$$f((x_i, y_i)_{i=1}^n)$$

Note

There can be different alternative estimators

Estimation Sketch

A **sample** $+$ An **estimator** $\Rightarrow$ An **estimate**

X	Y
0.64	3.93
0.59	3.76
...	...
0.07	2.20
0.14	2.42
0.95	4.85

$(x_i, y_i)_{i=1}^n$

- random
- representative
- drawn from the population

$f((x_i, y_i)_{i=1}^n)$

$(\hat{\beta}_0, \hat{\beta}_1)$

Estimation Sketch

A **sample** $+$ An **estimator** $\Rightarrow$ An **estimate**

X	Y
0.64	3.93
0.59	3.76
...	...
0.07	2.20
0.14	2.42
0.95	4.85

$(x_i, y_i)_{i=1}^n$

- random
- representative
- drawn from the population

$f((x_i, y_i)_{i=1}^n)$

$(\hat{\beta}_0, \hat{\beta}_1)$

Note

The parameters you estimate, (β_0, β_1), are called an **estimand**

Sample Regression Model

- Our sample regression model can be written as

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{u}_i,$$

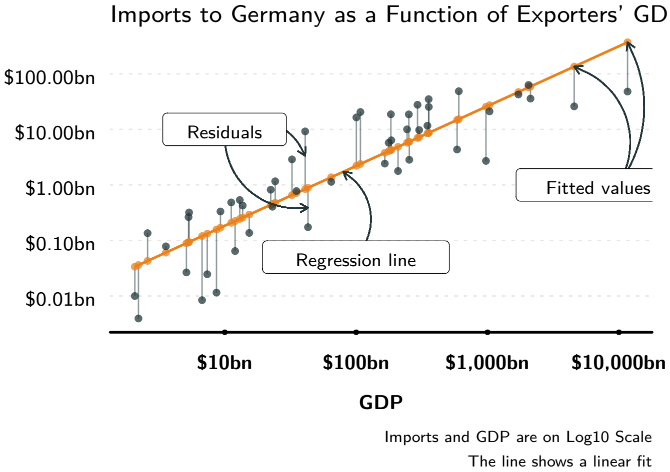
- The term

$$\hat{\beta}_0 + \hat{\beta}_1 x$$

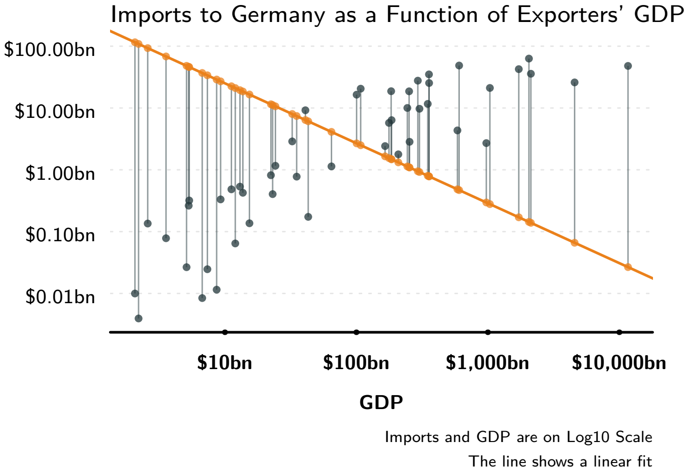
  is called the **sample regression function** or the **regression line**

- $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ are the **fitted** (or predicted) values and $\hat{u}_i = y_i - \hat{y}_i$ are the **residuals**

Trade Example



Bad Fit



Imports to Germany as a Function of Exporters' GDP
$100.00bn
$10.00bn
$1.00bn
$0.10bn
$0.01bn
$10bn
$100bn
$1,000bn
$10,000bn
GDP
Imports and GDP are on Log10 Scale
The line shows a linear fit

Ordinary Least Squares (OLS) Estimator

OLS Definition

- We can define the OLS estimator as the function that minimizes the **sum of squared residuals**
- The OLS estimates are then given by

$$(\hat{\beta}_0, \hat{\beta}_1) = \underset{(\beta_0, \beta_1)}{\arg\min} \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_i))^2$$

OLS Definition

- We can define the OLS estimator as the function that minimizes the **sum of squared residuals**
- The OLS estimates are then given by

$$(\hat{\beta}_0, \hat{\beta}_1) = \underset{(\beta_0, \beta_1)}{\arg\min} \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_i))^2$$

- If we define $g(\beta_0, \beta_1) \equiv \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_i))^2$, then the FONC implies that at the optimum

$$\frac{\partial g}{\partial \beta_0} = 0$$
$$\frac{\partial g}{\partial \beta_1} = 0$$

OLS Definition

- We can define the OLS estimator as the function that minimizes the **sum of squared residuals**
- The OLS estimates are then given by

$$(\hat{\beta}_0, \hat{\beta}_1) = \underset{(\beta_0, \beta_1)}{\arg\min} \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_i))^2$$

- If we define $g(\beta_0, \beta_1) \equiv \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_i))^2$, then the FONC implies that at the optimum

$$\sum_{i=1}^{n} 2(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))(-1) = 0$$
$$\sum_{i=1}^{n} 2(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))(-x_i) = 0$$

OLS Definition

- We can define the OLS estimator as the function that minimizes the **sum of squared residuals**
- The OLS estimates are then given by

$$(\hat{\beta}_0, \hat{\beta}_1) = \underset{(\beta_0, \beta_1)}{\arg\min} \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_i))^2$$

- If we define $g(\beta_0, \beta_1) \equiv \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_i))^2$, then the FONC implies that at the optimum

$$\sum_{i=1}^{n} (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)) = 0$$

$$\sum_{i=1}^{n} x_i (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) = 0$$

OLS Definition

- We can define the OLS estimator as the function that minimizes the **sum of squared residuals**
- The OLS estimates are then given by

$$(\hat{\beta}_0, \hat{\beta}_1) = \underset{(\beta_0,\beta_1)}{\arg\min} \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_i))^2$$

- If we define $g(\beta_0, \beta_1) \equiv \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_i))^2$, then the FONC implies that at the optimum

$$\sum_{i=1}^{n} (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)) = 0$$
$$\sum_{i=1}^{n} x_i(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) = 0$$

- These are called the **normal equations**

OLS Derivation

- From the first equation, we get

$$\begin{aligned}
\sum_{i=1}^{n}(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)) &= 0 \\
\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \hat{\beta}_0 - \sum_{i=1}^{n} \hat{\beta}_1 x_i &= 0 \\
\sum_{i=1}^{n} y_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^{n} x_i &= 0 \\
n\hat{\beta}_0 &= \sum_{i=1}^{n} y_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i \\
\hat{\beta}_0 &= \frac{1}{n}\sum_{i=1}^{n} y_i - \hat{\beta}_1 \frac{1}{n}\sum_{i=1}^{n} x_i \\
\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x}
\end{aligned}$$

- $\bar{y} \equiv \frac{1}{n}\sum_{i=1}^{n} y_i$ and $\bar{x} \equiv \frac{1}{n}\sum_{i=1}^{n} x_i$ are the **sample means** of Y and X

OLS Derivation

- Plugging this result into the second equation, we get

$$\sum_{i=1}^{n} x_i(y_i - \bar{y} + \hat{\beta}_1\bar{x} - \hat{\beta}_1 x_i) = 0$$

$$\sum_{i=1}^{n} x_i(y_i - \bar{y}) + \sum_{i=1}^{n} x_i(\hat{\beta}_1\bar{x} - \hat{\beta}_1 x_i) = 0$$

$$\sum_{i=1}^{n} x_i(y_i - \bar{y}) - \hat{\beta}_1 \sum_{i=1}^{n} x_i(x_i - \bar{x}) = 0$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

OLS Estimates

- The OLS estimates are given by

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

OLS Estimates

- The OLS estimates are given by

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$
$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Note

The function that takes the data $(x_i, y_i)_{i=1}^{n}$ as an input and returns the estimates $(\hat{\beta}_0, \hat{\beta}_1)$ is the OLS **estimator**. The value of the function for a given sample are the OLS **estimates**

OLS Estimates

- The OLS estimates are given by

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$
$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Note

Recall that $\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$ is the **sample covariance** between Y and X and $\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ is the **sample variance** of X

OLS Estimates

- The OLS estimates are given by

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$
$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

- Then we can rewrite $\hat{\beta}_1$ as

$$\hat{\beta}_1 = \frac{\widehat{Cov}(X, Y)}{\widehat{Var}(X)}$$

OLS Estimates

- The OLS estimates are given by

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$
$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

- Then we can rewrite $\hat{\beta}_1$ as

$$\hat{\beta}_1 = \frac{\widehat{Cov}(X, Y)}{\widehat{Var}(X)}$$

Assumption

The values of the independent variable X are not all identical, i.e., the sample variance is non-zero

OLS and the Method of Moments

- The OLS estimator is equivalent to some other popular estimators
- One of such estimators is the **Method of Moments** (MM) estimator
- MM works by specifying some **population moment conditions** and then replacing them with **sample moments**

OLS and the Method of Moments

- The OLS estimator is equivalent to some other popular estimators
- One of such estimators is the **Method of Moments** (MM) estimator
- MM works by specifying some **population moment conditions** and then replacing them with **sample moments**

Example

Population moment: $\mathbb{E}[X]$

Sample moment: $\frac{1}{n}\sum_{i=1}^{n} x_i$

OLS and the Method of Moments

- The OLS estimator is equivalent to some other popular estimators
- One of such estimators is the **Method of Moments** (MM) estimator
- MM works by specifying some **population moment conditions** and then replacing them with **sample moments**

Note

For this replacement to work, we need to assume that our data are a **random sample** from the population. Random sampling means that the values are pairwise **independent** and **identically distributed** (iid).

OLS and the Method of Moments

- The OLS estimator is equivalent to some other popular estimators
- One of such estimators is the **Method of Moments** (MM) estimator
- MM works by specifying some **population moment conditions** and then replacing them with **sample moments**
- Recall that our assumptions about the CEF imply two conditions

$$\mathbb{E}(u) = 0$$
$$Cov(u, X) = 0$$

OLS and the Method of Moments

- The OLS estimator is equivalent to some other popular estimators
- One of such estimators is the **Method of Moments** (MM) estimator
- MM works by specifying some **population moment conditions** and then replacing them with **sample moments**
- Recall that our assumptions about the CEF imply two conditions

$$
\begin{aligned}
\mathbb{E}[u] &= 0 \\
\mathbb{E}[uX] - \underbrace{\mathbb{E}[u]\mathbb{E}[X]}_{=0} &= 0
\end{aligned}
$$

OLS and the Method of Moments

- The OLS estimator is equivalent to some other popular estimators
- One of such estimators is the **Method of Moments** (MM) estimator
- MM works by specifying some **population moment conditions** and then replacing them with **sample moments**
- Recall that our assumptions about the CEF imply two conditions

$$\mathbb{E}[u] = 0$$
$$\mathbb{E}[uX] = 0$$

OLS and the Method of Moments

- The OLS estimator is equivalent to some other popular estimators
- One of such estimators is the **Method of Moments** (MM) estimator
- MM works by specifying some **population moment conditions** and then replacing them with **sample moments**
- Now replace the population moments with sample moments

$$\frac{1}{n}\sum_{i=1}^{n}\hat{u}_i = 0$$

$$\frac{1}{n}\sum_{i=1}^{n}\hat{u}_i x_i = 0$$

OLS and the Method of Moments

- The OLS estimator is equivalent to some other popular estimators
- One of such estimators is the **Method of Moments** (MM) estimator
- MM works by specifying some **population moment conditions** and then replacing them with **sample moments**
- Substitute for $\hat{u}_i$ and multiply by n

$$\sum_{i=1}^{n}(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)) = 0$$
$$\sum_{i=1}^{n}x_i(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)) = 0$$

OLS and the Method of Moments

- The OLS estimator is equivalent to some other popular estimators
- One of such estimators is the **Method of Moments** (MM) estimator
- MM works by specifying some **population moment conditions** and then replacing them with **sample moments**
- Substitute for $\hat{u}_i$ and multiply by n

$$\sum_{i=1}^{n}(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)) = 0$$
$$\sum_{i=1}^{n} x_i(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)) = 0$$

- These are the normal equations

Next Time on *Introductory Econometrics*...

Properties of the OLS estimator