EC508: Econometrics
Linear Regression with a Single Regressor

Jean-Jacques Forneron

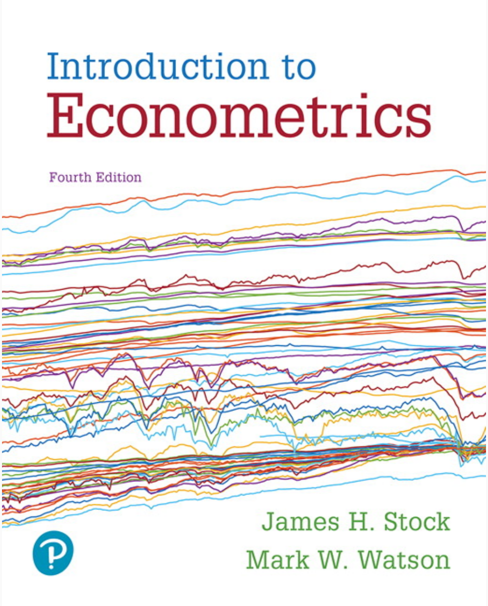
Spring, 2023

Boston University

Syllabus

- Lectures: (syllabus);
  Office Hours: Tu 5:00-6:00pm/Th 11:00-12:00pm
- Discussion Session: (syllabus)
  TA Office Hours: (syllabus)
- Problem Sets, Exams, Grading
  - Problem Sets: can work in groups, <u>individual submission</u>.
  - Midterm: TBA. Final: TBA.
    Both online, $\sim$1 week to complete
  - Final Grade =
    Problem Sets (20%) + Midterm (30%) + Final (50%).

Textbook: Introduction to Econometrics, Stock & Watson



Reference Textbook: Econometrics, Hayashi (more advanced)

Software: R using Rstudio

- Can be downloaded freely at https://www.rstudio.com/products/rstudio/download/#download
- Review session with TA

Linear Regression with a Single Regressor (SW 4-5)

- Regression Analysis: analyse <u>relationships</u> between economic variables *in a meaningful way*
- Interested in how some variables $X_1, \ldots, X_k$ affect a variable of interest y. Some examples:
  1. Demand for a good:
     - $Y_i =$ demand for gasoline for individual i
     - $X_{1,i} =$ price of gasoline, $\#$ of cars owned by i
     - Question: how much would demand vary if prices were to increase by 1%?
  2. Test Scores:
     - $Y_i =$ test score for student i
     - $X_i =$ class size, teacher's experience, ...
     - Question: how much would i's test score vary if the class size was reduced by half?

General Regression Models

- Regression model: describes the relationship between y and x

$$Y_i = g(X_i, \beta, u_i), \quad i = 1, \ldots, n$$

- $i =$ individuals, $n =$ number of observations
- $Y_i =$ dependent variable
- $X_i =$ vector of independent variables or regressors
- $\beta =$ vector of parameters

$$X_i = \begin{pmatrix} X_{1,i} \\ \vdots \\ X_{k,i} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}$$

- $u_i =$ regression error

The Linear Regression Model with a Single Regressor

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \quad i = 1, \ldots, n$$

- We have n observations, $(X_i, Y_i), i = 1, \ldots, n$.
- X is the independent variable or regressor
- Y is the dependent variable
- β_0 = intercept
- β_1 = slope
- u_i = the regression error; it consists of omitted factors. In general, these omitted factors are other factors that influence Y, other than the variable X. The regression error also includes error in the measurement of Y.

Linear regression lets us estimate the slope of the population regression line

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \quad i = 1, \ldots, n$$

- The slope β_1 of the population regression line is the expected effect on Y of a unit change in X.
- Ultimately our aim is to estimate the causal effect on Y of a unit change in X – but for now, just think of the problem of fitting a straight line to data on two variables, Y and X.

Statistical, or econometric, inference about the slope entails

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \quad i = 1, \ldots, n$$

- Estimation:
  - How should we draw a line through the data to estimate the population slope β_1?
    Answer: ordinary least squares (OLS).
  - What are the advantages and disadvantages of OLS?
- Hypothesis testing:
  - How to test if the slope β_1 is zero?
- Confidence intervals:
  - How to construct a confidence interval for the slope β_1?

Example: Test Scores

$$\text{Test Score}_i = \beta_0 + \beta_1 \text{STR}_i + u_i, \quad i = 1, \ldots, n$$

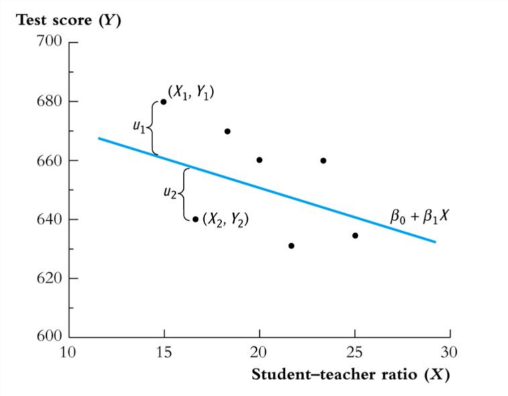
- STR = Student to Teacher Ratio
- β_1 = slope of population regression line:

$$\beta_1 = \frac{\Delta \text{Test Score}}{\Delta \text{STR}}$$

- Why are β_0 and β_1 "population" parameters?
- We would like to know the population value of β_1.
- We don't know β_1, so must estimate it using data.

The population regression model in a picture

Observations on Y and X ($n = 7$); the population regression line; and the regression error (the "error term"):



The Ordinary Least Squares Estimator (SW 4.2)

- We will focus on the least squares ("ordinary least squares" or "OLS") estimator of the unknown parameters β_0 and β_1.
- The OLS estimator solves:

$$\min_{b_0, b_1} \sum_{i=1}^{n} (Y_i - [b_0 + b_1 X_i])^2$$