

Econ 103: Introduction to Econometrics

Lecture 1 — What Is Econometrics?

Ryan Longmuir

UCLA

Summer Session C, 2026

Reading: Stock & Watson (4th ed.), Ch. 1, §1.1–1.3; Hill, Griffiths & Lim (5th ed.), Ch. 1, §1.1–1.5.

Welcome to Econ 103

Goal of the course. Learn the core tool of empirical economics — [linear regression](#) — and how to use it with real data in R.

Over six weeks we build one idea, in three stages:

- **Probability review** (L2–L4): the language of uncertainty.
- **Simple regression** (L5–L12): one explanatory variable.
- **Multiple regression** (L13–L20): many explanatory variables, plus causality and treatment effects.

Today (Lecture 1)

The big picture: what econometrics *is*, what it is *for*, and the single distinction that runs through the whole course — [correlation vs. causation](#).

Today's plan

- 1 Course mechanics
- 2 What is econometrics?
- 3 Three goals: explain, predict, optimize
- 4 Correlation vs. causation
- 5 A first look at data

Part 1

Course mechanics

The essentials

Meetings (on Zoom)

- **Lecture:** Tue & Thu, 3:15–5:20 pm
- **Lab:** Tue, 1:00–2:15 pm

Recordings posted when technically possible.
You must enroll in **both** the lecture and its paired lab.

Staying in touch

- Office hours: Wed & Fri, 11:30–1:00, Bunche 2265
- Email: rlongmuir@g.ucla.edu (put “ECON 103” in the subject)
- Questions → **Campuswire** (code 6251)

Textbook

Principles of Econometrics, Hill, Griffiths & Lim (5th ed.). Main text; we cover roughly the first 8 chapters. Slides also posted. Stock & Watson is a useful second reference and is mapped chapter-by-chapter in the lecture notes.

Software

R and **RStudio** (free). Install before Lab 1; campus / virtual desktops have it too. *No prior coding required.*

Example (Prerequisites)

Econ 11 and Econ 41. We lean heavily on [Econ](#)

How you're graded

Your grade is a weighted average of homework, the midterm, and the final — and we take **whichever option is better for you**:

- ~ 4 homework assignments (online; due Mondays at midnight). Late = half credit absent an accommodation or prior arrangement.
- Some **coding questions** on every assignment (R recommended).
- Exams are multiple choice, online, closed-book / closed-internet.

Component	Opt. 1	Opt. 2
Homework	20%	20%
Midterm	30%	40%
Final	50%	40%

Exam dates

- **Midterm:** Thu, Aug 20 (covers L1–L10)
- **Final:** Thu, Sep 10 (cumulative)

Part 2

What is econometrics?

What is econometrics?

Ask a half-dozen econometricians and you get a half-dozen answers: the science of testing theories; a toolkit for forecasting; fitting models to data; making numerical policy recommendations.

They're all right.

Working definition

Econometrics is the science and art of using **economic theory** and **statistical methods** to analyze **economic data** — to estimate relationships, test hypotheses, and predict outcomes.

It bridges the gap between being “a student of economics” and being a *practicing* economist. Theory says “a price increase lowers quantity demanded.” Econometrics answers **by how much**.

Economics runs on “how much?” questions

Theory gives the *direction* of a relationship; decisions need a *number*.

- Cut class size by 5 students — how much do test scores rise?
- Raise cigarette prices 1% — how much does consumption fall?
- A Pizza Hut buys more newspaper ads — how much do sales rise?
- The Fed raises the discount rate — how much does inflation slow?

The recurring structure

Each question is about a **parameter** — an unknown number (an elasticity, a slope, a multiplier) describing how one variable relates to another. The parameter's value is **unknown** and must be **estimated from data**.

Because the answer comes from data, it carries **uncertainty**: a different sample gives a different number. So we always report an estimate *and* how precise it is.

From an economic model to an econometric model

Economic theory writes relationships as exact functions, e.g. a demand equation

$$Q^d = \beta_1 + \beta_2 P + \beta_3 P^s + \beta_4 P^c + \beta_5 \text{INC}.$$

But real outcomes are *not* exact — countless small factors are omitted. Econometrics adds a **random error** e :

$$Q^d = \underbrace{\beta_1 + \beta_2 P + \beta_3 P^s + \beta_4 P^c + \beta_5 \text{INC}}_{\text{systematic part (from theory)}} + \underbrace{e}_{\text{random "noise"}}.$$

- The β 's are the **unknown parameters** we want to learn.
- e collects everything we left out and the intrinsic randomness of behavior.

This “systematic part + error” template is the skeleton of *every* model in the course. We meet it formally in Lecture 5.

Part 3

Three goals: explain, predict, optimize

Why are we building the model? Three goals

The same regression can serve different purposes — and the purpose changes what we must assume.

1. Explain (causal)

Estimate the **effect** of X on Y , *holding all else equal*. “How much do test scores rise *because of* smaller classes?”

2. Predict (forecast)

Use X to make a good **guess** of Y . Causality *not* required. “What will GDP growth be next year?”

3. Optimize (decide)

Plug estimates into a **decision**. “*How much* should the Fed raise rates? How much should we advertise?”

Example (The goal sets the bar)

*For prediction, you only need a stable association. For explanation and sound decisions, you need the parameter to be **causal** — a much higher bar, and most of the work of this course.*

Part 4

Correlation vs. causation

The central distinction of the course

Consider a regression of a student's *grade* on the fraction of lectures they *skip*:

$$\text{GRADE} = \beta_1 + \beta_2 \text{SKIP} + e.$$

We expect $\beta_2 < 0$: more skipping goes with lower grades.

But does skipping *cause* low grades?

Maybe not. A demanding job, or low motivation, could drive **both** skipping and poor grades. Then SKIP and GRADE are **correlated** — useful for *prediction* — yet β_2 is **not** the causal effect of skipping. It also soaks up the influence of those omitted factors.

Correlation is not causation. A good predictor (umbrellas predict rain) need not be a cause (umbrellas don't *make* it rain).

What would pin down a causal effect?

The ideal: a randomized experiment

Split units into a **treatment group** and a **control group** by **random assignment**, then compare average outcomes.

Randomization makes the groups comparable in every other way, so the only systematic difference is the treatment. The **causal effect** is *defined* as the difference such an experiment would reveal.

- *Agronomy*: fertilize randomly chosen plots; compare yields.
- *Education*: Tennessee's **Project STAR** randomly assigned students to small vs. regular classes (we revisit it in L20).

Why we usually can't

Real experiments are often **unethical**, **infeasible**, or **too expensive**. So most economic data are **observational** — we watch the world, we don't run it.

Example (The confounding trap)

*Districts with small classes also tend to be wealthier. So a raw "small classes → higher scores" comparison mixes the class-size effect with the income effect. Untangling them is the job of **multiple regression** (Part II).*

Part 5

A first look at data

Data come in three types

Cross-section

Many **entities**, one time period. 420 California districts in 1999; a CPS wage survey. # entities = n .

Time series

One **entity**, many periods. U.S. GDP growth, quarterly, 1960–2017. # periods = T .

Panel

Many entities, *each* over many periods. 48 states \times 11 years of cigarette sales. $n \times T$ observations.

Data are also **experimental** (from a designed experiment) or **observational** (surveys, administrative records) — and most economic data are observational. We start, and spend most of the course, with **cross-sectional** data.

A first look: food expenditure and income

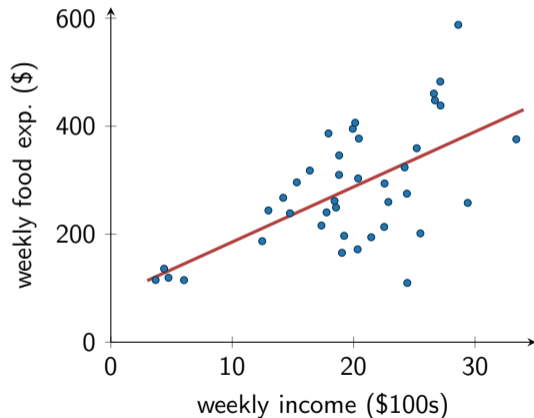
Our running example (Hill, Griffiths & Lim): a cross-section of 40 households — weekly *income* and weekly *food expenditure* (both in \$100s and \$, resp.).

A **scatterplot** is the natural first look at the relationship between two variables.

- Higher income **tends to** go with higher food spending — an upward cloud (correlation ≈ 0.62).
- But the points don't lie on a line: spending varies a lot at any given income. That scatter is the **random error** e .

The course in one picture

Lectures 5–6 make “draw the best line through this cloud” precise — that line is **ordinary least squares**.



Dots: 40 households; red line previews OLS (L5–6).

Recap

What econometrics is

Using economic theory + statistics to turn *data* into *numbers*: estimating parameters, testing hypotheses, predicting outcomes.

Three goals

- Explain (causal effects)
- Predict (forecast)
- Optimize (decisions / policy)

Data types

- Cross-section, time series, panel
- Experimental vs. observational

Next time (Lecture 2): the language of uncertainty — [random variables and distributions](#).

The big idea

- Model = systematic part + random error e .
- **Correlation \neq causation.**
- Randomized experiments *define* causal effects; with observational data we work to approximate them.

Questions?