# Chapter 1

Introduction

# Learning Objectives

- Define and describe the basics of econometrics
- Describe how to do an econometric study

# Step 1: What Do We Want to Do?

- The first step in doing econometrics is to define the purpose of the modeling. It is easy to skip this step, but doing so means your analysis is unlikely to be useful.

- Your purpose should be concrete and concise.

- Often, if you can state your purpose in the form of a question, you will see whether you have defined it adequately.

# Example: Do Good Teachers Produce Better Student Outcomes?

*American Economic Review 2014, 104(9): 2633–2679*
*http://dx.doi.org/10.1257/aer.104.9.2633*

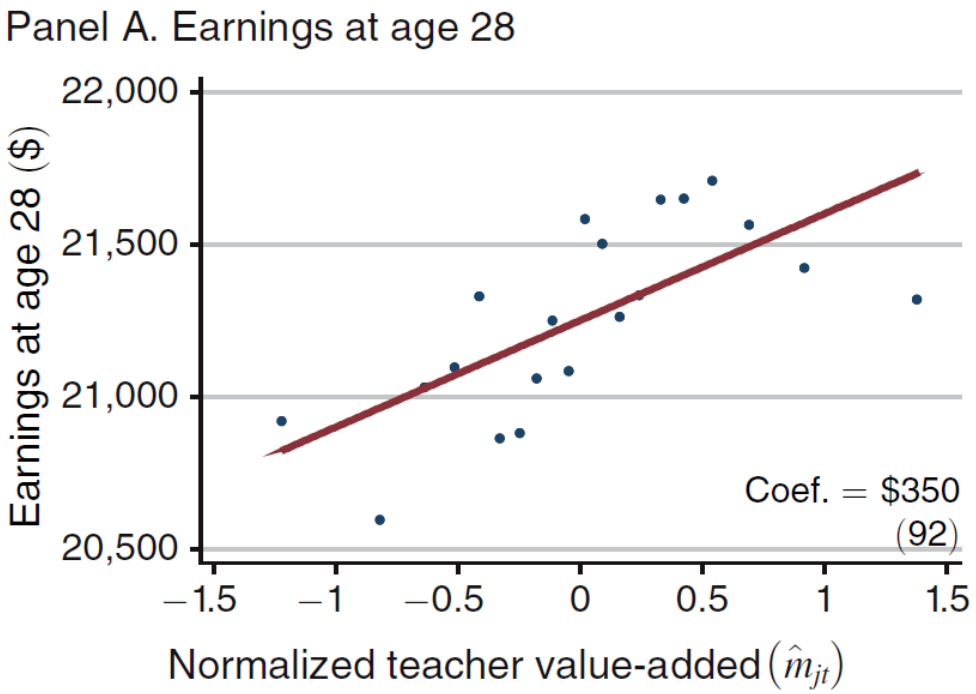
## Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood†

*By* Raj Chetty, John N. Friedman, and Jonah E. Rockoff*

*Are teachers' impacts on students' test scores (value-added) a good measure of their quality? This question has sparked debate partly because of a lack of evidence on whether high value-added (VA) teachers improve students' long-term outcomes. Using school district and tax records for more than one million children, we find that students assigned to high-VA teachers are more likely to attend college, earn higher salaries, and are less likely to have children as teenagers. Replacing a teacher whose VA is in the bottom 5 percent with an average teacher would increase the present value of students' lifetime income by approximately $250,000 per classroom.* (*JEL* H75, I21, J24, J45)

https://www.aeaweb.org/articles?id=10.1257/aer.104.9.2593

# YES! Good Teachers Raise Future Income




Replacing an average teacher with a teacher in top 5% would increase students' earnings later in life by 2.8%.

- The average 12 year old in the United States can expect lifetime earnings of $522,000.
  - 2.8% earnings bump is worth about $14,500 per student.
  - Multiply that by 20 kids per classroom and an excellent teacher is really valuable.

# Example: Does the Law of Demand Hold for Electricity?

*American Economic Review 2014, 104(4): 1417—1438*
*http://dx.doi.org/10.1257/aer.104.4.1417*

## Knowledge is (Less) Power: Experimental Evidence from Residential Energy Use†

*By* KATRINA JESSOE AND DAVID RAPSON*

*Imperfect information about product attributes inhibits efficiency in many choice settings, but can be overcome by providing simple, low-cost information. We use a randomized control trial to test the effect of high-frequency information about residential electricity usage on the price elasticity of demand. Informed households are three standard deviations more responsive to temporary price increases, an effect that is not attributable to price salience. Conservation extends beyond pricing events in the short and medium run, providing evidence of habit formation and implying that the intervention leads to greenhouse gas abatement. Survey evidence suggests that information facilitates learning.* (*JEL* D12, D83, L11, L94, Q41, Q54)

https://www.aeaweb.org/articles?id=10.1257/aer.104.4.1417

# YES! Law of Demand Holds for Electricity – but only if people know how much they're using

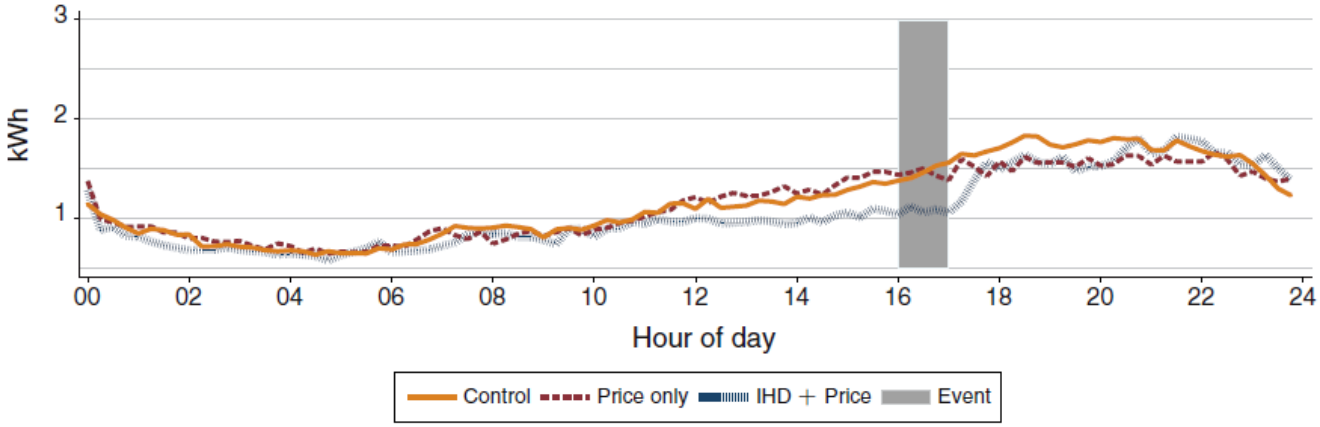
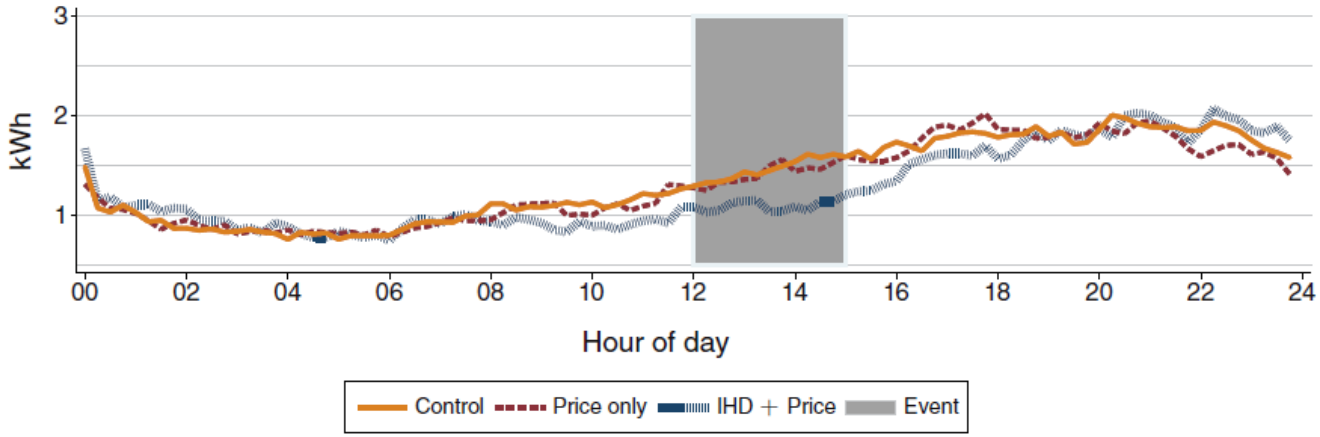



FIGURE 5. AUGUST 17, 2011: 2HR $1.25 INCREASE, 30-MIN NOTICE



# Example: Is it Possible to Forecast Stock Returns?


## A Comprehensive Look at The Empirical Performance of Equity Premium Prediction

**Ivo Welch**
Brown University Department of Economics NBER

**Amit Goyal**
Emory University Goizueta Business School

Our article comprehensively reexamines the performance of variables that have been suggested by the academic literature to be good predictors of the equity premium. We find that by and large, these models have predicted poorly both in-sample (IS) and out-of-sample (OOS) for 30 years now; these models seem unstable, as diagnosed by their out-of-sample predictions and other statistics; and these models would not have helped an investor with access only to available information to profitably time the market. (*JEL* G12, G14)

https://academic.oup.com/rfs/article/21/4/1455/1565737

# No! Stock Market Prediction Models are Unreliable?

| | Variable | Data | IS $\overline{R}^2$ | IS for OOS $\overline{R}^2$ | OOS $\overline{R}^2$ | OOS ΔRMSE |
|---|---|---|---|---|---|---|
| **Full Sample, Not Significant IS** | | | | | | |
| **dfy** | Default yield spread | 1919–2005 | −1.18 | | −3.29 | −0.14 |
| **infl** | Inflation | 1919–2005 | −1.00 | | −4.07 | −0.20 |
| **svar** | Stock variance | 1885–2005 | −0.76 | | −27.14 | −2.33 |
| **d/e** | Dividend payout ratio | 1872–2005 | −0.75 | | −4.33 | −0.31 |
| **lty** | Long term yield | 1919–2005 | −0.63 | | −7.72 | −0.47 |
| **tms** | Term spread | 1920–2005 | 0.16 | | −2.42 | −0.07 |
| **tbl** | Treasury-bill rate | 1920–2005 | 0.34 | | −3.37 | −0.14 |
| **dfr** | Default return spread | 1926–2005 | 0.40 | | −2.16 | −0.03 |
| **d/p** | Dividend price ratio | 1872–2005 | 0.49 | | −2.06 | −0.11 |
| **d/y** | Dividend yield | 1872–2005 | 0.91 | | −1.93 | −0.10 |
| **ltr** | Long term return | 1926–2005 | 0.99 | | −11.79 | −0.76 |
| **e/p** | Earning price ratio | 1872–2005 | 1.08 | | −1.78 | −0.08 |

- Negative means that prediction model is worse than historical mean.
- IS means “in-sample”. OOS means “out-of-sample”.

# Step 2: Formulate Your Research Design and Specify the Econometric Model

- Take your abstract objective from step 1 and convert it into an econometric model with data that can answer your questions.

- Requires some economic theory, common sense and a little cleverness.

- Making a good choice about which data to collect and use determines whether you will be able to meet your objective.

# Step 2 Example: Do Good Teachers Produce Better Student Outcomes?

- Who is a good teacher?
  - Can't just measure test scores. Some teachers may get high test scores because they are easy graders or because smart students take their class.
  - Answer:  a good teacher raises test scores more than average

- With this "value added" measure of teacher quality, Chetty, Friedman and Rockoff can estimate using econometrics whether students who took classes from high VA teachers earn more later in life.

# Step 2 Example: Does the Law of Demand Hold for Electricity?

- Everyone wants to use their AC on hot days
- Electricity is more expensive on hot days because demand is high on those days
- To test law of demand, you could give some people a high price and others a low price on the same day
- Jessoe and Rapson convinced an electric utility to let them raise prices for a random subset of customers
  - Then they estimated using econometrics how consumers responded to the high prices

# Step 2 Example: Is it Possible to Forecast Stock Returns?

- Everyone is a stock market expert after the fact
  - (or a bitcoin market expert)

- Example: historically, the stock market did better in January.
  - Does this mean that the stock market will do better in future Januarys?
  - Hint: no.

- Welch and Goyal build econometric models of stock returns using historical data and then see whether those models predict future returns.

# Step 3: Apply Statistical Theory

The Power of Statistics:

Statistics lets us make statements of probability about an **unobserved population** using data from a single **sample**.

***Econometrics is challenging because it integrates all three fields—economics, mathematics, and statistics—into one.***

# How Does Poverty Affect Student Performance in California Schools?

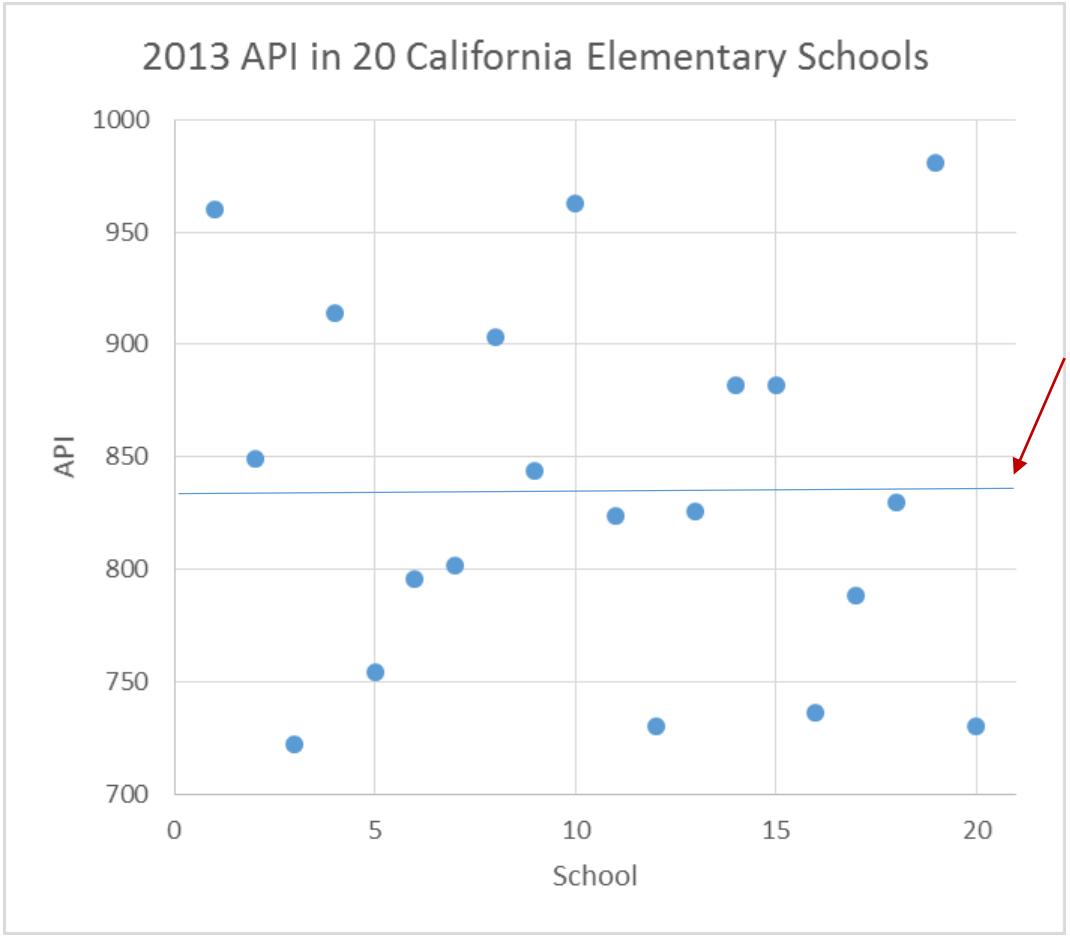
- To answer this question, we first need to measure “student performance” and “poverty.”
- Academic performance index (API)
  - The California state government uses it to measure school performance
  - Constructed from students’ scores on state-wide standardized tests
  - Ranges from 200 to 1000.
  - We can use the API as our measure of average student performance at each California school
- We do not know the household incomes for each student at each school
  - But U.S. National School Lunch Program provides free lunches to school children from low-income households
  - We know the share of students eligible for free lunches at each school
  - We can use free lunch eligibility (FLE) as an indicator of the share of students from households with incomes below the poverty line at each school.

# Table 1.1 Academic Performance Index (API) and Free Lunch Eligibility (FLE) at 20 Randomly Chosen California Elementary Schools in 2013 (from Total Population of 5,765 Schools)

| School | County | API | FLE |
|---|---|---|---|
| Joe A. Gonsalves Elementary | Los Angeles | 960 | 16 |
| Old River Elementary | Kern | 849 | 0 |
| Sierra Vista Elementary | Kern | 722 | 96 |
| West Portal Elementary | San Francisco | 914 | 44 |
| Isabelle Jackson Elementary | Sacramento | 754 | 83 |
| Rio Vista Elementary | Orange | 796 | 90 |
| Poplar Avenue Elementary | Butte | 802 | 80 |
| Cloverly Elementary | Los Angeles | 903 | 46 |
| Creative Arts Charter | San Francisco | 844 | 33 |
| Carolyn A. Clark Elementary | Santa Clara | 963 | 6 |
| Raymond Elementary | Orange | 824 | 69 |
| Fernangeles Elementary | Los Angeles | 730 | 100 |
| Rainbow Ridge Elementary | Riverside | 826 | 90 |
| Cyrus J. Morris Elementary | Los Angeles | 882 | 29 |
| Benjamin Franklin Elementary | Riverside | 882 | 36 |
| Salvador Elementary | Napa | 736 | 65 |
| Bowers Elementary | Santa Clara | 788 | 59 |
| Vintage Parkway Elementary | Contra Costa | 830 | 54 |
| Balboa Magnet Elementary | Los Angeles | 981 | 22 |
| Selby Lane Elementary | San Mateo | 730 | 80 |
| | Mean | 835.80 | 54.90 |
| | Std. Deviation | 81.14 | 31.00 |

Source: California Department of Education; http://www.cde.ca.gov/ta/ac/ap/apidatafiles.asp

**Figure 1.1** 2013 API in 20 California elementary schools. Academic performance varies widely across these 20 schools.



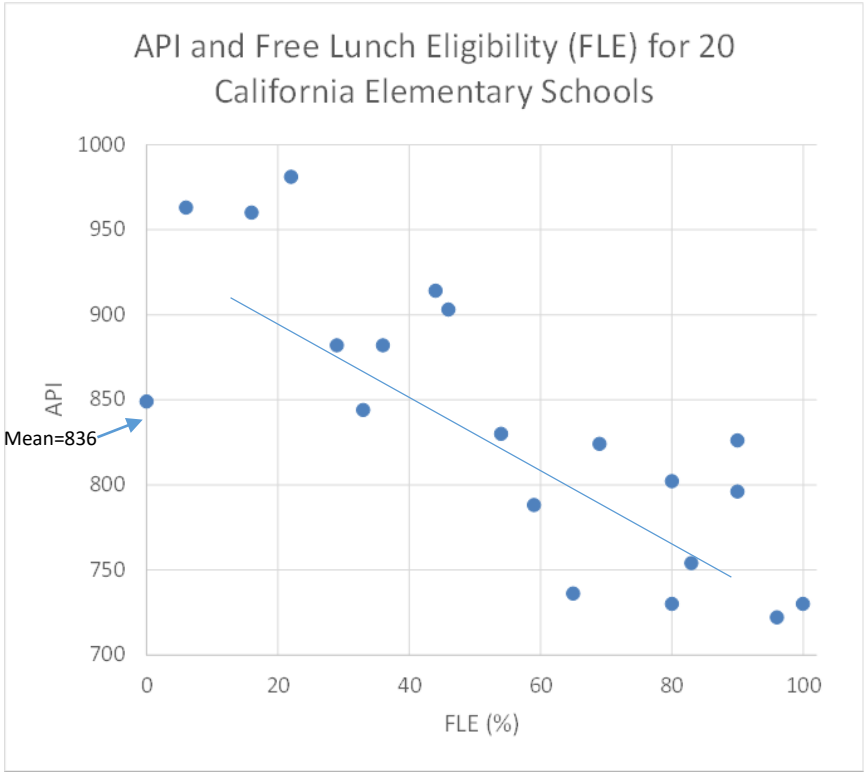

Mean API for this sample = 835.80.

Repeat for many random samples, and on average you would get $\mu$, the mean API of all schools. *But $\mu$ almost certainly does not equal 835.80!*

(BYAM: It's 813.70. But we don't know that yet!)

# Let Free School Lunch Eligibility "Proxy" for Poverty in School Districts

Figure 1.5. It sure looks like API decreases with FLE.




What we can learn from FLE to help predict API? Begin by estimating a linear model in which FLE is the only variable explaining API; that is:

$$Y_i = b_0 + b_1 X_i + e_i$$

$Y_i$: API of school $i$

$X_i$: FLE of school $i$

$e_i$: Error, how much we miss by if we use FLE to predict API for school $i$

# What We Learned

- Econometrics is about making sense of economic data.

- Three steps to conducting econometric analysis.
  1. State the purpose of the analysis.
  2. Formulate the research design and specify the econometric model.
  3. Apply statistical theory.