

# Data Section of a Structural Estimation Paper

**Dr. Richard W. Evans**

February 2020

# Insight from Physics

## Richard P. Feynman, *QED: The Strange Theory of Light and Matter*

*“Finally, there is this possibility: after I tell you something, you just can’t believe it. You can’t accept it. You don’t like it. A little screen comes down and you don’t listen anymore. **I’m going to describe to you how Nature is**—and if you don’t like it, that’s going to get in the way of your understanding it. It’s a problem that physicists have learned to deal with: They’ve learned to realize that whether they like a theory or they don’t like a theory is not the essential question. Rather, it is whether or not the theory gives predictions that agree with experiment. It is not a question of whether a theory is philosophically delightful, or easy to understand, or perfectly reasonable from the point of view of common sense. The theory of quantum electrodynamics describes Nature as absurd from the point of view of common sense. And it agrees fully with experiment. So I hope you can accept **Nature as She is**—absurd.”*

# The Data Section is...

- Data section is about “Data as She is”
- Sometimes (hopefully) you find surprises
- Demonstrate, display, describe key slices of your data
- Set up data for richer theoretical or empirical analysis
- “Every good paper should have a great data visualization”

# Where ideas come from

- Media article/report claim. “Is that true?”
  - Read the day’s news headlines, tweets, and posts
- New types of events/shocks
  - Great recession, Coronavirus, Syrian refugees, natural experiments
- New data sources allow for new questions and/or answers
- New methods allow for new questions and/or answers

# Inequality examples

- Piketty (2014), income inequality is increasing in developed countries.
  - High incomes rising at a fast rate
  - Middle-to-low incomes stagnant
- 2 recent papers deep dive in data to find different results/interpretations
  - Sacerdote: middle-to-low standard of living has been increasing
  - Guvenen and Kaplan: high incomes not increasing that much

# Inequality: Sacerdote

## Sacerdote (2017) “Fifty Years of American Growth...”

*“Despite the large increase in U.S. income inequality, consumption for families at the 25th and 50th percentiles of income has grown steadily over the time period 1960-2015. The number of cars per household with below median income has doubled since 1980 and the number of bedrooms per household has grown 10 percent despite decreases in household size. The finding of zero growth in American real wages since the 1970s is driven in part by the choice of the CPI-U as the price deflator; small biases in any price deflator compound over long periods of time. Using a different deflator such as the Personal Consumption Expenditures index (PCE) yields modest growth in real wages and in median household incomes throughout the time period. ... Meaningful growth in consumption for below median income families has occurred even in a prolonged period of increasing income inequality, increasing consumption inequality and a decreasing share of national income accruing to labor.”*

# Inequality: Guvenen and Kaplan

## Guvenen and Kaplan (2017) “Top Income Inequality in the 21st Century: Some Cautionary Notes”

- Incomes at the top increasing only because of shifting income from C-corps to S-corps
- “Such income is not ‘new’ income earned by top earners but is simply income that was previously labeled as corporate income rather than household income.”

## News, tweets, posts

- Look through days news and tweets



# Parts of a Data Section

- ① Describe how to access data, where stored, who curates
- ② Cite other key papers that have used this data
- ③ Describe how the data were collected
- ④ Describe key variables
  - Mean, std. dev., max., min., count
  - Correlations
  - Conditional statistics (slices), cross tabs
  - Time series
  - Visualizations, composition charts

## Parts of a Data Section

- ⑤ Map the variables from the data to the concepts of your model
  - Model/theory concepts often require transformations of data variables
- ⑥ Set up the data for the richer analyses you are going to do
- ⑦ You can often demonstrate the key result of your analysis with one visualization or table
  - Sometimes, this has to appear later in the paper because it requires your methods to see
  - But sometimes you can show it with a basic descriptive table/visualization utilizing a key slice

## Parts of a Data Section

- 8 You may want to have a data appendix
  - More detail, definitions, slices, visualizations can go here
- 9 Make sure your code for accessing, cleaning, organizing the data is available
  - GitHub, website, etc.
- 10 Which should come first, data section or model section?

## Examples: From Literature and Students

- Altonji, Joseph G., Anthony A. Smith, Jr., and Ivan Vidangos, “Modeling Earnings Dynamics,” *Econometrica*, pp. 1395-1454 (July 2013)
- Rust, John, “Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher,” *Econometrica*, 55:5, pp. 999-1033 (Sep. 1987)
- Bobae and DP paper
- Student taxi paper

Let's use one of your papers

Go through a student's project and outline data section.