

# Empirical Research in Modern Economics

Nurmukhammad Yusupov

TSUE

March 31, 2022

# Outline

- To Do or Not To Do [Research]? The Why?
- Types of Data
- Types of Estimations
- Getting Creative w/ Data Analysis
- Q&A

# To Do or Not To Do [Research]?

## The Why?

- The ultimate goal of research is to answer [unanswered] questions, hopefully to inform important decisions.
- For many academics, research is part of everyday job.
  - Must lead to a publication. Does not have to have practical implications.
  - The name of the game is “publish or perish”.
- For students, it is part of hands-on experience.
  - May earn you a good recommendation letter.
  - May help you land a job / internship.
  - Another line in your CV. ← as a student, you are a in a CV Building business
- For practitioners, (firms, governments, households, etc.), it may produce tangible / applicable results.

# To Do or Not To Do [Research]?

## The Why?

- Theoretical research offers new theory, and/or refutes existing [bad] theory.
  - Theory is simplified representation of reality.
  - It is based on a set of assumptions. You want weak assumptions, not strong assumptions.
  - Theory can be written verbally, through mathematical equations or schematic depiction. Must follow the norms of logic.
- Empirical research is based on data analysis, typically to find statistical support in favor of a hypothesis/idea or against it.
- Conceptual research is mostly, if not fully, verbal and develops concepts. ← not recommended for young researchers.
- Literature review takes stock of the extant results on a given issue or branch of the field to present ideas in a systematized way.

# Types of Data

- Data analysis is a set of procedures for discovering structure and patterns in data, typically aimed at
  1. Generating predictive models
  2. Classification / clustering of things
  3. Description of the underlying phenomena
- Those procedures are mostly:
  - Founded on statistical tools
  - Getting more computer-intensive, e.g. Big Data

# Types of Data

- Cross Sectional data is a collection of variables for a set of units at a single point in time, e.g. individuals, firms, households, etc.
  - Most frequently used
- Longitudinal (or Panel) data is a collection of the same variables for the same set of units at multiple points in time.
  - Most useful, but hard to implement.
- Repeated Cross Section is a collection of the same variables for comparable, but not the same set of units, e.g. new random samples each time.
- Time-series data is observations about the same units over time.

# Types of Data

- Cross Sectional data is a collection of variables for a set of units at a single point in time, e.g. individuals, firms, households, etc.
  - Most frequently used
- Longitudinal (or Panel) data is a collection of the same variables for the same set of units at multiple points in time.
  - Most useful, but hard to implement.
- Repeated Cross Section is a collection of the same variables for comparable, but not the same set of units, e.g. new random samples each time.
- Time-series data is observations about the same units over time.

Time series + Cross section = Panel

# Types of Data

- Household surveys:
  - LSMS ← Some are panels
  - DHS ← Some could be panels
  - Country level surveys (e.g. IFLS, RLMS, KLIPS)
  - Regional household surveys (e.g. LiT, CALISS)
  - ...
- Social surveys:
  - \*GSS, ESS
  - ...
- Enterprise surveys:
  - BEEPS, WBES
  - GEM
  - Workplace Surveys
  - ...



# Types of Estimations

- Basic statistics:
  - Measures of central tendency (mean, mode, median, ...)
  - Measures of dispersion (st. deviation, variance, partial moments, ...)
  - Visualization of data (histograms, pie charts, distribution graphs, ...)
- Regressions are empirical estimations of the specific types of relationships between the dependent variable and one or multiple independent variables.
  - Most often, and perhaps the simplest, used regression method is that of ordinary least squares (OLS), which assumes a linear relationship.
  - Statistical software is used, e.g. Stata, SPSS, etc. (even MS Excel)

# Types of Estimations

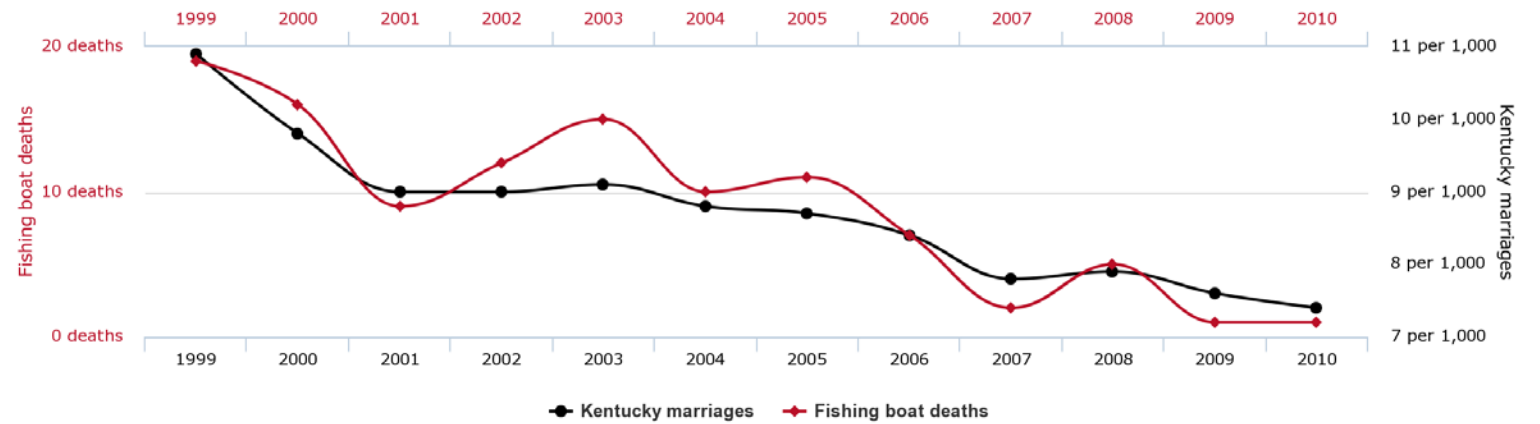
- Correlation vs causality is the biggest headache of all researchers!
- Problems of reverse causality:
  - Height and gender
  - Institutions and development
- Problems of endogeneity, e.g. multicollinearity.
- While there are statistical tools, the bottom-line is human judgement. **Causality has to make sense!**
-

# Types of Estimations

- Correlation vs causality is the biggest headache of all researchers!
- Problems of reverse causality:
  - Height and gender
  - Institutions and development
- Problems of endogeneity, e.g. multicollinearity.
- While there are statistical tools, the bottom-line is human judgement. **Causality has to make sense!**
- Spurious correlations:

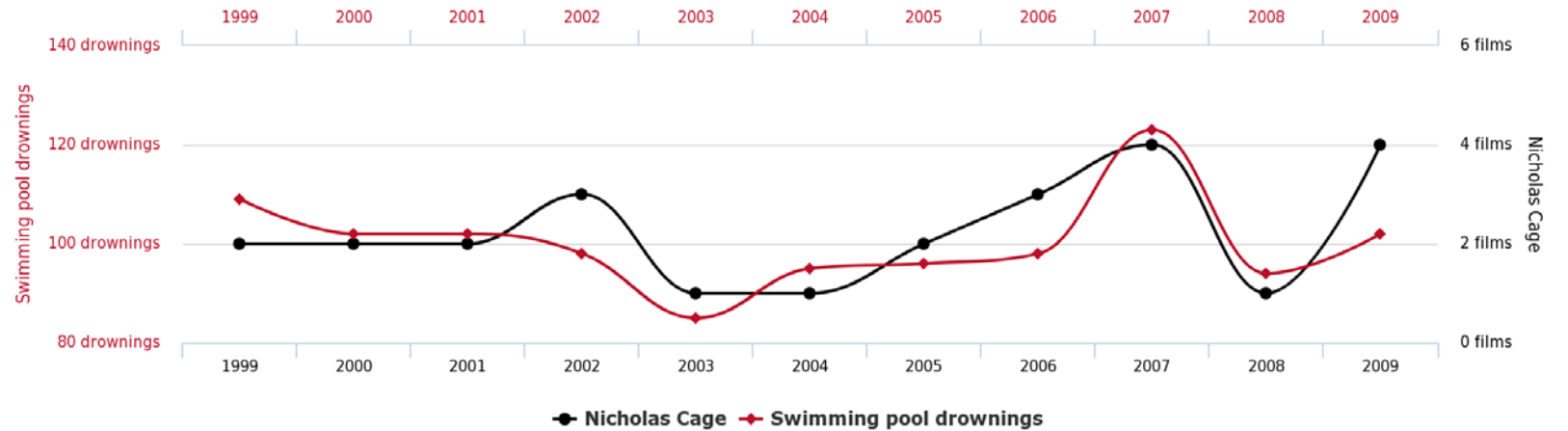
# Spurious correlations

## People who drowned after falling out of a fishing boat correlates with Marriage rate in Kentucky



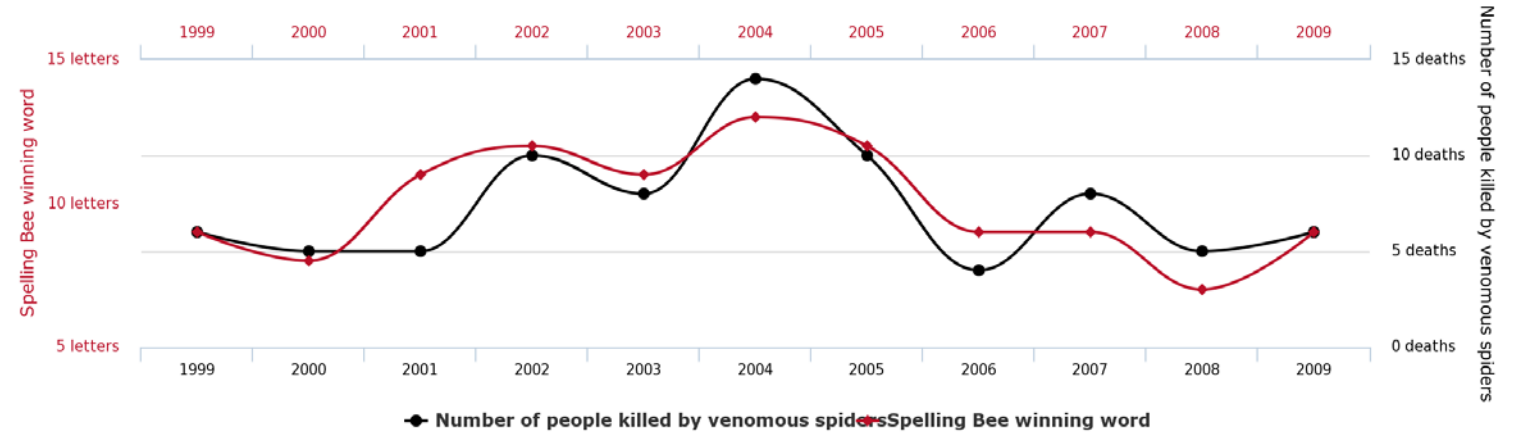
# Spurious correlations

## Number of people who drowned by falling into a pool correlates with Films Nicolas Cage appeared in



# Spurious correlations

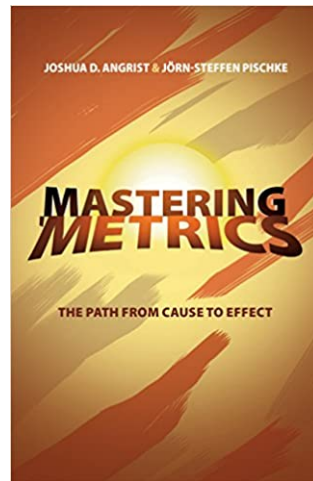
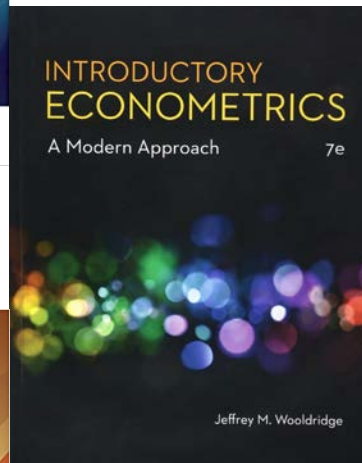
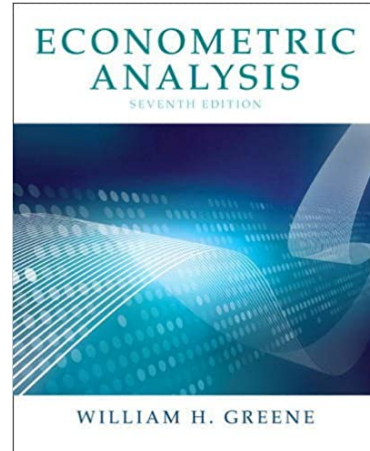
## Letters in Winning Word of Scripps National Spelling Bee correlates with Number of people killed by venomous spiders



# Getting creative

- People do not like to report their income. Therefore, we use consumption data as a proxy for income.
- Causality:
  - Design your research in a way that enables you to answer the counterfactual question, e.g. treatment and control groups.
  - Use panel data. Repeated cross-section can be turned into a panel with some aggregation.
  - Laboratory, or controlled trials.
  - Always try to achieve randomization to address possible selection bias.
  - Try to find an instrumental variable.

# Textbooks



- Greene "Econometric Analysis" ← probably the most complete and advanced level textbook
- Wooldridge "Introductory Econometrics" ← good introductory text
- Angrist and Pischke "Mastering 'Metrics'" ← focus on causality



# Getting creative

- Often, establishing causality requires experimenting with interventions, i.e. control-treatment approach.
  - Laboratory experiments
  - Randomized control trials → selection bias
  - Natural experiments
- However, modern tools allow for causality analysis of observational data\* as well, e.g. based on longitudinal datasets. But one may have to get creative.

\* 2021 Nobel Prize in Economics was awarded exactly for that.

# Getting creative

- Acemoglu and Robinson (2001). “Colonial origins of comparative development,” *American Economic Review*, 91 (5): 1369–1401.
  - colonists mortality as an instrument for institutions.
- Jayachandran and Kuziemko (2011). “Why Do Mothers Breastfeed Girls Less than Boys? Evidence and Implications for Child Health in India,” *The Quarterly Journal of Economics*, 126(3): 1485-1538.
  - breastfeeding duration as a proxy for son preference.
- Vogl (2013). “Marriage Institutions and Sibling Competition: Evidence from South Asia,” *The Quarterly Journal of Economics*, 128(3): 1017–1072.
  - birth spacing of siblings as a factor in arranged marriage.

# Getting creative

- Are entrepreneurs are made or born? - Twin studies
- Measuring efficiency of institutions across countries – mail experiment.
- Measuring economic development across countries – satellite image study.
- Differential treatment of boys vs girls – outlay equivalence analysis.
- Measuring corruption – parking tickets in downtown NYC.

# Final thoughts

- What to do to write a good empirical paper?
  - Have a puzzle. No puzzle – no paper.
  - Be innovative. Try combining multiple datasets.
  - Develop a solid theoretical foundation. Remember, computer does not care what to compute. Results must make sense and lend themselves to interpretation.
  - Learn the idiosyncrasies of the language in your subfield. Speak the same language as your audience does.
  - Your references are your spare parts
    - → cite reputable sources, avoid predatory journals and events
    - → know the panorama of journals, events and institutions.
    - See <http://doctorant.tsue.uz/uslubiy-korsatma/>
  - Be patient:

*Long is the way and hard,  
that out of hell leads up to light.*

# Q&A

- Comments?
- Questions?