Causality in Economic History

James Fenske

University of Oxford

In this lecture

- I summarize Bin Wong's views on how causation has been and should be used in history.
- I summarize a textbook approach to the Neyman-Rubin causal model, which I believe is the most widely used framework for discussing causation in the social sciences.
- I discuss Angrist and Pischke's views on the "credibility revolution" in empirical economics and the critiques that have been made of them.
- I briefly discuss some of methods that I will teach in Quantitative Methods 2 and that are popular in the estimation causal effects.

Introduction

- R. Bin Wong, "Causation", *A Concise Companion to History*.
- "Causation of one form or another is present in most narratives." Most histories tell stories.
- But: historians focus less on causation today than in the 1960s and 1970s, with exceptions.
- Until the 1960s, historians wrote about major events by presenting reasons why they occurred (e.g. causes of the industrial revolution).
- These histories often left out the lives of common people, leading to the social histories of the 1960s and 1970s.
- The turn to cultural history since the 1980s has led to a greater split from economics and political science and greater use of anthropology and literary studies - causation no longer central.

Introduction

- "Causation, isn't that something historians worried about in the 1970s?"
- Global history since 2000 has replaced earlier attempts to track historical changes over long periods.
- "How do we navigate between individual stories and larger accounts of history that avoid the pitfalls of grand narratives with teleological assumptions?"
- Answering this can help us consider the types of scholarship historians can do in the future.
- Wong's chapter argues that renewed consideration of causation can help connect local stories with larger narratives.

Causal questions historians used to ask

- 1960s-1970s: historians tried to explain significant events, e.g. American revolution, French Revolution, Chinese Revolution.
- They debated what factors they deemed most important, e.g. economic relationships and material conditions, ideas and institutions. Because revolutions are composed of "complex sequences of events in varied contexts" the causal significance of each factor was difficult to disentangle.
- Other reasons to turn away: Chinese revolution no longer appeared as successful; recognition of historical continuities; work on how revolutions were remembered in understanding cultural sensibilities in recent times.
- Similar issues in assigning causes to the Industrial Revolution, and its significance was diminished by economic historians in the 1980s. Part of a longer-run story of economic change in Europe.

Causation in large scale historical narratives

- Comparative work has sought to identify causes of economic development outside Europe by looking at causes identified for North America and Western Europe: legal systems with written contracts and Western-style firms, banking institutions, encouraging entrepreneurs...
- Formation of national states and industrial economies: two processes as backdrop of social change. Social history in 1960s and 1970s considered how common people "lived the big changes" (Tilly). Used quantitative data and focused on ordinary people.
- If actors had little influence over their conditions, a focus on causes of their conditions became irrelevant.
- "Studies of causes for a set of conditions can complement but cannot contradict the experiences of people who lived under those conditions."
- People's actions will have consequences, "and in a very simple and direct sense be caused by their efforts," intended or not. Example of panicked house sales.
- Price theory, insofar as it can predict outcomes independent of explaining preferences, "succeeds at certain kinds of causal explanation."

Causation in large scale historical narratives

- Historians may try to recreate these "sensibilities" without causation, but these reconstructions cannot substitute for causal accounts.
- Work in the 1980s and 1990s focused on reconstructing historical sensibilities, not causation.
- Study of large-scale processes became less popular: some historians had less belief in unilinear models of development, others were concerned that explanations for events could serve the beliefs or interests of specific actors.
- Historians began to doubt how much existed independently of the subjective perspectives of individuals.
- "Unless historians grant the existence of a world out there in which things happen that can be explained without reference to the observer's position, epistemological variety undermines belief in causal explanations."
- Difference from science is not the fact that theories are socially constructed, but degree to which a set of propositions gains general acceptance.

Causation in large scale historical narratives

- Scientists use theory to "simplify phenomena into manageable sets of features," while historians seek to "celebrate the complexity and uncertainty of their topics, as well as the creativity and uniqueness of the actors they present"; devotion to affirming multiple causes makes identifying a few causes unattractive.
- Causation eclipsed: grand narratives lose their appeal; limitations of materialism as an approach; complexity of historical change; topics not suited to causation.
- This turn happened at the same time as the "conceptual poverty" of the "end of history" and "clash of civilizations" theses; Wong believes historians could seek to replace these with accounts of historical change that consider causality.

Recovering causation after the linguistic and cultural turns

- Suspicion of European history's developmental narrative as context for understanding others $\Rightarrow$ studies like Chakrabarty's history of Bengali intellectuals in the 1920s and 1930s, and how they created a subjective universe "separate from the world of power and wealth dominated by Westerners."
- Chakrabarty, like many scholars, sees the global spread of capitalism as defining much of the context for his subjects but, since Bengali intellectuals cannot influence it, issues of causation are not important.
- "A world made by materialist processes still existed ... but it no longer mattered directly to what historians preferred for their subjects of enquiry."
- Grand narratives rejected, but not replaced.
- Wong: we can still consider the choices people make and to what effect.
- "Such efforts can build on the cultural and linguistic turn in historical studies to renew and reframe efforts to address the changing patterns of political, economic and social experiences in different places and periods of history."

Recovering causation in history after the cultural turn

- It is plausible that people's ideas about their states and economies influence beliefs and institutions, which then affects political and economic activities.
- Chinese industrialization, for example, has depended on policy decisions and the choices of individuals.
- Why did the PRC in the 1950s move capital and technology to the villages to take advantage of underemployed rural labor, but India did not? The history of rural craft industries that had created experience of market networks, exchange, and production.
- Culturally-specific ideas and institutions can explain similar events with similar motivations are not part of a shared set of common historical changes, e.g. conflicts over food in Europe and China (different ideas of fairness, different priorities of states...).
- Comparable episodes in different paths of change $\Rightarrow$ similar causal chains within larger patterns of difference.

Bringing causation back into cultural and social history

- Can more explicit attention to causation enrich studies of culture? Many of these highlight both local and global elements.
- Wong digresses to argue that national units are rarely appropriate for studies of cultural and social practices.
- Subjects like food and music can be the products of causal processes.
- E.g. European classical music underwent innovation, some due to technology (e.g. harpsichord to piano forte) and others due to recombination and extension of ideas.
- Other genres of music have their own patterns of change in different locations. e.g. post-WW2 U.S. rock music has its own antecedents and can be traced into other cultural settings. There are "evolving taxonomies" that reflect local tastes.
- Global and local: one artist may sing in many languages, record in one country, be represented in another, etc...

Bringing causation back into cultural and social history

- Similar discussion of cuisine. Point? "As we discover and explore the diversity and complexities of music and cuisines in various places and many periods of history, we can also, should we wish, show the causal mechanisms responsible for the changes over time as well as the similarities and differences across space."

Constructing causal accounts in future historical writing

- Ability to construct multiple interpretations and recover contingencies have helped motivate a move away from simple narratives.
- Shift from grand narratives to global framings has allowed historians to connect local subjects to larger themes without causal arguments.
- Similar patterns in multiple sites may be either parallels or connected.
- Regions helpful for organizing important aspects of a situation need not be national or extremely local.
- Tension between global and local related to that between wanting coherent narratives of long periods and wanting nuance and complexity.
- One way to address the gap: look at "turning points," identified after the fact, from the vantage points of the actors themselves.
- Result of consciously looking for causation: leave behind national narratives as organizing principle; incorporate European past into larger history rather than rejecting Eurocentric studies.

Causality, Internal Validity, and Threats

- I will lecture mostly from Guo and Fraser (2009) "Propensity Score Analysis: Statistical Methods and Applications," ch. 2 "Conceptual framework and assumptions."
- The key question in program evaluation: "To what extent can the net difference observed in outcomes between treated and non-treated groups be attributed to the intervention, all else held constant?"
- Rubin (1986, JASA): "No causation without manipulation."
- A correlated with B may result from:
  - C determines A and B.
  - A causes B.
  - B causes A.
- Paul Lazarsfeld's definition of causation:
  - A must precede B in time (similar to Hume).
  - A must be empirically correlated with B.
  - The correlation of A and B cannot be explained away as the result of a third variable.

Causality, Internal Validity, and Threats

- Internal validity (Campbell, 1957, PB): inferences about whether correlation between A and B.
- GF list nine threats to internal validity. Wikipedia lists 13:
  1. Ambiguous temporal precedence
  2. Confounding (a third variable).
  3. Selection bias (pre-treatment group differences).
  4. History (events outside the study).
  5. Maturation (subjects change during the experiment).
  6. Repeated testing (e.g. Hawthorne, learning)
  7. Instrument change (i.e. change in measurement devices)
  8. Regression toward the mean
  9. Mortality/differential attrition
  10. Selection-maturation interaction
  11. Diffusion (SUTVA).
  12. Compensatory rivalry/resentful demoralization (response of control group).
  13. Experimenter bias (double blind).

Counterfactuals and the Neyman-Rubin counterfactual framework

- "A has caused B" as "B would not have occurred if it were not for A."
- "Potential outcomes" framework from Neyman-Rubin: the outcome that person i would have under treatment (1) or control (0) is given by Y_{0i} and Y_{1i}.
- Let treatment be denoted by $W_i \in \{0, 1\}$. Then the measured outcome is:

$$Y_i = W_i Y_{1i} + (1 - W_i) Y_{0i}$$

- "The fundamental problem of causal inference": Y_{0i} not observed for the treated (nor Y_{1i} for the untreated).
- Consider a very simple case, with no covariates or threats to internal validity. Then the *standard estimator for the average treatment effect* is:

$$\tau = E(Y_1|W = 1) - E(Y_0|W = 0)$$

- That is, use $E(Y_0|W = 0)$ to estimate the counterfactual $E(Y_0|W = 1)$.

Counterfactuals and the Neyman-Rubin counterfactual framework

- First issue: Note that this approach requires many assumptions, including the *ignorable treatment assignment assumption* and the SUTVA (discussed below).
- Second issue: The standard estimator is a weighted average of the treatment effects for the treatment group and the control group:

$$
\begin{aligned}
\bar{\tau} &= \pi E(\bar{\tau}|W=1) + (1-\pi)E(\bar{\tau}|W=0) \\
&= \pi[E(Y_1|W=1) - E(Y_0|W=1)] \\
&+ (1-\pi)[E(Y_1|W=0) - E(Y_0|W=0)] \\
&= \pi E(Y_1|W=1) + (1-\pi)E(Y_1|W=0) \\
&- \pi E(Y_0|W=1) - (1-\pi)E(Y_0|W=0) \\
&= E(Y_1) - E(Y_0)
\end{aligned}
$$

- Here, π is the proportion treated.

Counterfactuals and the Neyman-Rubin counterfactual framework

- Not everything here is observed. But, if we assume that $E(Y_1|W = 1) = E(Y_1|W = 0)$ and $E(Y_0|W = 1) = E(Y_0|W = 0)$, then the above can be rewritten as:

$$E(Y_1|W = 1) - E(Y_0|W = 0)$$

- So, this randomization assumption is sufficient for the standard estimator to consistently estimate the true treatment effect.
- Third issue: can the randomization assumption be extended to analyses in social and health sciences? Selection bias.
- Fourth issue: Rubin extended the counterfactual framework to be more general, and apply to observational studies. These require dealing with assignment to treatment that is not exogenous.
- Fifth issue: treatment may be heterogeneous across units; the average may not well represent the treatment for any one unit.

Counterfactuals and the Neyman-Rubin counterfactual framework

- Sixth issue: Means have limitations. Heckman and others have suggested focusing on other summary measures, including proportion of treated who benefit relative to an alternative, or distribution of gains at base state values.
- Seventh issue: this can be extended to more complicated situations, e.g. stratification:

$$Y_{si} = W_{si} Y_{s1i} + (1 - W_{si}) Y_{s0i}$$

- Eighth issue: Though the Neyman-Rubin framework is a statistical tool, it does not rule out using theory.

The ignorable treatment assignment assumption

- Other names: unconfoundedness, selection on observables, conditional independence, and exogeneity:

$$(Y_0, Y_1) \perp W|X$$

- Key idea: assignment to one condition or another is independent of potential outcomes if observable covariates are held constant.
- Reasonable in a randomized experiment, often violated in quasi-experimental designs or observational studies.
- One common approach: Test whether W is correlated with any of the variables in X.
- This is the same as the assumption of contemporaneous independence or exogeneity in an OLS regression:

$$Y_i = \alpha + \tau W_i + X_i'\beta + e_i$$

- If e_i is correlated with W_i or X_i (e.g. due to non-random assignment, omitted variables, or measurement error), estimates of τ or β are biased

The SUTVA

- The stable unit treatment value assumption.
- Formally, consider N units indexed $i =, 1, ..., N$. There are T treatments indexed by $w = 1, ..., T$. The outcome variable Y is indexed $Y_{iw}(w = 1, ..., T; i = 1, ..., N)$. SUTVA is the assumption that the value of Y for unit i exposed to treatment w is the same no matter what treatments other units receive, for all i and all w.
- Violated when unrepresented versions of treatment exist or there is interference between units.
- Examples: spillover of fertilizer across plots in agricultural research. General equilibrium effects of job training programs.

Methods for estimating treatment effects

- GF summarize the seven models they will teach in the textbook. These are a mix of explicit modeling of the selection process (e.g. Heckman's selection model), propensity score matching, and propensity score weighting.
- James' view: these were more popular ten years ago, but still make sensible robustness checks.
- GF also discuss instrumental variables and regression discontinuity, which I will discuss below.
- GF discuss the underlying logic of statistical inference, which is not my focus.

Types of treatment effects

- 1. *Average treatment effect (ATE)*:

$$ATE = \tau = E(Y_1|W = 1) - E(Y_0|W = 0)$$

- Under certain assumptions:

$$ATE = E[(Y_1|W = 1) - (Y_0|W = 0)|X]$$

- 2. *Intent-to-treat (ITT) effect* .
- Think in terms of compliers, defiers, always-takers, and never-takers.
- The difference between the treatment and control group means is an unbiased estimator of the ITT (under randomization).
- 3. *Efficacy effect (EE)*.
- i.e. effect under conditions of ideal application; requires careful monitoring.
- 4. *Average treatment effect for the treated (TT)*:

$$E[(Y_1 - Y_0)|X, W = 1]$$

Types of treatment effects

- 5. *Average treatment effect for the untreated (TUT)*:

$$E[(Y_1 - Y_0)|X, W = 0]$$

- Useful for thinking about program extension.
- 6. *Marginal treatment effect (MTE)*.
- Useful special case: people at the margin of indifference.
- 7. *Local average treatment effect (LATE)*. This is the average causal effect for compliers, and is what is estimated by IV.
- Note:

$$EE \neq ITT(ATE) \neq TT$$

- GF outline the example of an intervention encouraging people with pulmonary disease to exercise, then measuring their lung function.

Treatment effect heterogeneity

- The ATE can be decomposed, showing that it is biased in the presence of pre-treatment heterogeneity and treatment heterogeneity:

$$
\begin{aligned}
ATE =& E(Y_0 - Y_1) \\
=& [E(Y_1|W=1) - (Y_0|W=0)] \\
& - [E(Y_0|W=1) - (Y_0|W=0)] \\
& - (TT - TUT)q \\
=& \text{difference in means} \\
& - \text{pretreatment heterogeneity bias} \\
& - \text{difference in treatment effect} \times \text{treatment rate}
\end{aligned}
$$

- Three motivations for considering:
  - Are effects significant for all subsamples?
  - Are effects the same for all subsamples?
  - Is there evidence for the strong ignorability assumption?

Checking the plausibility of the unconfoundedness assumption

- Not testable, but indirect approaches exist to assess plausibility.
- One approach, from Heckman and Hotz (1989):
  - Partition the covariates X into one variable V and the remainder, Z.
  - Treat V as an outcome, W as the treatment, and Z as the covariates.
  - There should be no treatment effect.
  - Particularly useful: lagged values of the outcome.
- Another method from Rosenbaum (1997).
  - Partition the control group in two.
  - Take one as treatment, and one as control.
  - There should be no treatment effect.

Taking the con out of econometrics?

- Angrist and Pischke (2010) "The Credibility Revolution in Empirical Economics" *JEP*: very brief summary, since this paper is on your reading list.
- They focus on four changes that have reduced the "con" in econometrics:
  - More and better data
  - Fewer distractions (e.g. functional forms, heteroskedasticity, serial correlation).
  - Better research design (e.g. Angrist and Lavy on class size, Tennessee STAR experiment).
  - More transparent discussion of research design.

Better research design

In applied micro fields such as development, education, environmental economics, health, labor, and public finance, researchers seek real experiments where feasible, and useful natural experiments if real experiments seem (at least for a time) infeasible. In either case, a hallmark of contemporary applied microeconometrics is a conceptual framework that highlights specific sources of variation. These studies can be said to be design based in that they give the research design underlying any sort of study the attention it would command in a real experiment.

Critiques

- Leamer:
  - External validity: or, can the Mariel boatlift tell us anything about building a wall along the southern US border?
  - Randomization is not enough: additive confounders matter in small samples, and interactive confounders always matter – heterogeneity and external validity.
  - More sensitivity analyses - "three valued logic."
- Keane:
  - Experimentalism is not a panacea for specification searches: all statistical inference relies on untestable assumptions.
  - Take using Maimonides' rule to estimate the effects of class size: what if good schools attract students? What if good parents react to class sizes when choosing schools? What if teachers assigned to new classes are not random?
  - We have a causal estimate - so what? Not useful if we have not estimated the structural parameters in the production of cognitive ability.
  - Specification searching impossible in structural work.

Long quote from Keane

Interestingly, it is easy to do natural experiments in marketing. Historically, firms were quite willing to manipulate prices experimentally to facilitate study of demand elasticities. But it is now widely accepted by firms and academics that such exercises are of limited use. Just knowing how much demand goes up when you cut prices is not very interesting. The interesting questions are things like: Of the increase in sales achieved by a temporary price cut, what fraction is due to stealing from competitors vs. category expansion vs. cannibalization of your own future sales? How much do price cuts reduce your brand equity? How would profits under an every-day-low-price policy compare to a policy of frequent promotion? It is widely accepted that these kinds of questions can only be addressed using structural models – meaning researchers actually need to estimate the structural parameters of consumers' utility functions. As a result, the "experimentalist" approach has never caught on.

Critiques

- Sims:
  - "Because economics is not an experimental science, economists face difficult problems of inference. The same data generally are subject to multiple interpretations. It is not that we learn nothing from data, but that we have at best the ability to use data to narrow the range of substantive disagreement."
  - "What the essay says about macroeconomics is mainly nonsense." VAR, SVAR, and estimated DSGE models have, for example, led to consensus on the effects of monetary policy in a literature where claims of natural experiments have not generally been successful.
- Nevo and Whinston:
  - "When sources of credible identification are available, structural modeling can provide a way to extrapolate observed responses to environmental changes to predict responses to other not-yet-observed changes."
  - Mergers as an extended example.

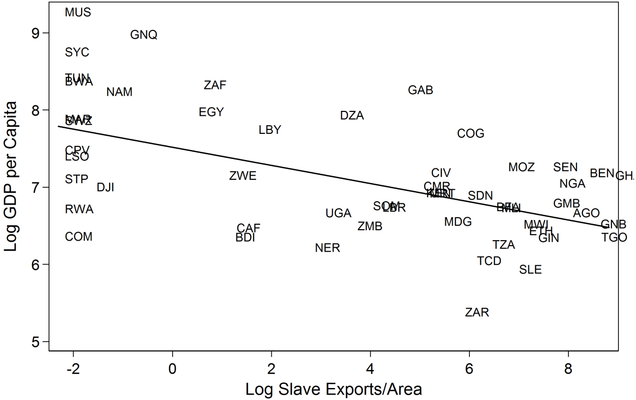
Simple OLS

- If you fit a line through y and x, you are assuming the two have a linear relationship. Wooldridge uses notation similar to:

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

  - y_i is the "dependent variable," "regressand," or "left-hand-side variable." In the next example this is log GDP per capita.
  - β_0 is the "intercept parameter" or "constant."
  - β_1 is the "slope parameter" or "beta" or "the coefficient on x."
  - x_i is the "independent variable," "regressor," or "right-hand-side variable." In the next example this is log slave exports per unit area.
  - u_i (often ϵ_i) is the "error term."
  - i subscripts index the observations $i = 1, 2, ...N$.
- Ordinary least squares (OLS) estimates β_0 and β_1 by minimizing the sum of the squared u_i implied by choices of β_0 and β_1.
- This is only unbiased if $E(u|x) = 0$.

Example from Nunn (2008)



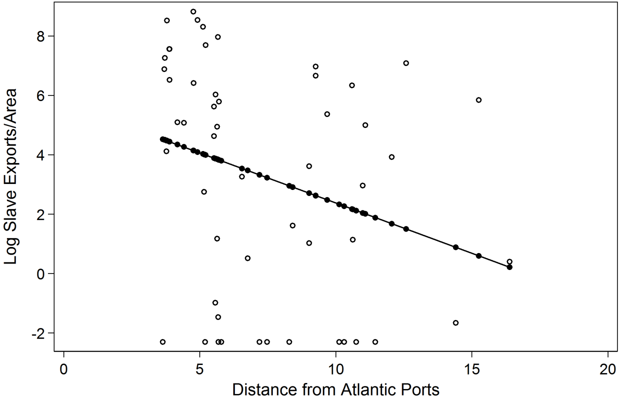
What to do about omitted variable bias?

- Approach 1: Ignore it. Sometimes correlations are informative.
  - Take the paper for the Week 1 computer practical, Ashraf and Galor's "Dynamics and Stagnation in the Malthusian Epoch".
  - Greater land productivity $\Rightarrow$ greater population density in 1500, but not higher income.
  - This is a "stylized fact" (broad generalization) consistent with Malthusian theory; it is a prediction of a model in data.
  - Another example: Acemoglu, Johnson, and Robinson's "Reversal of fortune".
  - Former colonies that were more densely populated in 1500 are poorer today.
  - They need not claim that population density in 1500 has a causal impact on income today in order for their results to show that a "reversal of fortune" has occurred, or to rule out a naive "geography explains everything" view.
- Approach 2: Control for more variables (Lecture 1).

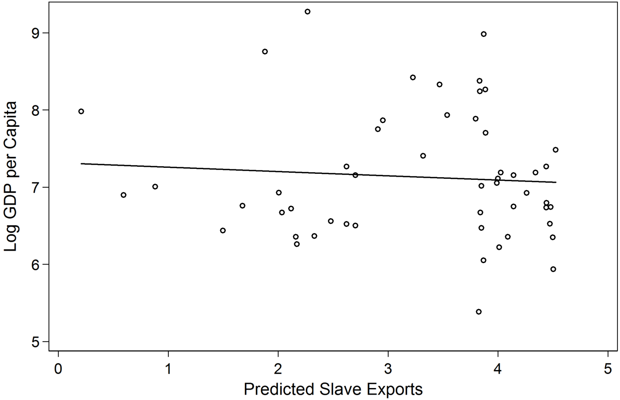
What to do about omitted variable bias?

- Approach 3: Use a randomized control trial.
  - Modern example: Miguel and Kremer's "Worms."
  - The effect of randomly-assigned deworming pills on school attendance in Kenya.
  - An historical example: Bleakley and Ferrie's "Shocking behavior" and "Up from Poverty?"
  - The impact of a land lottery in Georgia on fertility and long-run wealth.
- Approach 4: Use instrumental variables (Lecture 3).
  - Intuition: Use a variable that predicts your x of interest but is uncorrelated with u and has no direct effect on y to produce quasi-experimental variation in x.
  - Modern example: Angrist's "Lifetime earnings and the Vietnam era draft lottery" uses the Vietnam draft lottery to predict military service, in order to find the causal effect of military service on earnings.
  - Historical example: Nunn's "The long-term effects of Africa's slave trades" uses distance from ports of slave demand to predict slave exports, in order to find the causal effect of slave exports on long-run development.

Instrumental variables: Example of first stage



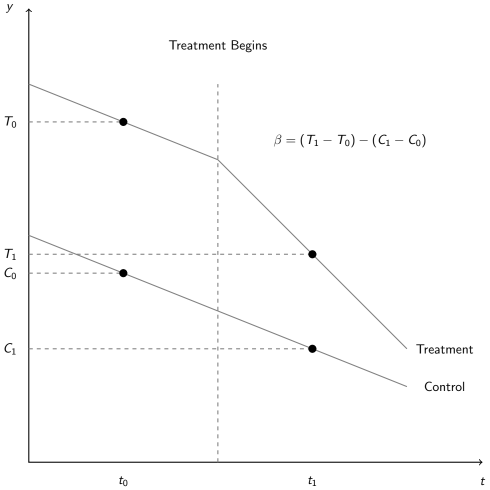
Instrumental variables: Example of second stage



What to do about omitted variable bias?

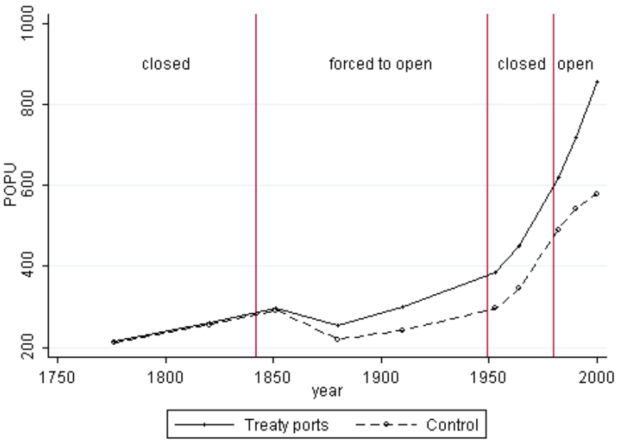
- Approach 5: Use difference-in-difference (Lecture 4).
    - Intuition: Compare the difference between the treatment and control groups before and after the treatment is instituted.
    - Historical example: Jia's "The Legacies of Forced Freedom: China's Treaty Ports."
    - Did the creation of treaty ports in China lead to greater population growth in treaty port prefectures, compared with other riverine and coastal prefectures?
- Approach 6: Use regression discontinuity (Lecture 5).
    - Intuition: Find a running variable where individuals just on one side of a cutoff receive treatment and individuals on the other side do not.
    - Modern example: Bharadwaj et al.'s "Early Life Health Interventions and Academic Achievement."
    - Babies just below 1500g birthweight receive extra medical treatment in Chile. How does this affect test scores in school?
    - Historical example: Dell's "Mining Mita."
    - Are areas just inside the boundary where forced labor was practiced in colonial Peru poorer today?

Difference-in-Difference: Hypothetical

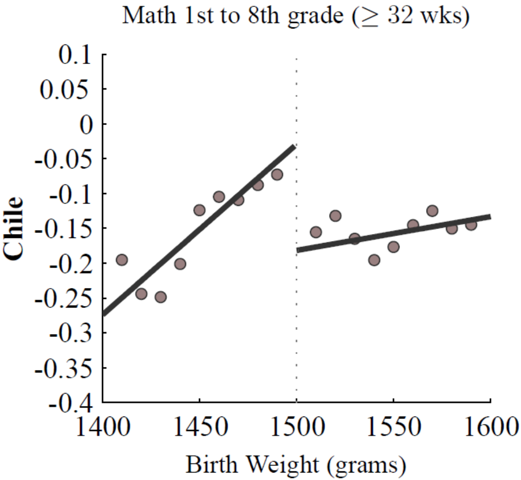


Difference-in-difference: Real example

FIGURE 2B. – TRENDS OF POPULATION SIZE FROM 1776 TO 2000: COMBINING WAVI



Regression discontinuity: Real example



What to do about omitted variable bias?

- Approach 7: Use some other tools, like propensity score matching and synthetic control (Lecture 6).
  - Intuition: Find the best approximation to a "control group" in non-experimental data.
  - Historical example: Abadie et al's "California's Tobacco Control Program."
  - How did Proposition 99 affect cigarette sales in California, compared with the most comparable parts of the US?
- Approach 8: Identify structural breaks in a time series (Lecture 7).
  - Intuition: Find jumps and trend breaks in a time-series variable, and look in the historical record to find what might have caused them.
  - Historical example: Guinnane et al's "Turning Points in the Civil War."
  - What events during the US civil war led investors to revise their views on the value of Greenbacks?
- Point? There are many strategies for dealing with the problem that correlation is not causation. My goal in QM2 is to teach you some of the more popular ones.

Synthetic control: Real example

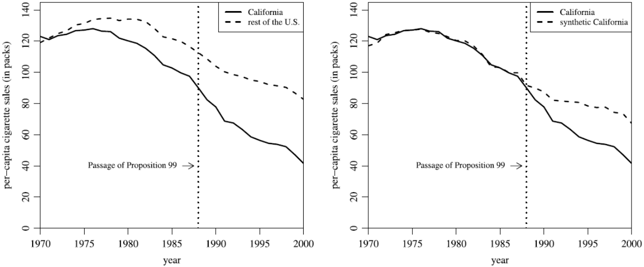


Figure 1. Trends in per-capita cigarette sales: California vs. the rest of the United States.

Figure 2. Trends in per-capita cigarette sales: California vs. synthetic California.

Structural breaks: Real example

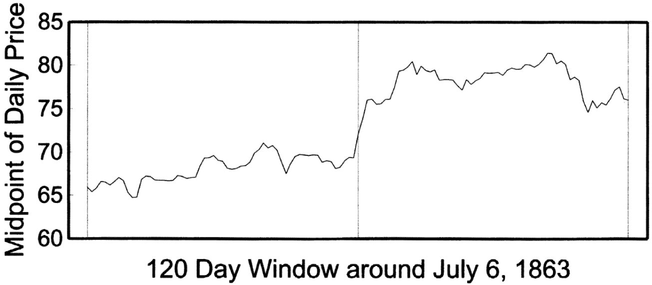


FIGURE 2. EXAMPLE OF A MEAN SHIFT IN THE GREENBACK/GOLD EXCHANGE RATE