Key Considerations and Questions in Research Design

Reflections and Comments Based on Angrist and Pischke (2009)
Chapter 1&2

Angrist, J. D., & Pischke, J. S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Before responding… How is archival research typically conducted in Accounting or Marketing?

- It is similar to Applied Economics
- It primarily uses reduced form approaches.

- ->>Focus
- Analyze the consequences of A (e.g., policy interventions)

Board composition and firm performance;

Incentive schemes and employee productivity;

Marketing strategies and firm profits;

…

Before responding… How is archival research typically conducted in Accounting or Marketing?

- A’s consequences on B> imply “effect”
- So, one of our core jobs is finding and tease out causality.

For this book:

What does this book suggest?
(Looks like Toolkits for..?)

What tools does it provide for intuitively establishing causality?
(DID,IV..?)

This book is like a cookbook-guide us to find causality. But why this is important?

- You establish a framework where the data is consistent.
- What if other frameworks produce the same data patterns?
- What if your assumptions are overly strong, imposing too many restrictions to fit the data?”

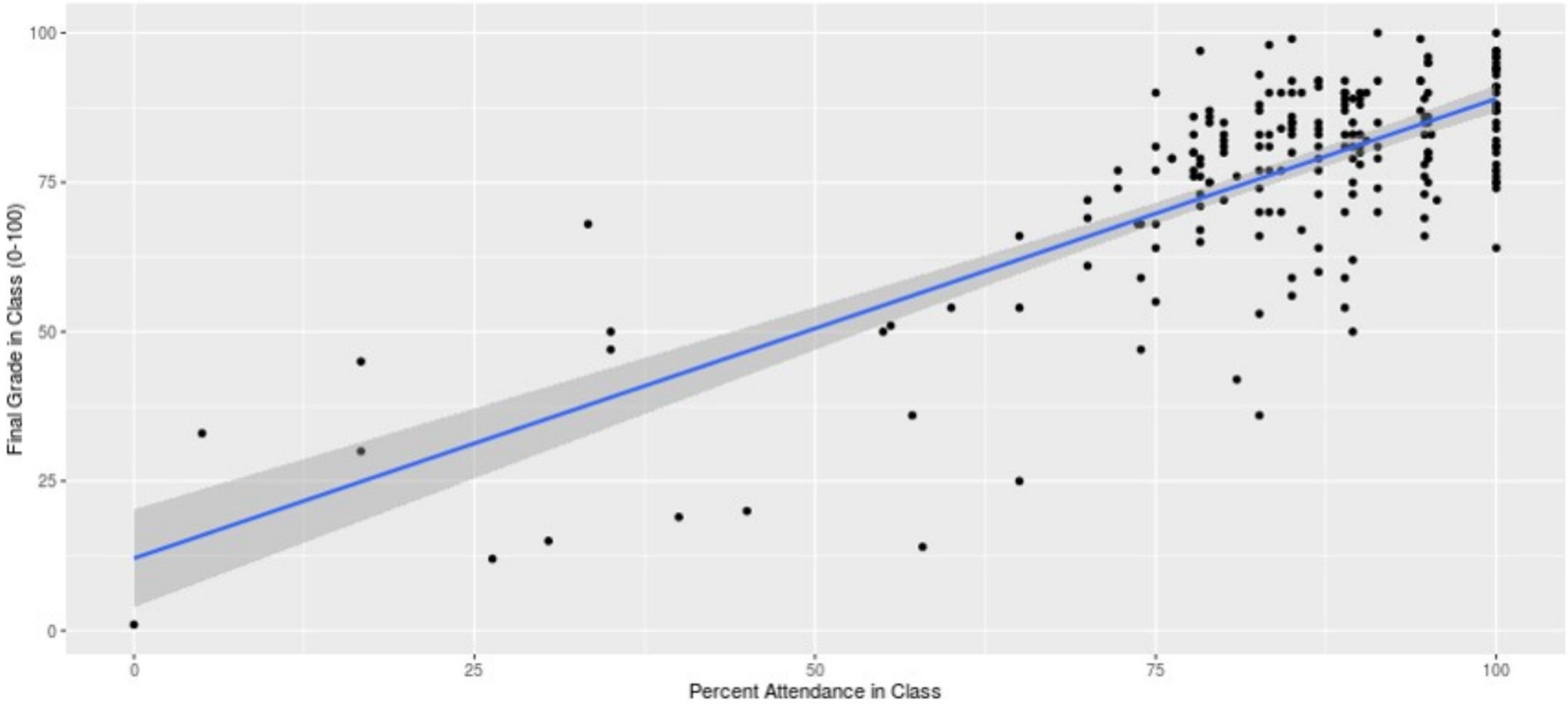
In response, it is better to let the data speak.

Their request: Provide **solid evidence** from the data to support and strengthen your conceptual framework.

Evidence for cause and effect might be the central thing if theory is not new

- First question this book suggest us to answer in our research design
- 1. **Causal** relationship of interest
- Regression analysis is a very useful tool to uncover basic statistical relationship in the data → Correlations
- The fact that A is correlated with B does not inform us about whether A causes B
- Correlations might be useful for predictions. . . but are of limited use in making policy recommendations
- Reason is …You want to measure the most likely outcome(consequence) associated with a (potential) policy intervention → Causality

Grade and Attendance



Can we say the attendance(A) effect on grade(B) in confidence?

Ability and motivation(C) correlated with A and B simultaneously.

Key point is to isolate the A we are interested in.

- How to isolate or only manipulate A?

- *"Causation is something that makes a difference, and the difference it makes must be a difference from what would have happened without it." – David Lewis*

Key point is to isolate the A we are interested in.

- How to isolate or only manipulate A?

- *“Causation is something that makes a difference, and the difference it makes must be a difference from what would have happened without it.” – David Lewis*

- >>**counterfactual**. Counterfactuals are neither past nor future. They are alternative histories created by thought experiments but we use them as framing devices to decipher causality in our timeline.

Key point is to isolate the A we are interested in.

- How to find counterfactuals in reality? Could we design experiment to approximate it?
- Second question we need to answer is …
- 2. **Ideal experiment type**
- First come up my mind is…what we learn from clinical studies

Clinical Study: Multicenter Randomized Controlled Trial (RCT)

- We can learn from their standards and methodology:
- Setting up treatment vs. placebo allows for comparison, with the placebo approximating counterfactuals.
- To achieve a clean effect:
- Ensure double-blinded random assignment, perfect compliance, uncontaminated subjects, no early medication, unit-treatment and replicable results, among other factors.
- In accounting or social science, see key assumptions when using treatment-control group in Armstrong et al. (2022)'s paper.

Armstrong, C. S., Glaeser, S. A., & Huang, S. (2022). Contracting with controllable risk. *The Accounting Review*, *97*(4), 27-50.

Clinical study: Multicenter RCT

- But setting up trials like clinical studies are time and money consuming, they need funding from National Institute of Health (NIH) and ethical approval (many paperwork…). For Econ, some funding are from NBER.
- ?Externality and Generalization
- Third one Angrist may let us know is
- 3. Identification strategy in using natural or quasi-experiment
- With archival data in reality, how to identify what we are interested in.
- (A's effect)

- We may already have heard Difference in Differences (DID), Regression Discontinuity(RD), PSM, IV, etc.

- If you are interested, several papers:
- DID, IV, Matching, Synthetic Control: (Border opening; Immigration shock; SEC Tick Size Pilot Program; newspaper closure)
- Beerli, A., Ruffner, J., Siegenthaler, M., & Peri, G. (2021). The abolition of immigration restrictions and the performance of firms and workers: Evidence from Switzerland. American Economic Review, 111(3), 976-1012.
- Dustmann, C., Schönberg, U., & Stuhler, J. (2017). Labor supply shocks, native wages, and the adjustment of local employment. The Quarterly Journal of Economics, 132(1), 435-483.
- Hope, O. K., & Liu, J. (2023). Does stock liquidity shape voluntary disclosure? Evidence from the SEC tick size pilot program. Review of Accounting Studies, 28(4), 2233-2270.
- Heese J., Perez-Cavazos G., Peter C. (2022). When the local newspaper leaves town: The effect of local newspaper closures on corporate misconduct. Journal of Financial Economics
- Issues in using natural experiments and staggered/dynamic DID:
- Roth, J., Sant'Anna, P. H., Bilinski, A., & Poe, J. (2023). What's trending in difference-in-differences? A synthesis of the recent econometrics literature. Journal of Econometrics, 235(2), 2218-2244.
- Heath, D., Ringgenberg, M. C., Samadi, M., & Werner, I. M. (2023). Reusing natural experiments. The Journal of Finance, 78(4), 2329-2364.
- RD (Voting 50%/School Attendance Seasons):
- Cuñat, V., Gine, M., & Guadalupe, M. (2012). The vote is cast: The effect of corporate governance on shareholder value. The journal of finance, 67(5), 1943-1977.

- Also, clinical study use Mendelian Randomization besides RCT.
- Key: use genetic variation as IV to set up random assignment groups.
- Sanderson, E., Glymour, M. M., Holmes, M. V., Kang, H., Morrison, J., Munafò, M. R., ... & Davey Smith, G. (2022). Mendelian randomization. *Nature Reviews Methods Primers*, *2*(1), 6.

- 4. Mode of statistical inference
- Considerations about clustering...

Today's focus: ideal experiment setting (Chapter II)

- Back to causality
- Does Drug X reduce blood pressure? If we try to figure out, how to design an experiment?

Today’s focus: ideal experiment setting

- Back to causality
- Does Drug X reduce blood pressure? How to design?

❑1. Find a set of 50 subjects that suffer from high blood pressure but not using medications for a window period. Excluding existing conditions that will have interactions with Drug X.

❑2. Randomly divide the subjects into two groups of equal size.

❑3. Drug X will be administrated to individuals in group one and a placebo pill to individuals in group two. But neither patients nor I will know the group allocation.

❑4. Measure the blood pressure after a reasonable number of weeks. Ensure they follow the standard of compliance

❑5. Take average blood pressure of each group. Use test for the difference between two means to verify if Drug X reduces blood pressure.

Does having a health insurance improve your health?

Can we compare two groups then find the effect?

- Does having a health insurance improve your health?

- We cannot randomly assign who can get insurance.
- Main problem: if the group of individuals with no insurance has different characteristics from the group with insurance, then our measurement of the treatment effect may be mistaken. → selection bias
- The drug trial example didn't have this problem because the subjects in both groups were randomly selected

Selection bias

$$\underbrace{E\left[\mathrm{Y}_i|\mathrm{D}_i = 1\right] - E\left[\mathrm{Y}_i|\mathrm{D}_i = 0\right]}_{\text{Observed difference in average health}} = \underbrace{E\left[\mathrm{Y}_{1i}|\mathrm{D}_i = 1\right] - E\left[\mathrm{Y}_{0i}|\mathrm{D}_i = 1\right]}_{\text{average treatment effect on the treated}}$$

$$+\underbrace{E\left[\mathrm{Y}_{0i}|\mathrm{D}_i = 1\right] - E\left[\mathrm{Y}_{0i}|\mathrm{D}_i = 0\right]}_{\text{selection bias}}$$

It represents the difference in average outcomes between the treated group and the untreated group

It is the difference between the expected outcome if treated versus if not treated, but only for those who were actually treated.

This difference arises because the groups may differ in ways that affect the outcome, independent of the treatment (e.g., healthier individuals might be more likely to opt into treatment).

Individuals’ (rational) choice always translate into selection bias

I chose to get a PhD because I didn’t like my life – i.e., Y^0 maybe was different for me than others
I chose to get a PhD because I thought it would help me – i.e., Y^1 maybe was different for me than others

Bias- IFRS mandate for R's disclosure

- A Firms-can earn t when disclosure R--50% in simulated economy
- B Firms-can loss t when disclosure R--50% in simulated economy
- Rational: Maximization of earning; have right to choose in Pre

Acct Research	Pre	Post	Difference
A-placebo	λ+t	λ+t	0
B-treated	λ	λ-t	-t
Lab Setting	Pre	Post	Difference
placebo	λ	λ	0
treated	λ	λ	0

Bias- IFRS mandate for R's disclosure

- Effect of IFRS mandate of diversity rule on performance
- ~~Effect of board diversity on performance~~
- E.g. Only add 1 female into the board--tokenism

Acct Research	Pre	Post	Difference
A-placebo	λ+t	λ+t	0
B-treated	λ	λ-t	-t
Lab Setting	Pre	Post	Difference
placebo	λ	λ	0
treated	λ	λ	0

Shock is not necessarily to be random
random is the core

Why randomization can address the bias? Independence assumption

Treatment is assigned to a population independent of that population's potential outcomes

$$(Y^0, Y^1) \perp\!\!\!\perp D$$

This is random or quasi-random assignment and ensures mean potential outcomes for the treatment group and control group are the same. Also ensures other variables are distributed the same for a large sample.

$$\begin{aligned} E[Y^0|D=1] &= E[Y^0|D=0] \\ E[Y^1|D=1] &= E[Y^1|D=0] \end{aligned}$$

Key takeaways

- Causality is defined by potential outcomes, not by realized (observed) outcomes
- Observed association is neither necessary nor sufficient for causation (problem of selection bias)
- Estimation of causal effects of a treatment (usually) starts with studying the assignment mechanism

Further discussion:
Now that We’ve shown data/variable causality

- Our research design needs to answer our research questions
- Does the variable we construct accurately reflect what our conceptual framework intends to capture?

- How do we balance causality with the impact of our ideas?
- “For example, on one hand, there is the risk of misinterpreting or misusing correlations from studies lacking proper identification, especially in policy decisions. On the other hand, we might have studies focusing on 'cute' settings that provide precise estimates for relatively minor questions.”

Thanks!

Critics and constructive feedbacks are welcomed.