Week 10: Overview and Outlook

PUBL0050 Causal Inference

Julia de Romémont

Term 2 2023-24

UCL Departement of Political Science

A reminder about the use of AI in your assessment

Can I use AI tools in my quantitative methods assignments?

Yes, but only for certain tasks:

- **Coding**: To correct errors in your code or to solve specific, common coding problems and to improve on the appearance of tables and figures.
- **Writing**: To help improve your writing, including greater clarity or more accurate grammar.
- **Generally**: To support your efforts to resolve conceptual queries, although you should always make use of your classes, support and feedback hours, and moodle forums first.

A reminder about the use of AI in your assessment

Can I use AI tools in my quantitative methods assignments?

This means you cannot use it:

- To write parts or all of an assessment;
- To write parts or all of your code;
- To generate outlines, structures and high-level arguments for essays;
- For rewriting or paraphrasing text from other sources for use in written work.
- **Under no circumstances** should you upload any course material to ChatGPT or other other GenAI tools.

Can I use AI tools in my quantitative methods assignments?

All use of AI must be acknowledged, described and referenced in your essay.

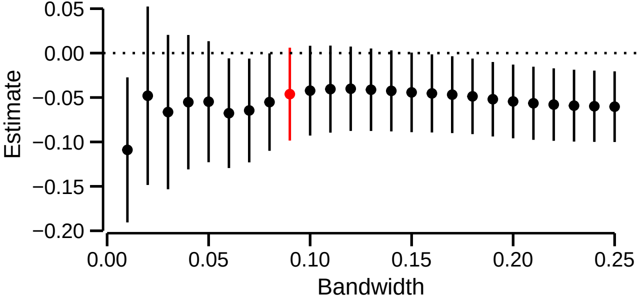
For example, when using ChatGPT to improve a figure:

```
# Plot (OpenAI 2024)
ggplot(results, aes(x= bw, y= est, color=opt)) +
  geom_point() +
  geom_linerange(aes(ymin=lo,ymax=hi)) +
  geom_hline(yintercept = 0, linetype="dotted") +
  scale_x_continuous("Bandwidth",breaks = seq(0,0.25,.05)) +
  scale_color_manual(values = c("black","red")) +
  ylab("Estimate") +
  theme_clean() +
  lemon::coord_capped_cart(bottom="both",left="both") +
   theme(plot.background = element_rect(color=NA),
        panel.grid.major.y = element_blank(),
        legend.position = "none",
        axis.ticks.length = unit(2,"mm"))
```

Can I use AI tools in my quantitative methods assignments?

All use of AI must be acknowledged, described and referenced in your essay.

For example, when using ChatGPT to improve a figure:



Source: Author's calculation, OpenAI (2024)

A reminder about the use of AI in your assessment

Can I use AI tools in my quantitative methods assignments?

All use of AI must be acknowledged, described and referenced in your essay.

For example, when using ChatGPT to improve a figure:

Acknowledgement
I acknowledge the use of ChatGPT in the coding of figure XX. The prompt I entered was: "How can I add a horizontal dotted line where the Y-axis is zero?", and ChatGPT suggested the following code: `geom_hline(yintercept = 0, linetype="dotted")`.

References
OpenAI (2024). ChatGPT (20th March version). Large Language model, https://chat/.openai.com/chat

This Lecture

Review and Summary

Criticisms of "causal empiricism"

What next?

Conclusion

Review and Summary

Potential outcomes

> *The strength of [the potential outcomes framework] is that it allows us to make assumptions more explicit than they usually are. When they are explicitly stated, the analyst can then begin to look for ways to evaluate or partially test them.*
>
> *– Holland, 1986*

- The potential outcomes model provides an organising framework for thinking about causal analysis.
- Each method we have studied has been motivated by the counterfactual comparisons that lie at the heart of that framework.
- The framework provides clarity over the requirements for making causal statements on the basis of quantitative data.

A common logic

Each topic of the course can be broadly decomposed into four parts:

- **Estimands**
  - What is the treatment effect we are trying to estimate?
  - ATE; ATT; LATE; etc
- **Assumptions**
  - Which assumptions are required to give the resulting estimates a causal interpretation?
  - What tests or checks can and should be done to inspect the plausibility of these assumptions?
- **Data requirements**
  - What are the minimum data requirements for implementing the design?
- **Estimation**
  - Which techniques are used to estimate the causal effects?

Estimands

1. **Randomized Experiments**
   - ATE $=$ ATT
2. **Selection on Observables**
   - ATE or ATT or ATC (or weighted ATE for OLS)
3. **Difference-in-differences**
   - ATT
4. **Synthetic control**
   - Treatment effect for unit 1 at time t (TT_{1t})
   - Average treatment effect for unit 1 (ATT_1)
5. **Instrumental variables**
   - LATE (local to compliers)
6. **Regression discontinuity**
   - Sharp: LATE (local to units at the discontinuity)
   - Fuzzy: LATE (local to complying units at the discontinuity)

Assumptions

1. **Randomized Experiments**
   - Valid in expectation: Independence of Y_{0i}, Y_{1i} & D_i
2. **Selection on Observables**
   - Untestable: Independence of Y_{0i}, Y_{1i} & D_i, conditional on X_i
   - Testable: Common support $0 < \Pr(D = 1|X) < 1$ for all X
3. **Difference-in-differences**
   - Untestable: Parallel trends
4. **Synthetic control**
   - Untestable: Equal trends
5. **Instrumental variables & Fuzzy regression discontinuity**
   - Valid in expectation in experiment: Independence of instrument
   - Untestable in observational data: Conditional independence of instrument
   - Testable: First-stage
   - Untestable: No defiers & exclusion restriction
6. **Regression discontinuity**
   - Untestable: Smoothness of potential outcomes at the discontinuity.

1. **Randomized Experiments**
   - Independence of Y_{0i}, Y_{1i} & D_i
     - Check covariate balance on observables with logistic regression or Fisher's randomisation test
2. **Selection on Observables**
   - Independence of Y_{0i}, Y_{1i} & D_i, conditional on X_i
     - Matching: Check post-matching covariate balance visually and/or with mean absolute balance
     - Regression: Check for how "fragile a result is against the possibility of unobserved confounding"[1]
   - Common support $0 < \Pr(D = 1|X) < 1$ for all X
     - Calculate probablity of treatment (=propensity scores) dependent on covariates with logistic regression

[1] cf. Cinelli & Hazlett, 2020 via sensitivity analysis. A relatively recent R package is available for this: `sensemakr`.

Assumptions checks and tests

3. **Difference-in-differences**
   - Parallel trends
     - Check pre-treatment trends visually and/or via lags and leads
4. **Synthetic control**
   - Equal trends
     - Check pre-treatment fit between treated unit and synthetic control visually and/or pre-treatment RMSPE

Assumptions checks and tests

5. **Instrumental variables & Fuzzy regression discontinuity**
   - Independence of instrument
     - cf. randomised experiments
   - Conditional independence of instrument
     - cf. selection on observables with regression
   - First-stage
     - Test strength of first stage with F-test
   - No defiers
     - Discuss the likelihood of there being defiers
   - Exclusion restriction
     - Discuss possible alternative causal pathways that would have the instrument have an effect on the outcome

6. **Regression discontinuity**
   - Smoothness of potential outcomes
     - check for discontinuity in observable covariates around cut-off
     - check for discontinuity in density around cut-off with McCrary test
     - check for discontinuity at other levels of the running variable with placebo tests

Data requirements

1. **Randomized Experiments**
   - Cross-section of D_i and Y_i
2. **Selection on Observables**
   - Cross-section of D_i, Y_i, and **all relevant** X_i
3. **Difference-in-differences**
   - At least 2 repeated cross-sections or panel waves of D_i and Y_i
   - Strengthened by time-varying X_i
   - Strengthened by multiple pre-treatment period observations
4. **Synthetic control**
   - Several pre-treatment cross-sections or panel waves of D_i, Y_i and X_i
5. **Instrumental variables**
   - Cross-section with D_i, Y_i, Z_i and sometimes X_i
6. **Regression discontinuity**
   - Cross-section with D_i, Y_i and X_i where X_i is marked by a discontinuity

Estimation

1. **Randomized Experiments**
   - t-test; regression
2. **Selection on Observables**
   - Matching; regression
3. **Difference-in-differences**
   - Regression with first-difference Y
   - Fixed-effects regression
4. **Synthetic control**
   - Synthetic control method
5. **Instrumental variables**
   - Wald estimator
   - Two-stage least squares
6. **Regression discontinuity**
   - Sharp: Regression within bandwidth window around discontinuity
   - Fuzzy: 2SLS within bandwidth window around discontinuity

Causal inference designs – summary

	Estimand	Data	Assumptions	Estimation
Randomized Experiment	ATE	Cross-section of Y and D	Independence of Y_{0i}, Y_{1i} and D	t-test; regression
Selection on observables	ATE, ATT, ATC	Cross-section of Y, D & all X	Selection on obs. Common support	subclassification; matching; regression
Diff-in-Diff	ATT	At least 2 cross-sections of Y and D	Parallel trends	1^{st} diff regression; Two-way FE
Synthetic control	TE for unit i	Several cross-sections	Equal trends	SC method
Instrumental variables	LATE	Cross-section of Y, D & Z (& X)	Independence of Z First stage No defiers Exclusion restriction	Wald Estimator; 2SLS
Regression discontinuity	LATE	Cross-section of Y, X & $D = 1\{X > c\}$	Smoothness of Y_{0i}, Y_{1i} at threshold	OLS/2SLS within bandwidth

Criticisms of “causal empiricism”

"Causal empiricism is often understood in terms of deep consideration – some might say an obsession – with 'causal identification' and clear definition of counterfactual comparisons."

– Samii, 2016, p. 941

A (somewhat unfair) distinction

1. Conventional regression
    - Assemble data on interesting X and Y variables and run a regression
    - Interpret coefficients as average causal effects
    - "loosely specified and heroically interpreted" regressions **(Samii, 2016)**
    - "the magic regression machine" **(Deaton, 2015)**

A (somewhat unfair) distinction

2. Causal empiricism
    - Clear description of assumptions necessary for causal interpretation
    - Identification of variation in D needed to determine a causal effect
    - Use of various empirical techniques to exploit this variation
    - Careful description of the population to which estimated effects apply
    - The focus of this course!

Criticisms of "causal empiricism"

1. Internal vs external validity
2. Effects of causes vs causes of effects
3. Identification vs importance

Criticism

Causal empiricist approaches have higher levels of internal validity, but conventional approaches have higher levels of external validity.

- Internal validity → Are the causal claims valid in this particular study?
- External validity → Do the conclusions from this study generalize to populations of greater interest?

The external validity critique takes several forms:

- Samples are often not representative but based on convenience
- Treatments are often not what we care about, but what is plausibly random
- Treatment effects are often only defined for some units (compliers; units at the discontinuity; etc)

Internal vs external validity – Rebuttal

Conventional regressions do not automatically produce generalizable effects!

- When selection on observables holds, regression estimates a **conditional-variance weighted ATE (MHE, p. 74-76)**
- The effective sample implied by weighted regression effects can be *very* different from the population **(Aronow & Samii, 2016)**
- Implication: Even if the **data** is representative of the population, the estimated **treatment effects** may not be!

Internal vs external validity – Rebuttal

- Recall that in matching the ATE is a weighted averages of δ_x:

$$\tau_{\text{ATE}} = \sum_x P(X_i = x)\delta_x$$

  i.e. where the weights are the distribution of X_i in the population (τ_{ATE})

- But the estimates for β from an OLS regression of Y on D and X are a bit different:

$$\beta_{\text{OLS}} = \sum_x \frac{Var[D_i = 1|X_i = x]P(X_i = x)}{\sum_x Var[D_i = 1|X_i = x]P(X_i = x)}\delta_x$$

- Given this weighting scheme, regression estimates of β_1 are based on an effective sample that can be very different from the full sample.
- Aronow and Samii (2016) show that OLS implicitly weights each unit's contribution to the estimate of β_1 by:

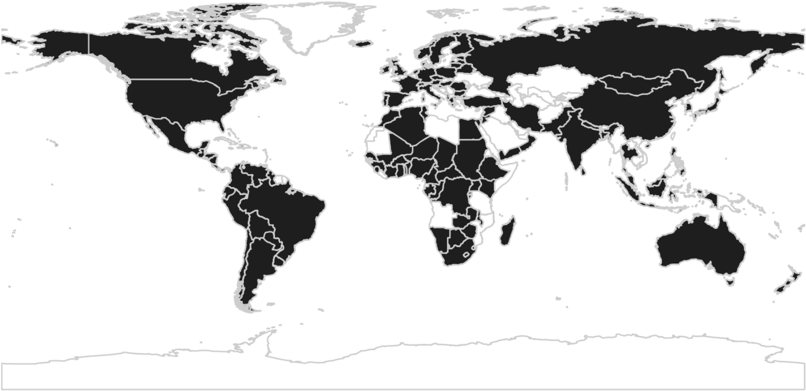
$$w_i = (D_i - E[D_i|X_i])^2$$

Implications:

- More weight on units whose treatment assignment is not well explained by the covariates. i.e. units in subsets of X_i with higher conditional treatment variance
- "Effective sample" may be very different from population of interest.

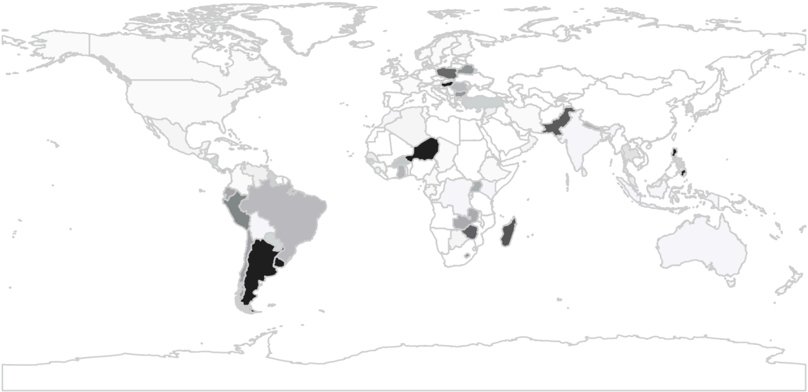
Internal vs external validity – Rebuttal

Nominal Sample



Internal vs external validity – Rebuttal

Effective Sample



Internal vs external validity – Rebuttal

There is no clear ordering of experiments, quasi-experiments, and observational studies that use regression or other control methods in terms of the generality of their findings.

–Samii, 2016, p. 945

Effects of Causes vs Causes of Effects – Criticism

- In this course, we have focused on questions of the type:

  What is the effect of D_i on Y_i?

  - This is known as an **"effects of causes"** question.

- A plausible alternative causal question type:

  What are the causes of Y_i?

  - This is known as a **"causes of effects"** question.

Criticism

Causal empiricist approaches in social science are too focused on effects-of-causes questions, when often we care about the many causes of a given phenomenon.

Effects of Causes vs Causes of Effects – Rebuttal

1. Social science before the 2000s developed almost exclusively "causes of effects" studies, many of which were poorly identified and resulted in the generation of "pseudo-facts" (Samii, 2018)
2. Even if individual causal studies address "effects of causes" questions, the collective endeavor of causal inference can lead to more satisfying answers about the many causes of effects
   - Pursue well-defined questions about the effects of some cause
   - Accumulate knowledge about which causes have effects on some outcome of interest
   - Build a full causal account of all the effects on that outcome

The counterfactual approach…is well suited to this pragmatic account of social science research, where progress results from credible advances rather than grand claims.

–Morgan and Winship, p. 443

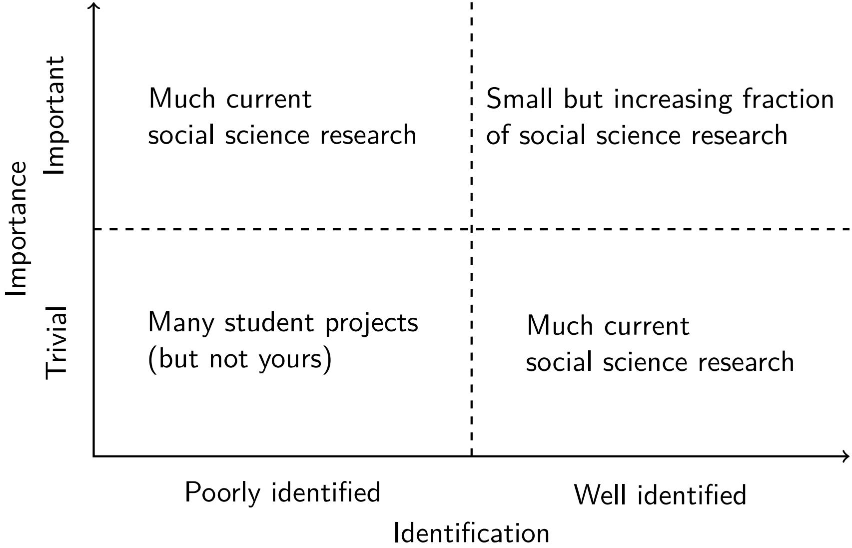
Identification vs Importance

- A more reasonable concern, in my view, is that there is potentially a trade-off between causal identification and research question importance.
- This particularly manifests in cases where data limitations prevent the application of the types of strategy we have employed on this course.
- Areas where causal empiricist approaches tend to struggle:
  - Questions about long-term effects of treatments
  - Questions about general equilibrium effects of treatments
  - Questions about large, structural, and/or slow moving treatments (i.e. institutions; climate change; norms; race)

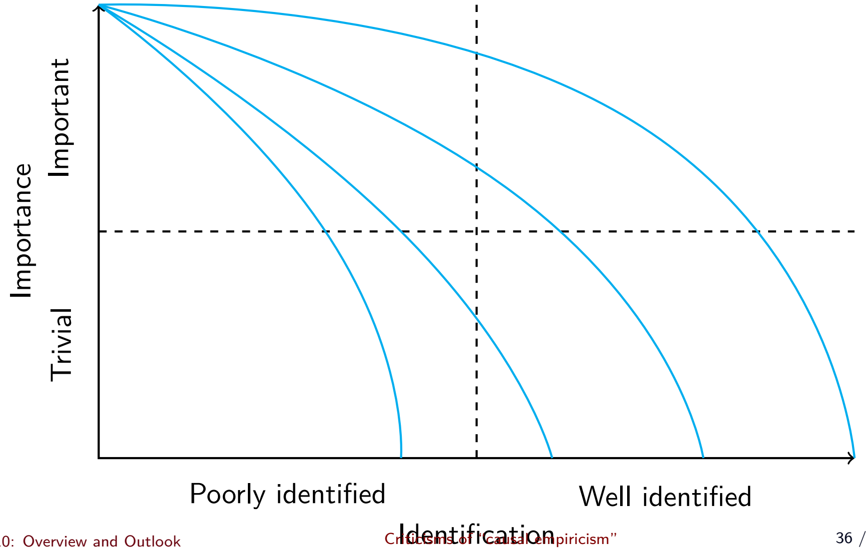
In many cases, cleanly identified research strategies answer relatively narrow questions.

–Ruhm, 2018

Identification vs Importance



Identification and Importance "Frontier"



Identification and Importance "Frontier"

- Causal empiricist approaches are more demanding in terms of data requirements and assumptions than conventional approaches
- By necessity, this implies a trade-off between identification and question importance in some cases
- By taking a course in these methods, you have shifted your production function, but this does not mean that you escape the trade-off

My suggestions:

- Identification assumptions should be transparent; methods appropriate; and conclusions cautious
- However we shouldn't give up on important policy issues for the sake of perfect identification
- In other words: **make sure you are on the frontier!**

What next?

Update your CV

You could all now legitimately add something like this to your CV:

> *Advanced quantitative methods training, including experience with: analysing randomized experiments; observational causal inference methods (e.g., regression, matching); quasi-experimental methods (e.g., instrumental variables and regression discontinuity designs); and with panel-data methods (e.g., difference-in-differences, fixed-effect regressions, and synthetic control methods).*

Methods we did not (directly) cover

- Several of the methods that we discussed can be combined with each other
- This can strengthen your identification strategy, but beware that this may involve slightly different assumptions!
- For instance:
  - Instrumented (or Fuzzy) Difference-in-Differences (de Chaisemartin & D'Haultfoeuille, 2017, Ye et al, 2020)
  - Difference-in-Discontinuities (e.g. Grembi et al, 2016)
  - Matching and panel data/DiD (see the work and `R` packages by Imai, Kim and Wang)

Further topics in causal inference

1. Graphical models
   - An alternative approach to thinking about causality
   - Complementary to, though distinct from, the potential outcomes framework
   - Foundational book on Structural Causal Models: *Causality* by **Judea Pearl**
   - See the **Morgan and Winship** book for an overview
2. Causal mechanisms
   - Theories that do not only tell us that D should affect Y, but also **why** this causal relationship occurs
   - Methods to assess mechanisms mostly focus on indirect tests for observable implications of different mechanisms
   - See **Imai et. al., 2011**

3. Experimental design
    - Huge literature on experimental design, for field, lab, and survey experiments
    - Topics include: treatment heterogeneity; conjoint experiments; clustered experiments; mediation
    - See the **Gerber and Green** book for more on this
    - Recent (excellent) book edited by **Druckman and Green**: Advances in Experimental Political Science
4. Designing research designs
    - Recent developments to advance systematic and integrated approaches for observational and experimental data, descriptive and causal inferences, qualitative and quantitative (or mixed method) research
    - Combines ideas from the structural causal modeling and potential outcomes traditions
    - See this paper by **Blair, et al** and the online book Research Design: Declaration, Diagnosis, Redesign

The unasked question of measurement

- Throughout this course we have largely accepted that we have, or can easily find, measures for D and Y (and Z and X).
- Quantitative measurement, however, is a difficult problem, particularly in the social sciences where many of our concepts are essentially qualitative.
- A course on measurement might focus on:
  - **Quantitative Text Analysis** $\rightarrow$ extract meaning from collections of texts
  - **Factor analysis/Ideal point models** $\rightarrow$ combining variables to form measures of 'latent' concepts
  - **Clustering** $\rightarrow$ inferring 'classes' of observations from data
  - **Missing data** $\rightarrow$ methods for imputing or otherwise handling missing data

Recommendations

- Books
  - Judea Pearl and Dana MacKenzie, *The Book of Why*
  - Paul Rosenbaum, *Observation and Experiment*
  - Matthew J Salganik, *Bit By Bit: Social Research in the Digital Age*
  - Scott Cunningham, *Causal Inference: The Mixtape*
- Podcasts
  - Causal Inference Podcast
  - Scope Conditions
  - Probable Causation
- Blogs
  - Andrew Gelman blog
  - DataColada

Conclusion

Last Seminars

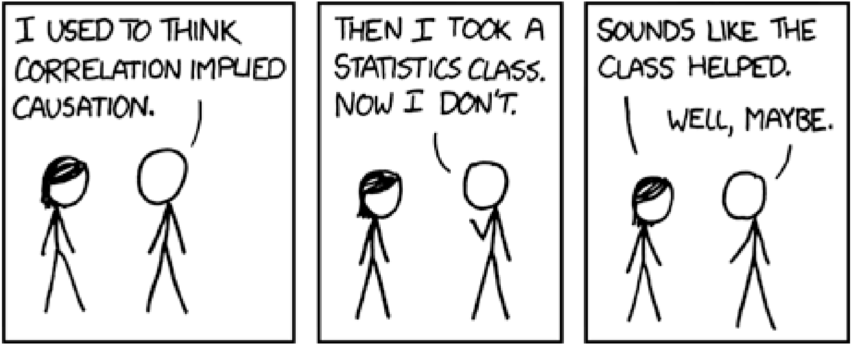
- No seminar exercises
- Make note and bring with you any and all questions about past *seminars* tasks to ask your seminar leader and discuss in class
- Also make note and bring with you to the seminar any questions about **lecture** content you may have
- These last seminars are also an opportunity for you to discuss your research proposal idea - please *use this opportunity*
  - Note that some may have already thought a lot about theirs, but some less - either is perfectly fine!
  - Just bring notes on what you have at this stage

To guide your discussion

All of your papers should have one thing in common – the **structure** of your research question: *Does D have a **causal** effect on Y?*

You then want to think about:

1. What is your treatment variable?
2. What is your outcome variable?
3. **Why** might we expect your treatment to have (or not) a **causal** effect on your outcome?
4. What data do you need and what can you find?
   - What is the unit of observation?
   - What is the treatment group?
   - What is the control group?
5. What design does your data allow?
   - What assumptions need to hold for you to be able to recover a credible causal estimate?
   - What threats to inference might remain?


I USED TO THINK CORRELATION IMPLIED CAUSATION.
THEN I TOOK A STATISTICS CLASS. NOW I DON'T.
SOUNDS LIKE THE CLASS HELPED.
WELL, MAYBE.

Thanks for participating all term, and have a great break!