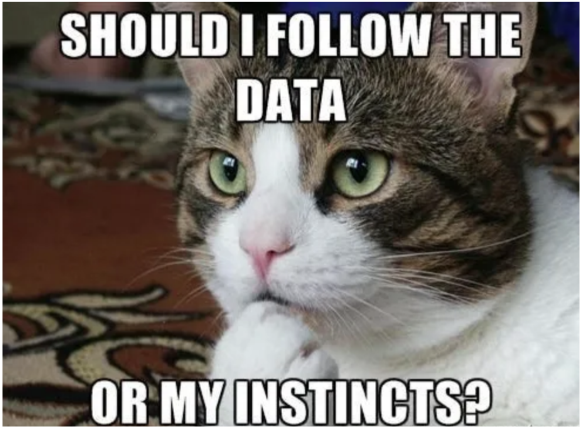
Causal Inference Workshop

Week 1 - Modeling Fundamentals

Causal Inference Workshop

January 19, 2024

Welcome to the Causal Inference Workshop!



Welcome to the Causal Inference Workshop!

- Extremely heavily based on Claire's (2022) and Anna's (2024) awesome courses
- Updated with a section on field experiments
- Slight modification for the DID section
- Slides and codes are uploaded to the Github repository
- Additional reference materials will be added to the `readings` folder.

Workshop outline

A. Causal inference fundamentals
  - Modeling assumptions matter too
  - Conceptual framework (potential outcomes framework)

B. Design stage: common identification strategies
  - IV + RDD [coding]
  - DiD, DiDiD, Event Studies, New TWFE Lit [coding]
  - Synthetic Control / Synthetic DiD [coding]

C. Analysis stage: strengthening inferences
  - Limitations of identification strategies, pre-estimation steps
  - Estimation [controls] and post-estimation steps [supporting assumptions]

D. Other topics in causal inference and sustainable development
  - Inference (randomization inference, bootstrapping)
  - Field experiments
  - Weather data regressions, remote sensing data [coding]

Workshop outline

A. Causal inference fundamentals
   - Modeling assumptions matter too
   - Conceptual framework (potential outcomes framework)

B. Design stage: common identification strategies
   - IV + RDD [coding]
   - DiD, DiDiD, Event Studies, New TWFE Lit [coding]
   - Synthetic Control / Synthetic DiD [coding]

C. Analysis stage: strengthening inferences
   - Limitations of identification strategies, pre-estimation steps
   - Estimation [controls] and post-estimation steps [supporting assumptions]

D. Other topics in causal inference and sustainable development
   - Inference (randomization inference, bootstrapping)
   - Field experiments
   - Weather data regressions, remote sensing data [coding]

Workshop outline

A. Causal inference fundamentals
   - Modeling assumptions matter too
   - Conceptual framework (potential outcomes framework)

B. Design stage: common identification strategies
   - IV + RDD [coding]
   - DiD, DiDiD, Event Studies, New TWFE Lit [coding]
   - Synthetic Control / Synthetic DiD [coding]

C. Analysis stage: strengthening inferences
   - Limitations of identification strategies, pre-estimation steps
   - Estimation [controls] and post-estimation steps [supporting assumptions]

D. Other topics in causal inference and sustainable development
   - Inference (randomization inference, bootstrapping)
   - Field experiments
   - Weather data regressions, remote sensing data [coding]

Workshop outline

A. Causal inference fundamentals
   - Modeling assumptions matter too
   - Conceptual framework (potential outcomes framework)

B. Design stage: common identification strategies
   - IV + RDD [coding]
   - DiD, DiDiD, Event Studies, New TWFE Lit [coding]
   - Synthetic Control / Synthetic DiD [coding]

C. Analysis stage: strengthening inferences
   - Limitations of identification strategies, pre-estimation steps
   - Estimation [controls] and post-estimation steps [supporting assumptions]

D. Other topics in causal inference and sustainable development
   - Inference (randomization inference, bootstrapping)
   - Field experiments
   - Weather data regressions, remote sensing data [coding]

Workshop Logistics

- 9am - 10am Fridays (before colloquium)
  - Office hours on demand 8am - 9 am, if you want to work through code / discuss anything
- If you’re taking the course for credit (pass/fail) and “research tools” requirement, attendance is expected every week
  - If you’re not taking it for credit, but want to get workshop-related emails, let me know!
- One assignment to turn it at the end (run and modify one of the coding exercises)

Outline

Workshop outline

Causal inference

Modeling assumptions of the Classical Linear Regression Model

Departures from usual assumptions

Summary

Mini Coding Exercise

Causal inference, identification, and modeling assumptions

What is causal inference?

- Process by which we use data to make claims about causal relationships
- *Potential outcomes* [framework]
  - Causal effect is difference between two potential outcomes
- *Identification* [application/implementation]
  - Identifying assumptions needed for a statistical estimate to have causal interpretation
  - Removing selection bias
  - E.g., RD, IV, ...
- *Estimation* [application/implementation]

RCT: Identification, Estimand, and Estimator

- **Estimand**
  - Average Treatment Effect (ATE):
  $$\tau = \mathbb{E}[Y(1) - Y(0)]$$
  - A population-level causal effect defined using potential outcomes.
- **Identification Assumptions (assumption about DGP)**
  - Random assignment:
  $$(Y(1), Y(0)) \perp D$$
  - Stable Unit Treatment Value Assumption (SUTVA).
- **Identifiable (under these assumptions)**
  - The estimand can be written using observable quantities:
  $$\mathbb{E}[Y(1)] - \mathbb{E}[Y(0)] = \mathbb{E}[Y \mid D = 1] - \mathbb{E}[Y \mid D = 0]$$
- **Estimator**
  - Difference in sample means:
  $$\hat{\tau} = \bar{Y}_{D=1} - \bar{Y}_{D=0}$$

2 by 2 DiD: Identification, Estimand, and Estimator

- **Estimand**
  - Average Treatment Effect on the Treated (ATT):

$$\tau_{\text{ATT}} = \mathbb{E}[Y(1)_{t=1} - Y(0)_{t=1} \mid D = 1]$$

- **Identification Assumptions (assumption about DGP)**
  - Parallel trends:

$$\mathbb{E}[Y(0)_{t=1} - Y(0)_{t=0} \mid D = 1] = \mathbb{E}[Y(0)_{t=1} - Y(0)_{t=0} \mid D = 0]$$

  - No anticipation assumption
- **Identifiable (under these assumptions)**
  - The counterfactual trend for treated units is recovered from controls:

$$\tau_{\text{ATT}} = (\mathbb{E}[Y_{1,1}] - \mathbb{E}[Y_{1,0}]) - (\mathbb{E}[Y_{0,1}] - \mathbb{E}[Y_{0,0}])$$

- **Estimator**
  - DiD estimator:

$$\hat{\tau}_{\text{DiD}} = (\bar{Y}_{1,1} - \bar{Y}_{1,0}) - (\bar{Y}_{0,1} - \bar{Y}_{0,0})$$

  - Equivalent to the coefficient on *Treatment* $\times$ *Post* in a 2×2 DiD regression.

Outline

Workshop outline

Causal inference

Modeling assumptions of the Classical Linear Regression Model

Departures from usual assumptions

Summary

Mini Coding Exercise

Today: Modeling assumptions

- In econometrics, focus of causal inference is usually on *identification*
- But we are *never* in a perfect identification setup in observational studies
  - Model specification has influence
  - Modeling assumptions must hold (*especially* because we make causal claims based on statistical significant (e.g., p-values))
  - May want more than just unbiasedness (e.g., precision, external validity)

→ Need to think about modeling assumptions and estimator properties

The Classical Linear Regression Model

- The Classical Linear Regression Model (CLRM):

$$y|X \sim \mathcal{F}(X\beta, \sigma^2 I)$$

- Represent specific distribution induced by the data generating process (DGP)
- Regression models focus on *conditional distribution* $y|X$
- We can therefore write them as *conditional mean* $+/\times$ *error*

$$y = \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k + e, \qquad e \overset{\text{iid}}{\sim} \mathcal{F}(0, \sigma^2 I)$$

$$y = \mathbb{E}[y|X] + e, \qquad \mathbb{E}[y|X] = \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k, \qquad e \overset{\text{iid}}{\sim} \mathcal{F}(0, \sigma^2 I)$$

RCT: Regression Estimator for ATE

- Write observed outcome in terms of potential Outcomes:

$$Y = XY_1 + (1 - X)Y_0$$

- Decompose potential outcomes into mean and deviation:

$$\begin{aligned} Y_1 &= \mathbb{E}[Y_1] + \epsilon_1, \qquad \mathbb{E}[\epsilon_1] = 0 \\ Y_0 &= \mathbb{E}[Y_0] + \epsilon_0, \qquad \mathbb{E}[\epsilon_0] = 0 \end{aligned}$$

- Substituting these back into the observed outcome equation:

$$\begin{aligned} Y &= X(\mathbb{E}[Y_1] + \epsilon_1) + (1 - X)(\mathbb{E}[Y_0] + \epsilon_0) \\ Y &= \mathbb{E}[Y_0] + X(\mathbb{E}[Y_1] - \mathbb{E}[Y_0]) + \underbrace{\epsilon_0 + X(\epsilon_1 - \epsilon_0)}_{e} \end{aligned}$$

- Identification assumption gives externality:

$$Y = \alpha + \beta X + e, \qquad \text{where } \beta = \mathbb{E}[Y_1 - Y_0] = \text{ATE}$$

$$\mathbb{E}[e|X = 0] = \mathbb{E}[\epsilon_0] = 0 \text{ and } \mathbb{E}[e|X = 1] = \mathbb{E}[\epsilon_1] = 0; \text{ thus } \mathbb{E}[e|X] = 0.$$

Assumptions of the CLRM

- A set of assumptions that describe a DGP (CLRM)
- The assumptions (aka Gauss-Markov assumptions), by decreasing order of importance:

Notation:	System of n equations	Matrix
Model:	$y_i = x_i'\beta + e_i \;\; (i = 1, ..., n)$	$y = X\beta + e$
Assumptions		

Assumptions of the CLRM

- A set of assumptions that describe a DGP (CLRM)
- The assumptions (aka Gauss-Markov assumptions), by decreasing order of importance:

Notation:	System of n equations	Matrix
Model:	$y_i = x_i'\beta + e_i \ \ (i = 1, ..., n)$	$y = X\beta + e$
Assumptions		
A1. linearity	model is linear in β	model is linear in β

Assumptions of the CLRM

- A set of assumptions that describe a DGP (CLRM)
- The assumptions (aka Gauss-Markov assumptions), by decreasing order of importance:

Notation:	System of n equations	Matrix
Model:	$y_i = x_i'\beta + e_i \;\; (i = 1, ..., n)$	$y = X\beta + e$
Assumptions		
A1. linearity	model is linear in β	model is linear in β
A2. rank condition	$\rho x_k, x_l \approx 1$	$X_{N\times K}$ has rank K

Assumptions of the CLRM

- A set of assumptions that describe a DGP (CLRM)
- The assumptions (aka Gauss-Markov assumptions), by decreasing order of importance:

Notation:	System of n equations	Matrix
Model:	$y_i = x_i'\beta + e_i \;\; (i = 1, ..., n)$	$y = X\beta + e$
Assumptions		
A1. linearity	model is linear in β	model is linear in β
A2. rank condition	$\rho x_k, x_l \approx 1$	$X_{N\times K}$ has rank K
A3. exogeneity	$\mathbb{E}[e_i \vert X] = 0$	$\mathbb{E}[e_i \vert X] = 0_{N\times 1}$

Assumptions of the CLRM

- A set of assumptions that describe a DGP (CLRM)
- The assumptions (aka Gauss-Markov assumptions), by decreasing order of importance:

Notation:	System of n equations	Matrix
Model:	$y_i = x_i'\beta + e_i \quad (i = 1, ..., n)$	$y = X\beta + e$
Assumptions		
A1. linearity	model is linear in β	model is linear in β
A2. rank condition	$\rho x_k, x_l \approx 1$	$X_{N\times K}$ has rank K
A3. exogeneity	$\mathbb{E}[e_i \| X] = 0$	$\mathbb{E}[e_i \| X] = 0_{N\times 1}$
A4. spherical errors	$e_i \| X \overset{\text{iid}}{\sim} (0, \sigma^2)$	$\mathbb{V}[e \| X] = \sigma^2 I_N$
-independent errors	$cov[e_i, e_j \| X] = 0$	
-homoskedastic errors	$\mathbb{V}[e_i \| X] = \sigma^2 \quad \cancel{\sigma_i^2}$	

Assumptions of the CLRM

- A set of assumptions that describe a DGP (CLRM)
- The assumptions (aka Gauss-Markov assumptions), by decreasing order of importance:

Notation:	System of n equations	Matrix
Model:	$y_i = x_i'\beta + e_i \quad (i = 1, ..., n)$	$y = X\beta + e$
Assumptions		
A1. linearity	model is linear in β	model is linear in β
A2. rank condition	$\rho x_k, x_l \approx 1$	$X_{N\times K}$ has rank K
A3. exogeneity	$\mathbb{E}[e_i \vert X] = 0$	$\mathbb{E}[e_i \vert X] = 0_{N\times 1}$
A4. spherical errors	$e_i \vert X \overset{\text{iid}}{\sim} (0, \sigma^2)$	$\mathbb{V}[e \vert X] = \sigma^2 I_N$
-independent errors	$cov[e_i, e_j \vert X] = 0$	
-homoskedastic errors	$\mathbb{V}[e_i \vert X] = \sigma^2 \quad \cancel{\sigma_i^2}$	
A5. normal errors	$e_i \vert X \sim \mathcal{N}(0, \sigma^2)$	$e_i \vert X \sim \mathcal{N}(0_{N\times 1}, \sigma^2 I_N)$

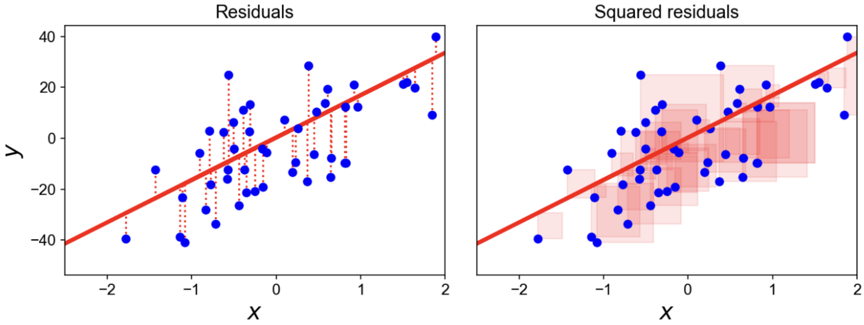
Properties of $\hat{\beta}_{OLS}$

$\rightarrow$ if these assumptions are met, we can make the powerful claim that $\hat{\beta}_{OLS}$ is the best linear model we can use

Properties of $\hat{\beta}_{OLS}$

$\rightarrow$ if these assumptions are met, we can make the powerful claim that $\hat{\beta}_{OLS}$ is the best linear model we can use

$$\hat{\beta}_{\text{OLS}} = \underset{\beta}{\text{argmin}} \sum_{i=1}^{n} r_i^2 = \sum_{j=1}^{n} (y_i - X_i'\beta)^2 = \ldots = (X'X)^{-1}X'y = \beta_0 + (X'X)^{-1}X'e$$



Properties of $\hat{\beta}_{OLS}$

$\rightarrow$ if these assumptions are met, we can make the powerful claim that $\hat{\beta}_{OLS}$ is the best linear model we can use

$$\hat{\beta}_{\text{OLS}} = \underset{\beta}{\text{argmin}} \sum_{i=1}^{n} r_i^2 = \sum_{j=1}^{n} (y_i - X_i'\beta)^2 = \ldots = (X'X)^{-1}X'y = \beta_0 + (X'X)^{-1}X'e$$

Based on the assumptions, $\hat{\beta}_{OLS}$ has the following:

- finite sample properties:

  (A1-A3) $\rightarrow$ $\hat{\beta}_{OLS}$ **unbiased**
  (A4) $\rightarrow$ $\hat{\beta}_{OLS}$ **efficient** (Best **Linear** Unbiased Estimator (BLUE))
  (A5) $\rightarrow$ $\hat{\beta}_{OLS}$ **efficient** ($\hat{\beta}_{MLE}$; Best Unbiased Estimator; normal)

- asymptotic properties:
  - $\hat{\beta}_{OLS}$ is consistent and is asymptomatically unbiased, normally distributed, and efficient

(Assumption 5) Normal errors

- Normal error assumption required for making inferences (computing confidence intervals or p-values)

- Without (A5), t and F tests are invalid
  - One-sample t-test for β ($\beta = 0$) assumes sampling distribution of $\hat{\beta}$ is normal (which means errors have to be normal)

- When (A5) is violated, appeal to asymptotics:
  - When n is large enough, Laws of Large Numbers (LLNs) and Central Limit Theorem (CLT) say that asymptotic sampling distribution of $\hat{\beta}$ is normal
  - If n is large, t and F tests are robust to departures from normality
  - In the case of highly non-normal error, may want to consider alternative (e.g., bootstrap)

Outline

Workshop outline

Causal inference

Modeling assumptions of the Classical Linear Regression Model

Departures from usual assumptions

Summary

Mini Coding Exercise

(A4) Non-spherical errors

Assuming (A1)-(A3), the asymptotic distribution of $\hat{\beta}_{OLS}$ is:

$$\hat{\beta}_{OLS} \overset{a}{\sim} \mathcal{N}\left(\beta_0, (X'X)^{-1} X' \Sigma X (X'X)^{-1'}\right)$$

$\rightarrow$ need a consistent estimate of the asymptotic vcov matrix in order to do sampling-based statistical inference $\rightarrow$ need Σ (the vcov matrix of the error term)

(A4) Non-spherical errors

Assuming (A1)-(A3), the asymptotic distribution of $\hat{\beta}_{OLS}$ is:

$$\hat{\beta}_{OLS} \overset{a}{\sim} \mathcal{N}\left(\beta_0, (X'X)^{-1}X'\Sigma X(X'X)^{-1'}\right)$$

$\rightarrow$ need a consistent estimate of the asymptotic vcov matrix in order to do sampling-based statistical inference $\rightarrow$ need Σ (the vcov matrix of the error term)

- Spherical e: $\Sigma = \sigma^2 I$, so we can simply consistently estimate the population variance σ^2 by the unbiased sample variance
- Heteroskedastic e: compute White SEs
- Autocorrelated e: compute Newey-West/Conley SEs if correlated in time/space...
- Clustered e: compute block-diagonal matrix using residuals

Notes on sandwich estimators

- If errors are autocorrelated in any way, it means your model is not capturing some feature of the DGP
  - Adjust for it after fitting the model (e.g., cluster SEs if autocorrelated by group)
  - Or incorporate into the structure of the model (e.g., multilevel structure)
- Sandwich estimators rely on asymptotics
  - Don't cluster SEs if you have too few clusters!
- These SEs only make sense for the linear regression model

Limited outcome models

A linear regression isn't appropriate when there is a *limited* y

- Binary: $y \in 0, 1$
- Count: $y \in 0, 1, 2, 3, \ldots$
- Censored

Limited outcome models

A linear regression isn't appropriate when there is a *limited y*

- Binary: $y \in 0, 1 \rightarrow$ probit, logit, ...
- Count: $y \in 0, 1, 2, 3, \ldots \rightarrow$ Poisson, negative binomial,...
- Censored $\rightarrow$ censored regression models

Use **generalized** linear models (GLMs) - flexible generalization of OLS:

$$g(\mathbb{E}[y|X]) = \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k, \quad e \overset{\text{iid}}{\sim} \mathcal{F}(0, \ldots)$$

- Invertible link function $g()$, which relates:
- $\mathbb{E}[y|X]$ to linear predictor vector $X\beta$
- Assume some data distribution $\mathcal{F}()$

Outline

Workshop outline

Causal inference

Modeling assumptions of the Classical Linear Regression Model

Departures from usual assumptions

Summary

Mini Coding Exercise

The Bottom-Line

In practice, causal inference in observational studies means, *both*

- A good design (identifying assumptions) $\rightarrow$ unbiasedness. Make causal effects recoverable from observed data.
- A model (modeling assumptions). Determines how we estimate and do inference for identified effects (functional form, SEs, efficiency)

You shouldn't overlook the modeling assumptions!

Outline

Workshop outline

Causal inference

Modeling assumptions of the Classical Linear Regression Model

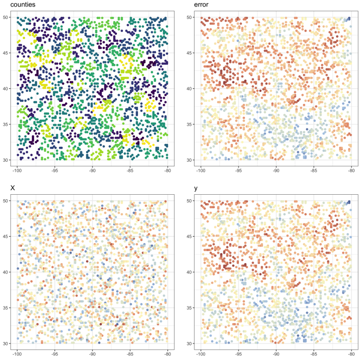
Departures from usual assumptions

Summary

Mini Coding Exercise

Mini coding exercise

- bit.ly/causal2026
- Create fake data with spatial autocorrelation
- Test clustered and Conley SEs
- You can play around with:
  - Parameters of DGP
  - Strength and type of spatial correlation
  - Conley SEs distance parameter
  - Clustered SEs (number of clusters, etc.)
  - Omitted variable bias
  - R vs. Stata functions



Questions? Comments?

Thank you!

References

Heavily based on Claire Palandri's 2022 version and Anna Papp's 2024 version of the Causal Inference Workshop.