Yale

Causal Inference for IR and IPE with Substantive Applications

Carlos Felipe Balcazar

MacMillan Center

January, 2024

EFFECTIVE REGRESSION WEIGHTS

Constant treatment effects? Problems with ATE

$$E(Y_i|D_i = 1, X_i) - E(Y_i|D_i = 0, X_i) = \tau?$$

- Treatment and assignment not the same within cells: $\tau_i \not\perp D_i|X_i$.

Constant treatment effects? Problems with ATE

$$E(Y_i|D_i = 1, X_i) - E(Y_i|D_i = 0, X_i) = \tau?$$

- Treatment and assignment not the same within cells: $\tau_i \not\perp D_i|X_i$.
- Observations are not effectively iid (sampling + design).

Constant treatment effects? Problems with ATE

$$E(Y_i|D_i = 1, X_i) - E(Y_i|D_i = 0, X_i) = \tau?$$

- Treatment and assignment not the same within cells: $\tau_i \not\perp D_i|X_i$.
- Observations are not effectively iid (sampling + design).
- Sampling weights would be still insufficient. Why?

Constant treatment effects? Problems with ATE

$$E(Y_i|D_i = 1, X_i) - E(Y_i|D_i = 0, X_i) = \tau?$$

- Treatment and assignment not the same within cells: $\tau_i \not\perp D_i|X_i$.
- Observations are not effectively iid (sampling + design).
- Sampling weights would be still insufficient. Why?
- Would IPW address this issue? Which weights?!
  - Effect of the treatment likely not the same within cells.

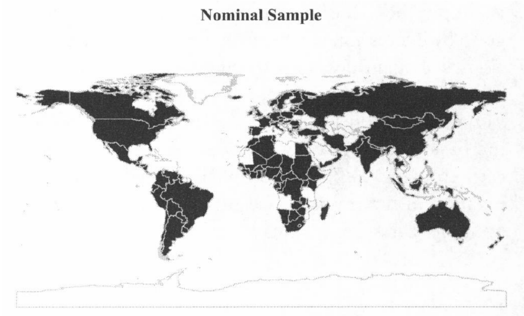
Yale

Constant treatment effects? Problems with ATE

$$E(Y_i|D_i = 1, X_i) - E(Y_i|D_i = 0, X_i) = \tau?$$

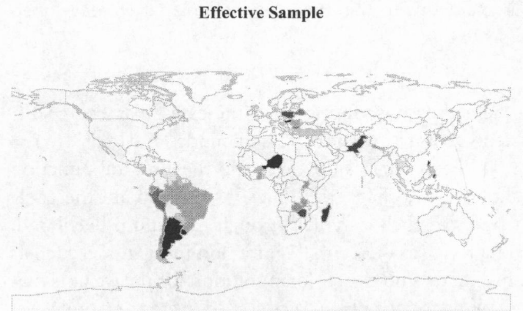
- Treatment and assignment not the same within cells: $\tau_i \not\perp D_i|X_i$.
- Observations are not effectively iid (sampling + design).
- Sampling weights would be still insufficient. Why?
- Would IPW address this issue? Which weights?!
  - Effect of the treatment likely not the same within cells.
- Post-stratification weights with IPW?

Effective regression weights



- Effective regression weights differ from actual ones! Why?

Effective regression weights



- Effective regression weights differ from actual ones! Why?

Effective regression weights

Note: Regression weights for the $\Delta Pr_a \times Pl_{t-4,a}$ variable from Model 3 of Table 3 in Caselli and Tesei (2016); computed following the procedure outlined in Aronow and Samii (2016). Darker shading denotes higher weight; white denotes missing values.

- Effective regression weights differ from actual ones! Why?

Effective regression weights

$$\hat{\beta} \xrightarrow{p} \frac{\mathrm{E}[w_i \tau_i]}{\mathrm{E}[w_i]}, \text{ where } w_i = (D_i - \mathrm{E}[D_i | X_i])^2. \quad (7)$$

- Effective regression weights differ from actual ones! Why?
- Average effects driven by few units; not representative! Why?

Effective regression weights

$$\hat{\beta} \xrightarrow{p} \frac{\mathrm{E}[w_i \tau_i]}{\mathrm{E}[w_i]}, \text{ where } w_i = (D_i - \mathrm{E}[D_i | X_i])^2. \quad (7)$$

$$\mathrm{E}[w_i | X_i] = \mathrm{Var}[D_i | X_i].$$

- Effective regression weights differ from actual ones! Why?
- Average effects driven by few units; not representative! Why?

Effective regression weights

$$\hat{\beta} \xrightarrow{p} \frac{E[w_i \tau_i]}{E[w_i]}, \text{ where } w_i = (D_i - E[D_i | X_i])^2. \quad (7)$$

$$E[w_i | X_i] = \mathrm{Var}[D_i | X_i].$$

- Effective regression weights differ from actual ones! Why?
- Average effects driven by few units; not representative! Why?
- Rejecting the null doesn't necessarily mean no effect! Why?

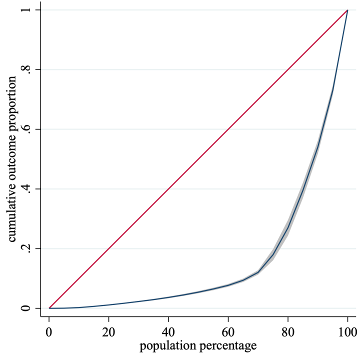
Yale

Diagnostics and considerations

$$\hat{\beta} \xrightarrow{p} \frac{\mathrm{E}[w_i \tau_i]}{\mathrm{E}[w_i]}, \text{ where } w_i = (D_i - \mathrm{E}[D_i | X_i])^2. \quad (7)$$

- Conditional sums or simple cloropeth maps.

Diagnostics and considerations



- Conditional sums or simple cloropeth maps.
- Distributional statistics: Gini; entropy measures.

Diagnostics and considerations

All obs	p90/p10	p90/p50	p10/p50	p75/p25
	47.632	17.079	0.359	12.747

Generalized Entropy indices GE(a), where a = income difference sensitivity parameter, and Gini coefficient

All obs	GE(-1)	GE(0)	GE(1)	GE(2)	Gini
	16.36012	1.20614	0.86768	1.11303	0.68322

Atkinson indices, A(e), where e > 0 is the inequality aversion parameter

All obs	A(0.5)	A(1)	A(2)
	0.40975	0.70065	0.97034

- Conditional sums or simple cloropeth maps.
- Distributional statistics: Gini; entropy measures.

Diagnostics and considerations

$$\hat{\beta} \xrightarrow{p} \frac{\mathrm{E}[w_i \tau_i]}{\mathrm{E}[w_i]}, \text{ where } w_i = (D_i - \mathrm{E}[D_i | X_i])^2. \quad (7)$$

- Conditional sums or simple cloropeth maps.
- Distributional statistics: Gini; entropy measures.
- Does IPW work in this case? IF so, how?

CLUSTERING

Why having large N may not really mean much?

- Recall heteroskedasticity - solution robust SE.

Why having large N may not really mean much?

- Recall heteroskedasticity - solution robust SE.
- Units may react similarly to treatment within groups or clusters.
  - Intra-cluster correlation $+$ (-) $\Rightarrow$ SE $\Uparrow$ ($\Downarrow$). Why?
  - In FE clusters matter for additional DF corrections (nesting).

Why having large N may not really mean much?

- Recall heteroskedasticity - solution robust SE.
- Units may react similarly to treatment within groups or clusters.
  - Intra-cluster correlation + (-) $\Rightarrow$ SE $\Uparrow$ ($\Downarrow$). Why?
  - In FE clusters matter for additional DF corrections (nesting).
- What matters is the number of clusters! (Effective observations).

Why having large N may not really mean much?

- Recall heteroskedasticity - solution robust SE.
- Units may react similarly to treatment within groups or clusters.
  - Intra-cluster correlation $+$ (-) $\Rightarrow$ SE $\Uparrow$ ($\Downarrow$). Why?
  - In FE clusters matter for additional DF corrections (nesting).
- What matters is the number of clusters! (Effective observations).
  - 5000 data points can effectively be 50 observations! Why?
  - <1 dimension, adjustments depend on $\min\{G_1, G_2, ...\}$.

Yale

Design-based uncertainty

$$\widehat{\theta} = \frac{1}{N_1} \sum_{i=1}^{n} R_i X_i Y_i - \frac{1}{N_0} \sum_{i=1}^{n} R_i (1 - X_i) Y_i.$$

- Imagine your data comes from a (quasi-)experimental design.
  - Sampling matters (uncertainty due to sampling frame).
    - Cluster selection and selection of observations within cluster.
  - EHW are overconservative because ignores design uncertainty.
    - Limiting condition when no heterogeneity in *ATE* across clusters.

Design-based uncertainty

$$V^{\text{design}}(N_1, N_0, n_1, n_0) = E\left[\text{var}(\widehat{\theta}|\mathbf{R}, N_1, N_0)\middle|N_1, N_0\right] = \frac{S_1^2}{N_1} + \frac{S_0^2}{N_0} - \frac{S_\theta^2}{N_0 + N_1}.$$

- Imagine your data comes from a (quasi-)experimental design.
  - Sampling matters (uncertainty due to sampling frame).
    - Cluster selection and selection of observations within cluster.
  - EHW are overconservative because ignores design uncertainty.
    - Limiting condition when no heterogeneity in *ATE* across clusters.

Design-based uncertainty

$$\widehat{\tau}_k = \frac{1}{N_{k,1} \vee 1} \sum_{i=1}^{n_k} R_{k,i} W_{k,i} Y_{k,i} - \frac{1}{N_{k,0} \vee 1} \sum_{i=1}^{n_k} R_{k,i} (1 - W_{k,i}) Y_{k,i},$$

- Imagine your data comes from a (quasi-)experimental design.
  - Sampling matters (uncertainty due to sampling frame).
    - Cluster selection and selection of observations within cluster.
  - EHW are overconservative because ignores design uncertainty.
    - Limiting condition when no heterogeneity in *ATE* across clusters.

Design-based uncertainty

$$\frac{1}{n_k}\sum_{i=1}^{n_k}\left(\frac{u_{k,i}^2(1)}{\mu_k}+\frac{u_{k,i}^2(0)}{1-\mu_k}\right)-p_k\frac{1}{n_k}\sum_{i=1}^{n_k}\Big(u_{k,i}(1)-u_{k,i}(0)\Big)^2.$$

- Imagine your data comes from a (quasi-)experimental design.
  - Sampling matters (uncertainty due to sampling frame).
    - Cluster selection and selection of observations within cluster.
  - EHW are overconservative because ignores design uncertainty.
    - Limiting condition when no heterogeneity in *ATE* across clusters.

Design-based uncertainty

$$\frac{1}{n_k}\sum_{i=1}^{n_k}\left(\frac{u_{k,i}^2(1)}{\mu_k}+\frac{u_{k,i}^2(0)}{1-\mu_k}\right)-p_k\frac{1}{n_k}\sum_{i=1}^{n_k}\Big(u_{k,i}(1)-u_{k,i}(0)\Big)^2.$$

- Imagine your data comes from a (quasi-)experimental design.
  - Sampling matters (uncertainty due to sampling frame).
    - Cluster selection and selection of observations within cluster.
  - EHW are overconservative because ignores design uncertainty.
    - Limiting condition when no heterogeneity in *ATE* across clusters.

Design-based uncertainty

$$\frac{1}{n_k}\sum_{i=1}^{n_k}\left(\frac{u_{k,i}^2(1)}{\mu_k}+\frac{u_{k,i}^2(0)}{1-\mu_k}\right)-p_k\frac{1}{n_k}\sum_{i=1}^{n_k}\Big(u_{k,i}(1)-u_{k,i}(0)\Big)^2.$$

- Imagine your data comes from a (quasi-)experimental design.
  - Sampling matters (uncertainty due to sampling frame).
    - Cluster selection and selection of observations within cluster.
  - EHW are overconservative because ignores design uncertainty.
    - Limiting condition when no heterogeneity in *ATE* across clusters.
- Sampling increases variance;

Yale

Design-based uncertainty

$$p_k(1-q_k)\frac{1}{n_k}\sum_{m=1}^{m_k} n_{k,m}^2(\tau_{k,m}-\tau_k)^2.$$

- Imagine your data comes from a (quasi-)experimental design.
  - Sampling matters (uncertainty due to sampling frame).
    - Cluster selection and selection of observations within cluster.
  - EHW are overconservative because ignores design uncertainty.
    - Limiting condition when no heterogeneity in *ATE* across clusters.
- Sampling increases variance; design uncertainty increases variance.

Design-based uncertainty

$$
\begin{aligned}
&-p_k\sigma_k^2\frac{1}{n_k}\sum_{i=1}^{n_k}\left(\frac{u_{k,i}(1)}{\mu_k}+\frac{u_{k,i}(0)}{1-\mu_k}\right)^2 \\
&+p_k\sigma_k^2\frac{1}{n_k}\sum_{m=1}^{m_k}\left(\sum_{i=1}^{n_k}1\{m_{k,i}=m\}\left(\frac{u_{k,i}(1)}{\mu_k}+\frac{u_{k,i}(0)}{1-\mu_k}\right)\right)^2 .
\end{aligned}
$$

- Imagine your data comes from a (quasi-)experimental design.
  - Sampling matters (uncertainty due to sampling frame).
    - Cluster selection and selection of observations within cluster.
  - EHW are overconservative because ignores design uncertainty.
    - Limiting condition when no heterogeneity in *ATE* across clusters.
- Sampling increases variance;

Design-based uncertainty (II)

$$v_k^{\text{cluster}} - v_k = \left(\frac{p_k n_k}{m_k}\right) q_k \left\{ \frac{1}{m_k} \sum_{m=1}^{m_k} \left(\frac{n_{k,m} m_k}{n_k}\right)^2 (\tau_{k,m} - \tau_k)^2 \right\}.$$

- Clustered standard errors are at least overconservative.
  - Many clusters; heterogeneity in *ATE* across clusters.
- Weights depends of design v. sampling.
  - Resampling weights to account for treatment effect heterogeneity.
  - Abadie et al. corrections if within-cluster heterogeneity assignment.
  - No correction if treatment assignment is at level of cluster ($\lambda = 0$).

Design-based uncertainty (II)

$$\widetilde{V}_k^{\mathrm{CCV}} = \widehat{\lambda}_k \widetilde{V}_k^{\mathrm{cluster}} + (1 - \widehat{\lambda}_k) \widetilde{V}_k^{\mathrm{robust}}$$

- Clustered standard errors are at least overconservative.
  - Many clusters; heterogeneity in *ATE* across clusters.
- Weights depends of design v. sampling.
  - Resampling weights to account for treatment effect heterogeneity.
  - Abadie et al. corrections if within-cluster heterogeneity assignment.
  - No correction if treatment assignment is at level of cluster ($\lambda = 0$).

Design-based uncertainty (II)

$$\widehat{\lambda}_k \approx 1 - q_k \frac{V(A_{k,m}(1 - A_{k,m}))}{E[(A_{k,m}(1 - A_{k,m}))^2]}.$$

- Clustered standard errors are at least overconservative.
  - Many clusters; heterogeneity in *ATE* across clusters.
- Weights depends of design v. sampling.
  - Resampling weights to account for treatment effect heterogeneity.
  - Abadie et al. corrections if within-cluster heterogeneity assignment.
  - No correction if treatment assignment is at level of cluster ($\lambda = 0$).

Yale

Insufficient for dyadic data!

Table 1. Assumed dependence by clustering type.

Dyad	US–UK 1	US–UK 2	US–FR 1	UK–ES 1
US–UK 1	Dependent	Dependent	Independent	Independent
US–UK 2	Dependent	Dependent	Independent	Independent
US–FR 1	Independent	Independent	Dependent	Independent
UK–ES 1	Independent	Independent	Independent	Dependent

(a) Clustering by repeated dyad only

Dyad	US–UK 1	US–UK 2	US–FR 1	UK–ES 1
US–UK 1	Dependent	Dependent	Dependent	Dependent
US–UK 2	Dependent	Dependent	Dependent	Dependent
US–FR 1	Dependent	Dependent	Dependent	Independent
UK–ES 1	Dependent	Dependent	Independent	Dependent

(b) Dyadic clustering

- Dyadic clusters (usually) underestimate standard errors! Why?

Insufficient for dyadic data!

Table 3. Clustering by repeated dyad.

ijt / $i'j't'$	111	112	211	212	221	222
111	$E[u_{111}^2]$	$E[u_{111}u_{112}]$	0	0	0	0
112	$E[u_{112}u_{111}]$	$E[u_{112}^2]$	0	0	0	0
211	0	0	$E[u_{211}^2]$	$E[u_{211}u_{212}]$	0	0
212	0	0	$E[u_{212}u_{211}]$	$E[u_{212}^2]$	0	0
221	0	0	0	0	$E[u_{221}^2]$	$E[u_{221}u_{222}]$
222	0	0	0	0	$E[u_{222}u_{221}]$	$E[u_{222}^2]$

- Dyadic clusters (usually) underestimate standard errors! Why?
- Dyadic clustering takes into account the network-like structure.

Insufficient for dyadic data!

Table 4. Dyadic clustering.

ijt / $i'j't'$	111	112	211	212	221	222
111	$\mathrm{E}[u_{111}^2]$	$\mathrm{E}[u_{111}u_{112}]$	$\mathrm{E}[u_{111}u_{211}]$	$\mathrm{E}[u_{111}u_{212}]$	0	0
112	$\mathrm{E}[u_{112}u_{111}]$	$\mathrm{E}[u_{112}^2]$	$\mathrm{E}[u_{112}u_{211}]$	$\mathrm{E}[u_{112}u_{212}]$	0	0
211	$\mathrm{E}[u_{211}u_{111}]$	$\mathrm{E}[u_{211}u_{112}]$	$\mathrm{E}[u_{211}^2]$	$\mathrm{E}[u_{211}u_{212}]$	$\mathrm{E}[u_{211}u_{221}]$	$\mathrm{E}[u_{211}u_{222}]$
212	$\mathrm{E}[u_{212}u_{111}]$	$\mathrm{E}[u_{212}u_{112}]$	$\mathrm{E}[u_{212}u_{211}]$	$\mathrm{E}[u_{212}^2]$	$\mathrm{E}[u_{212}u_{221}]$	$\mathrm{E}[u_{212}u_{222}]$
221	0	0	$\mathrm{E}[u_{221}u_{211}]$	$\mathrm{E}[u_{221}u_{212}]$	$\mathrm{E}[u_{221}^2]$	$\mathrm{E}[u_{221}u_{222}]$
222	0	0	$\mathrm{E}[u_{222}u_{211}]$	$\mathrm{E}[u_{222}u_{212}]$	$\mathrm{E}[u_{222}u_{221}]$	$\mathrm{E}[u_{222}^2]$

- Dyadic clusters (usually) underestimate standard errors! Why?
- Dyadic clustering takes into account the network-like structure.
  - Units belonging to a network may react similarly to treatment.

Insufficient for dyadic data!

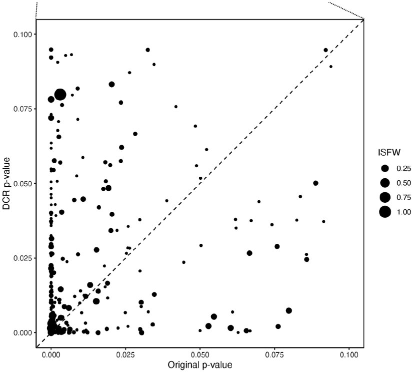
$$\hat{V}_r = \sum_{i=1}^{N} \hat{V}_{C,i} - \hat{V}_D - (N-2)\hat{V}_0,$$

where

$$\hat{V}_{C,i} = (\mathbf{X}'\mathbf{X})^{-1}\hat{\Sigma}_{C,i}(\mathbf{X}'\mathbf{X})^{-1},$$
$$\hat{V}_D = (\mathbf{X}'\mathbf{X})^{-1}\hat{\Sigma}_D(\mathbf{X}'\mathbf{X})^{-1},$$
$$\hat{V}_0 = (\mathbf{X}'\mathbf{X})^{-1}\hat{\Sigma}_0(\mathbf{X}'\mathbf{X})^{-1},$$

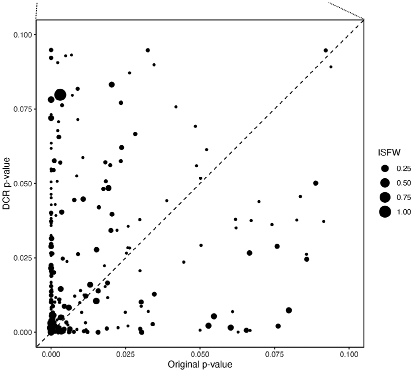
- Dyadic clusters (usually) underestimate standard errors! Why?
- Dyadic clustering takes into account the network-like structure.
  - Units belonging to a network may react similarly to treatment.

Insufficient for dyadic data!



- Type I and type II errors can emerge! Why?

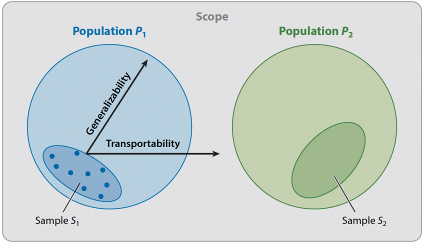
Insufficient for dyadic data!



- Type I and type II errors can emerge! Why?
- Significant changes in conclusions.

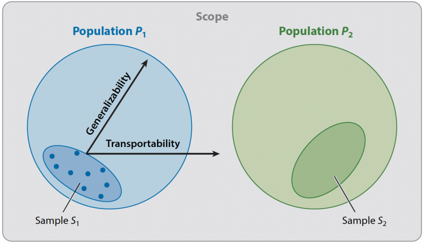
EXTERNAL VALIDITY

Transportability v. generalizability



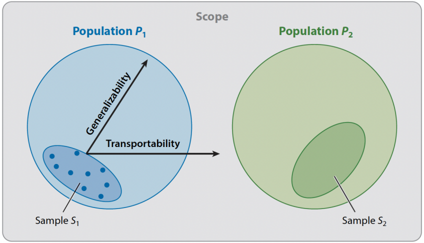
- We may want to use our results to make broad claims.

Transportability v. generalizability



- We may want to use our results to make broad claims.
- Causal inference allow us to obtain internal validity.

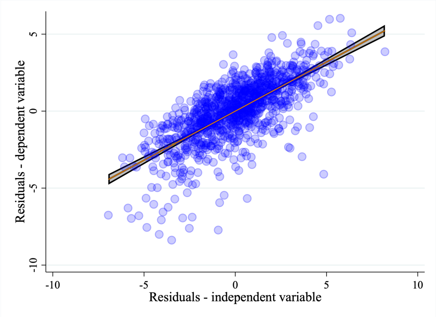
Transportability v. generalizability



- We may want to use our results to make broad claims.
- Causal inference allow us to obtain internal validity.
- To generalize/transport, our theory must be parsimonious!
  - Trade-off: assumptions v. flexibility for empirics.
  - Equilibrium may hold within the defined parameter space (scope).
  - Should not be a problem if clear about scope.

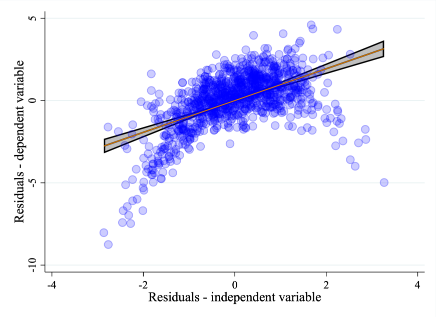
Yale

Theoretical assumptions v. flexibility in empirics



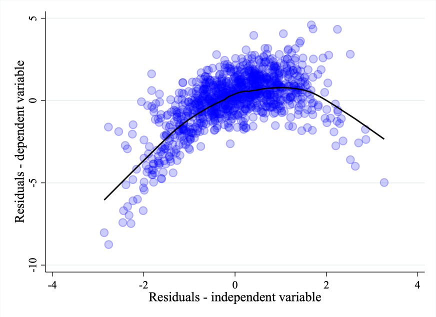
- New/extant theory assumes linear relationship (no controls).

Theoretical assumptions v. flexibility in empirics



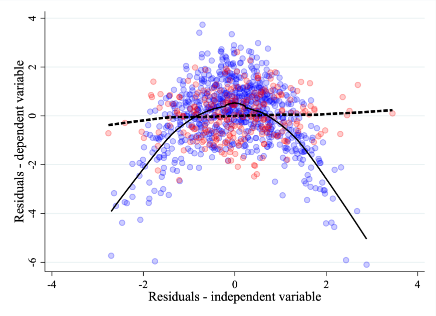
- New/extant theory assumes linear relationship (no controls).
- Relationship is quadratic!;

Theoretical assumptions v. flexibility in empirics



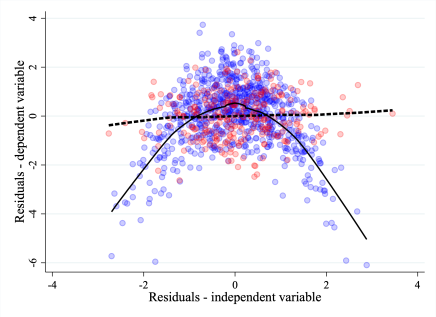
- New/extant theory assumes linear relationship (no controls).
- Relationship is quadratic!;
- Theory is not supported by evidence; theory needs refining.

Theoretical assumptions v. flexibility in empirics



- New/extant theory assumes linear relationship (no controls).
- Relationship is quadratic!; also heterogeneous effects.
- Theory is not supported by evidence; theory needs refining.

Theoretical assumptions v. flexibility in empirics



- New/extant theory assumes linear relationship (no controls).
- Relationship is quadratic!;
- Theory is not supported by evidence; theory needs refining.

THINKING LIKE AN EXPERIMENTALIST

Collecting data and cleaning data is as important

1. Study sample:
   a. What is the population of interest for the study?
   b. What is the sample of analysis?
   c. How does the sample differ from the population, and how will estimated effects generalize to the population?
2. Variables:
   a. What are the main variables of interest in your study?
   b. How will the variables be discretized or otherwise transformed for analysis?
3. Measurement:
   a. How (in)accurately are your variables measured?
   b. Is measurement error random or systematically biased? Is it correlated with other variables?
4. Missingness:
   a. What portion of your observations are missing for each variable?
   b. What is the mechanism giving rise to missingness?
   c. Are the observations missing at random, or is missingness correlated with other variables?
5. Statistical power:
   a. What is the variability of your outcome measure, and of your effect size?
   b. What dependencies exist among your observations?
   c. What is the minimum effect size you will be able to detect?

Recall commensurability

- Think about the limitations of your data beforehand. Why?

Recall commensurability

- Think about the limitations of your data beforehand. Why?
- What will these problems entail for research design. Why?

Recall commensurability

- Think about the limitations of your data beforehand. Why?
- What will these problems entail for research design. Why?
- What will these problems entail for commensurability. Why?

Recall commensurability

- Think about the limitations of your data beforehand. Why?
- What will these problems entail for research design. Why?
- What will these problems entail for commensurability. Why?
- Draft a pre-analysis plan (PAP) to tie your hands.

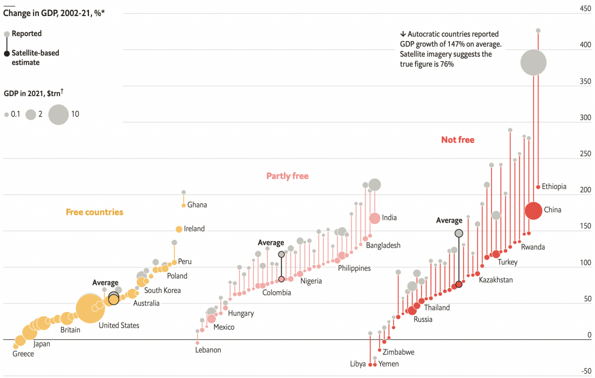
Recall commensurability

- Think about the limitations of your data beforehand. Why?
- What will these problems entail for research design. Why?
- What will these problems entail for commensurability. Why?
- Draft a pre-analysis plan (PAP) to tie your hands.
- Forces to think about theory and about data limitations.
    - Measurement error; self selection.

Recall commensurability

- Think about the limitations of your data beforehand. Why?
- What will these problems entail for research design. Why?
- What will these problems entail for commensurability. Why?
- Draft a pre-analysis plan (PAP) to tie your hands.
- Forces to think about theory and about data limitations.
  - Measurement error; self selection.
- Less latitude for fishing for results and bad inference.

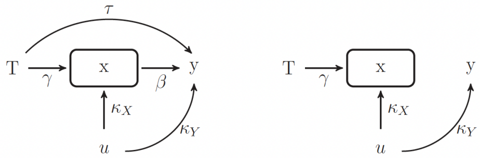
Self-selection spills over to data collection



- Quality/presence of data can be a function of treatment!
  - Deliberate and/or capacity constraints.

Self-selection spills over to data collection

FIGURE 2 Causal Graph When Covariate Is a Posttreatment Variable



(a) Researcher-assumed causal model (b) Actual causal model

- Quality/presence of data can be a function of treatment!
  - Deliberate and/or capacity constraints.
- Missing: controlled directed effect $\Rightarrow$ no generalizable.

Self-selection spills over to data collection

$$= \tau + \kappa_y \Bigg[\underbrace{\Pr(\mathbf{x}=0)}_{\text{Prob. low interest}} \underbrace{[\mathbb{E}(\mathbf{u}|T=1,\mathbf{x}=0) - \mathbb{E}(u|T=0,\mathbf{x}=0)]}_{\text{Imbalance when} x=0}$$

$$+ \underbrace{\Pr(\mathbf{x}=1)}_{\text{Prob. high interest}} \underbrace{[\mathbb{E}(\mathbf{u}|T=1,\mathbf{x}=1) - \mathbb{E}(\mathbf{u}|T=0,\mathbf{x}=1)]}_{\text{Imbalance when} x=1} \Bigg].$$

- Quality/presence of data can be a function of treatment!
  - Deliberate and/or capacity constraints.
- Missing: controlled directed effect $\Rightarrow$ no generalizable.
- Measurement error: bias due to imbalance. Hard to sign!

Self-selection spills over to data collection

$$= \tau + \kappa_y \Bigg[\underbrace{\Pr(\mathbf{x}=0)}_{\text{Prob. low interest}} \underbrace{[\mathbb{E}(\mathbf{u}|T=1,\mathbf{x}=0) - \mathbb{E}(u|T=0,\mathbf{x}=0)]}_{\text{Imbalance when}x=0}$$

$$+ \underbrace{\Pr(\mathbf{x}=1)}_{\text{Prob. high interest}} \underbrace{[\mathbb{E}(\mathbf{u}|T=1,\mathbf{x}=1) - \mathbb{E}(\mathbf{u}|T=0,\mathbf{x}=1)]}_{\text{Imbalance when}x=1} \Bigg].$$

- Quality/presence of data can be a function of treatment!
  - Deliberate and/or capacity constraints.
- Missing: controlled directed effect $\Rightarrow$ no generalizable.
- Measurement error: bias due to imbalance. Hard to sign!

Next class...

International trade, conflict and democratization!