Transfer Learning Between U.S. Presidential Elections

How can we learn from a 2020 ad campaign to inform 2024 ad campaigns?

Xinran Miao, Jiwei Zhao, Hyunseung Kang

arXiv:2411.01100

University of Wisconsin-Madison

New England Statistics Symposium (NESS)
June 2, 2025, New Haven, CI

Motivation: 2020 Digital Ad Campaign [Aggarwal et al. (2023) Nature Human Behavior]

Motivation: 2020 Digital Ad Campaign [Aggarwal et al. (2023) Nature Human Behavior]

2020 Digital Ad Campaign in U.S. Presidential Election: How would online ads against Donald Trump affect voter turnout in five battle-ground states: AZ, MI, NC, PA, and WI?

Motivation: 2020 Digital Ad Campaign [Aggarwal et al. (2023) Nature Human Behavior]

2020 Digital Ad Campaign in U.S. Presidential Election: How would online ads against Donald Trump affect voter turnout in five battle-ground states: AZ, MI, NC, PA, and WI?

- A stratified randomized experiment from Feb. 2020 to Nov. 2020 on nearly 2 million voters
  - Treatment group: an average of 754 ads against Trump by Acronym
  - Control group: no ads from Acronym
  - Outcome: voted in 2020 U.S. election? (**binary**)

Motivation: 2020 Digital Ad Campaign [Aggarwal et al. (2023) Nature Human Behavior]

2020 Digital Ad Campaign in U.S. Presidential Election: How would online ads against Donald Trump affect voter turnout in five battle-ground states: AZ, MI, NC, PA, and WI?

- A stratified randomized experiment from Feb. 2020 to Nov. 2020 on nearly 2 million voters
  - Treatment group: an average of 754 ads against Trump by Acronym
  - Control group: no ads from Acronym
  - Outcome: voted in 2020 U.S. election? (**binary**)
- Average treatment effect (ATE): difference in voter turnout between treated and control groups
- Estimate: $\widehat{\text{ATE}} = -0.06\%$, $\widehat{\text{SE}} = 0.12\%$ (**insignificant**)

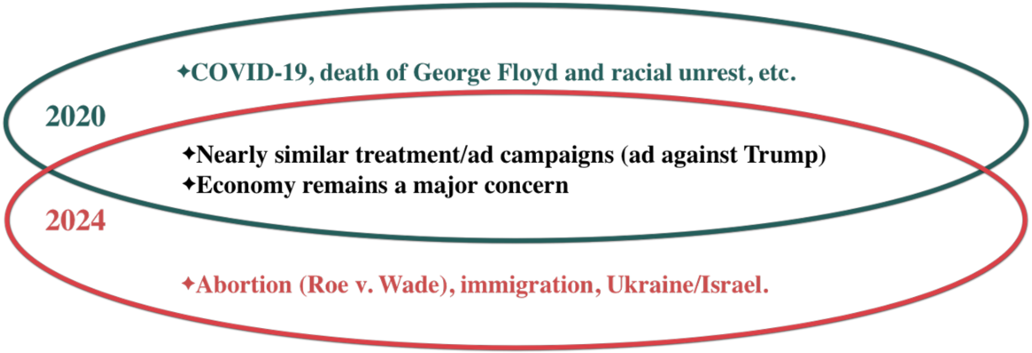
Motivation: 2020 Digital Ad Campaign [Aggarwal et al. (2023) Nature Human Behavior]

2020 Digital Ad Campaign in U.S. Presidential Election: How would online ads against Donald Trump affect voter turnout in five battle-ground states: AZ, MI, NC, PA, and WI?

- A stratified randomized experiment from Feb. 2020 to Nov. 2020 on nearly 2 million voters
  - Treatment group: an average of 754 ads against Trump by Acronym
  - Control group: no ads from Acronym
  - Outcome: voted in 2020 U.S. election? (**binary**)
- Average treatment effect (ATE): difference in voter turnout between treated and control groups
- Estimate: $\widehat{\text{ATE}} = -0.06\%$, $\widehat{\text{SE}} = 0.12\%$ (**insignificant**)

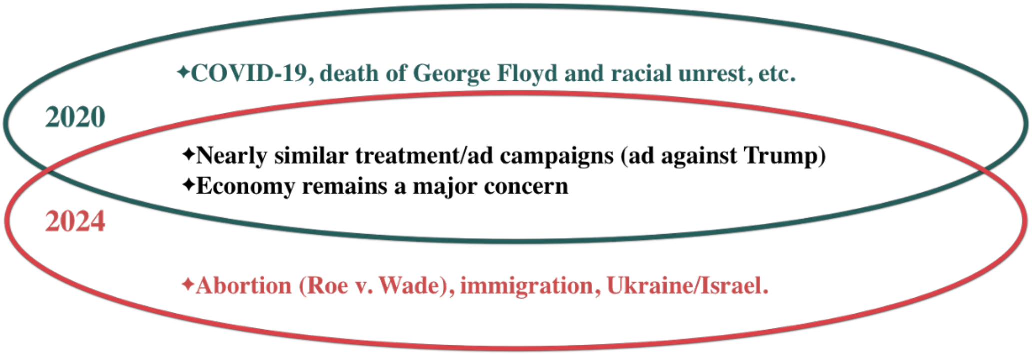
"One reasonable question... is how well ***our findings would generalize... to other electoral contexts****... it could be that the 2020 election was exceptional because of COVID... perhaps digital advertising would have larger effects in more typical settings..."*

This Talk: Would the Negative Ad Against Trump Remain Ineffective in 2024?



Ideal approach: running a randomized experiment in 2024.

This Talk: Would the Negative Ad Against Trump Remain Ineffective in 2024?



Ideal approach: running a randomized experiment in 2024.

However, ad campaigns are expensive: [Aggarwal et al. 2023]'s experiment costed $8.9 million USD

This work: a faster and cheaper solution by **transfer learning** + **sensitivity analysis**.

Our Setup

- Estimand: Ad effect in 2024,

$$\theta = \mathrm{E}[Y(1) - Y(0) | 2024 \text{ (i.e. target)}]$$

  - $Y(1)$ is the potential outcome (voted or not) if a voter received ads against Trump.
  - $Y(0)$ is the potential outcome (voted or not) if a voter did not receive ads against Trump.
- Design
  - Source: 2020 experiment from [Aggarwal et al. 2023]
  - Target: 2024 voters in Pennsylvania (∼4.8 million from 67 counties as of April 15, 2024)
  - 2024 covariates $\subset$ 2020 covariates
- Allows
  - Shift in voter demographics between 2024 and 2020 (i.e., covariate shift)
  - Shift in voter turnout between 2024 and 2020 (sensitivity analysis):

$$\underbrace{\mathrm{pr}(Y(a) \mid \textbf{target (2024)}, \text{ voter demographics})}_{\text{unobserved}} \neq \mathrm{pr}(Y(a) \mid \textbf{source (2020)}, \text{voter demographics}), \; a = 0, 1$$

Our Approach: Transfer Learning With Sensitivity Analysis

- Step I: Conduct inference under [Robins, Rotnitzky, and Scharfstein 2000]'s sensitivity model

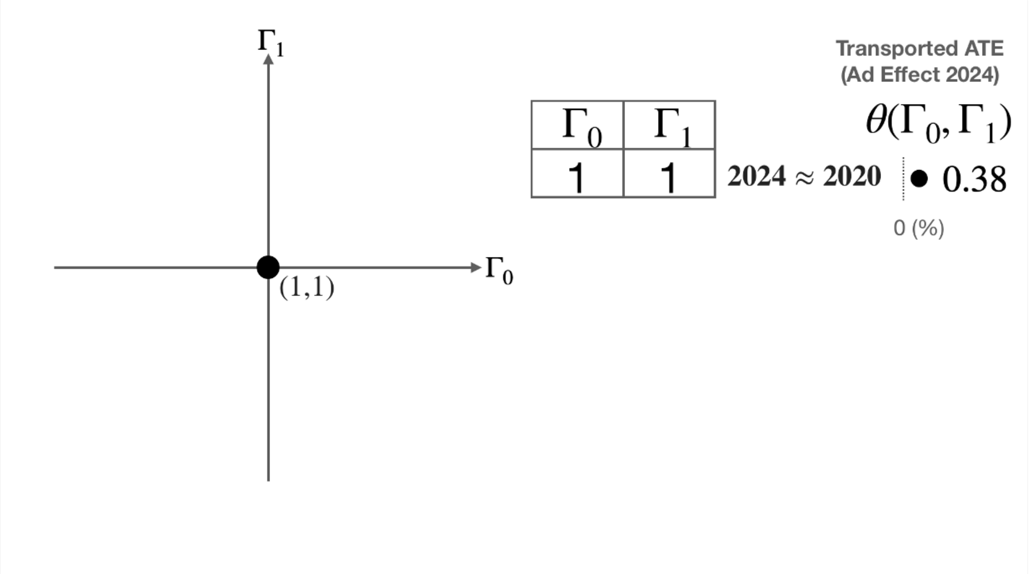
- Step II: Find a plausible range of sensitivity parameters (i.e., calibration)

Step I: Conduct Inference Under The Sensitivity Model

- (Γ_0, Γ_1) quantifies the unobserved shift in $(Y(0), Y(1))$ between 2024 and 2020 via an odds ratio model

$$\Gamma_a = \frac{\text{Odd}(Y(a) \mid \text{ voter demographics, target (2024)})}{\text{Odd}(Y(a) \mid \text{ voter demographics, source (2020)})}, \quad \Gamma_a > 0, \quad a = 0, 1.$$

  - $(\Gamma_0, \Gamma_1) = (1, 1)$ means $2024 \approx 2020$ (**"transportability"**, a widely adopted assumption);
  - Otherwise, $2024 \neq 2020$ (**sensitivity analysis**).

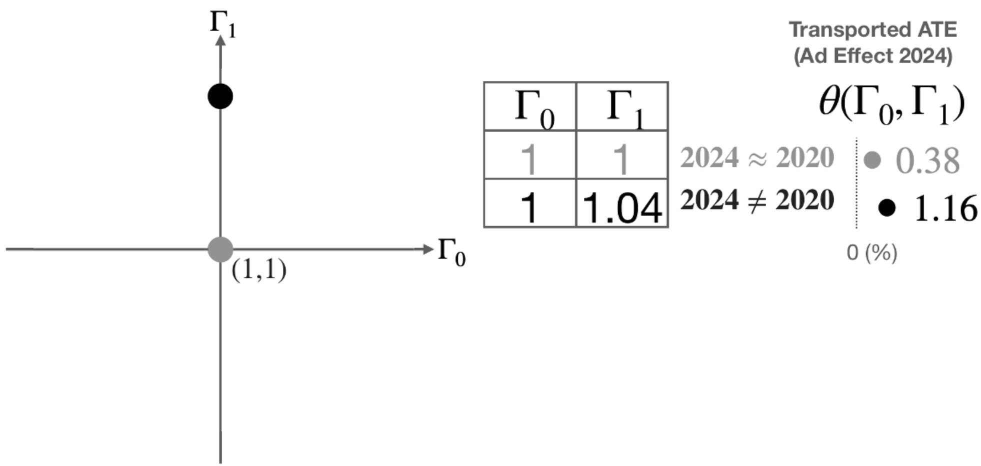


Step I: Conduct Inference Under The Sensitivity Model

- (Γ_0, Γ_1) quantifies the unobserved shift in $(Y(0), Y(1))$ between 2024 and 2020 via an odds ratio model

$$\Gamma_a = \frac{\text{Odd}(Y(a) \mid \text{ voter demographics, target (2024)})}{\text{Odd}(Y(a) \mid \text{ voter demographics, source (2020)})}, \quad \Gamma_a > 0, \quad a = 0, 1.$$

  - $(\Gamma_0, \Gamma_1) = (1, 1)$ means $2024 \approx 2020$ (“**transportability**”, a widely adopted assumption);
  - Otherwise, $2024 \neq 2020$ (**sensitivity analysis**).

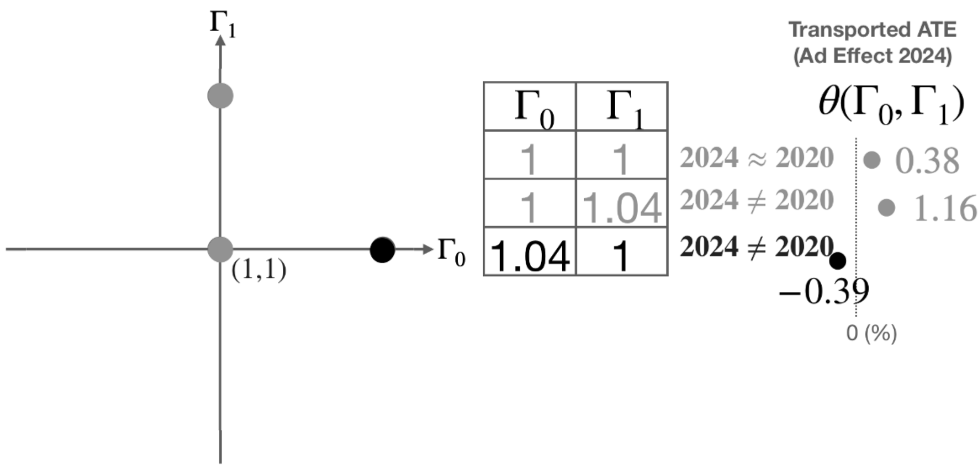


Step I: Conduct Inference Under The Sensitivity Model

- (Γ_0, Γ_1) quantifies the unobserved shift in $(Y(0), Y(1))$ between 2024 and 2020 via an odds ratio model

$$\Gamma_a = \frac{\text{Odd}(Y(a) \mid \text{ voter demographics, target (2024)})}{\text{Odd}(Y(a) \mid \text{ voter demographics, source (2020)})}, \quad \Gamma_a > 0, \quad a = 0, 1.$$

  - $(\Gamma_0, \Gamma_1) = (1, 1)$ means $2024 \approx 2020$ (**"transportability"**, a widely adopted assumption);
  - Otherwise, $2024 \neq 2020$ (**sensitivity analysis**).



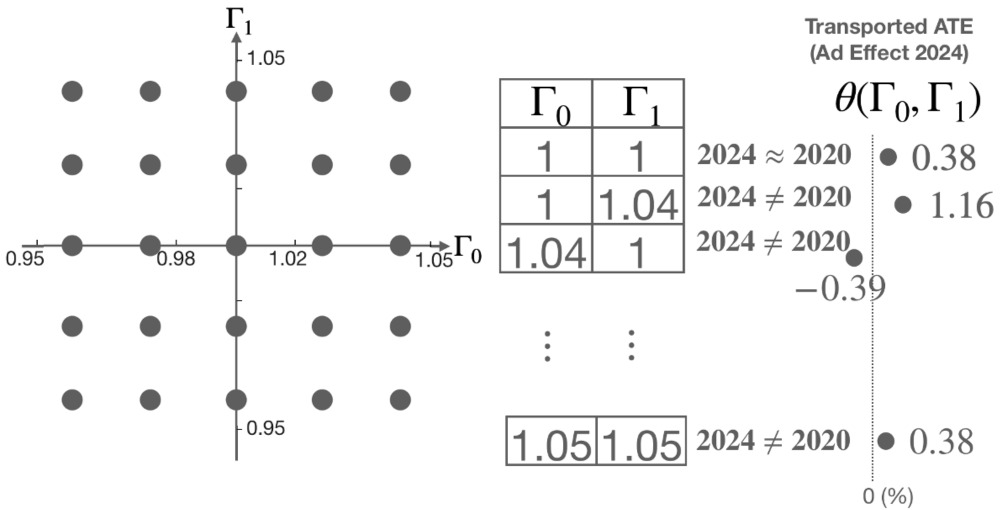
Γ_0	Γ_1
1	1
1	1.04
1.04	1

Step I: Conduct Inference Under The Sensitivity Model

- (Γ_0, Γ_1) quantifies the unobserved shift in $(Y(0), Y(1))$ between 2024 and 2020 via an odds ratio model

$$\Gamma_a = \frac{\text{Odd}(Y(a) \mid \text{ voter demographics, target (2024))}}{\text{Odd}(Y(a) \mid \text{ voter demographics, source (2020))}}, \quad \Gamma_a > 0, \quad a = 0, 1.$$

  - $(\Gamma_0, \Gamma_1) = (1, 1)$ means $2024 \approx 2020$ (**"transportability"**, a widely adopted assumption);
  - Otherwise, $2024 \neq 2020$ (**sensitivity analysis**).



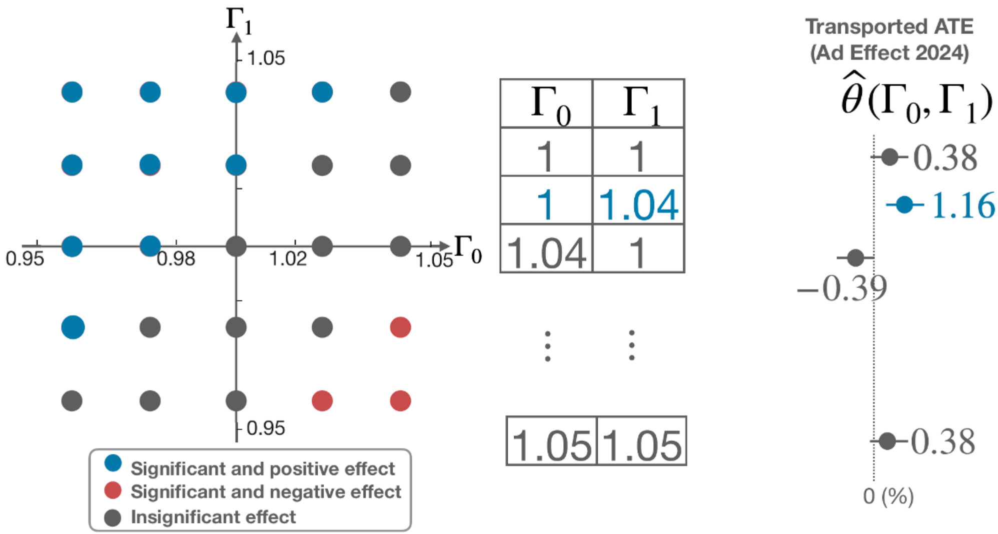
Step I: Conduct Inference Under The Sensitivity Model

- (Γ_0, Γ_1) quantifies the unobserved shift in $(Y(0), Y(1))$ between 2024 and 2020 via an odds ratio model

$$\Gamma_a = \frac{\text{Odd}(Y(a) \mid \text{ voter demographics}, \text{ target (2024)})}{\text{Odd}(Y(a) \mid \text{ voter demographics}, \text{ source (2020)})}, \quad \Gamma_a > 0, \quad a = 0, 1.$$

  - $(\Gamma_0, \Gamma_1) = (1, 1)$ means 2024 $\approx$ 2020 (**"transportability"**, a widely adopted assumption);
  - Otherwise, 2024 $\neq$ 2020 (**sensitivity analysis**).

- Calculate p-values using the proposed plug-in estimator (nonparametric) with bootstrap.



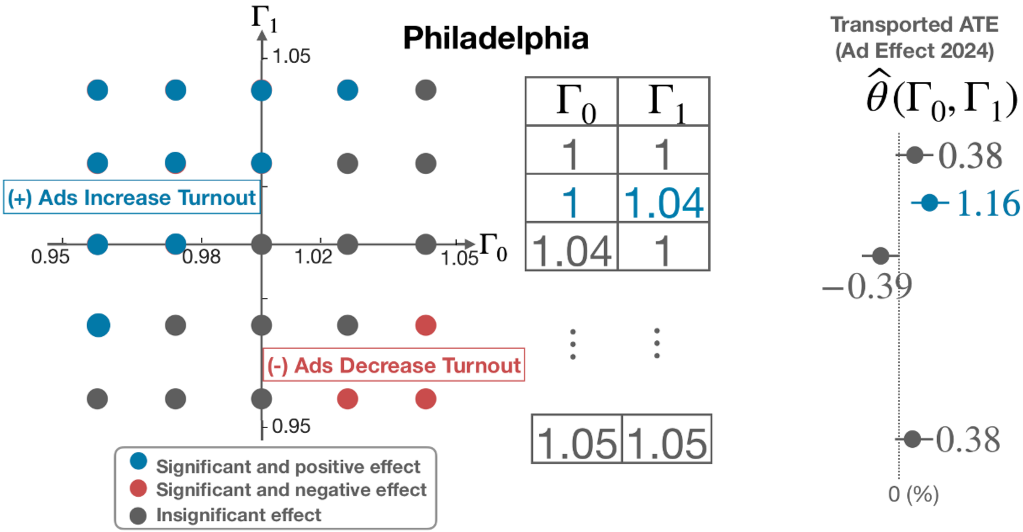
Step I: Conduct Inference Under The Sensitivity Model

- (Γ_0, Γ_1) quantifies the unobserved shift in $(Y(0), Y(1))$ between 2024 and 2020 via an odds ratio model

$$\Gamma_a = \frac{\text{Odd}(Y(a) \mid \text{ voter demographics, target (2024)})}{\text{Odd}(Y(a) \mid \text{ voter demographics, source (2020)})}, \quad \Gamma_a > 0, \quad a = 0, 1.$$

  - $(\Gamma_0, \Gamma_1) = (1, 1)$ means $2024 \approx 2020$ ("**transportability**", a widely adopted assumption);
  - Otherwise, $2024 \neq 2020$ (**sensitivity analysis**).

- Calculate p-values using the proposed plug-in estimator (nonparametric) with bootstrap.



Step II: Find A Plausible Range of Sensitivity Parameters (i.e., Calibration)

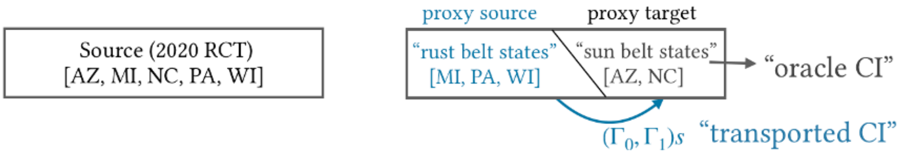
- Sensitivity parameter (Γ_0, Γ_1) is not identified since we did not run an experiment in 2024
- In sensitivity analysis, there is always a question about what is a “large”, “small”, or a “plausible” sensitivity parameter (Γ_0, Γ_1)
- Domain experts may specify a range of sensitivity parameters $\mathcal{C}$, and estimate the ad effect with $(\Gamma_0, \Gamma_1) \in \mathcal{C}$
- For investigators without such knowledge, we present one solution to this question by creating a **dis-similar** partition of the source data

Step II: Find a Plausible Range of Sensitivity Parameters (i.e., Calibration)

Key idea: Find (Γ_0, Γ_1)s that represent the unmeasured differences between two **dissimilar** source subsets.

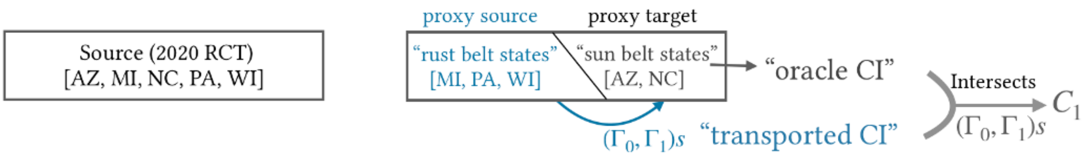
Step II: Find A Plausible Range of Sensitivity Parameters (i.e., Calibration)

Key idea: Find (Γ_0, Γ_1)s that represent the unmeasured differences between two **dissimilar** source subsets.



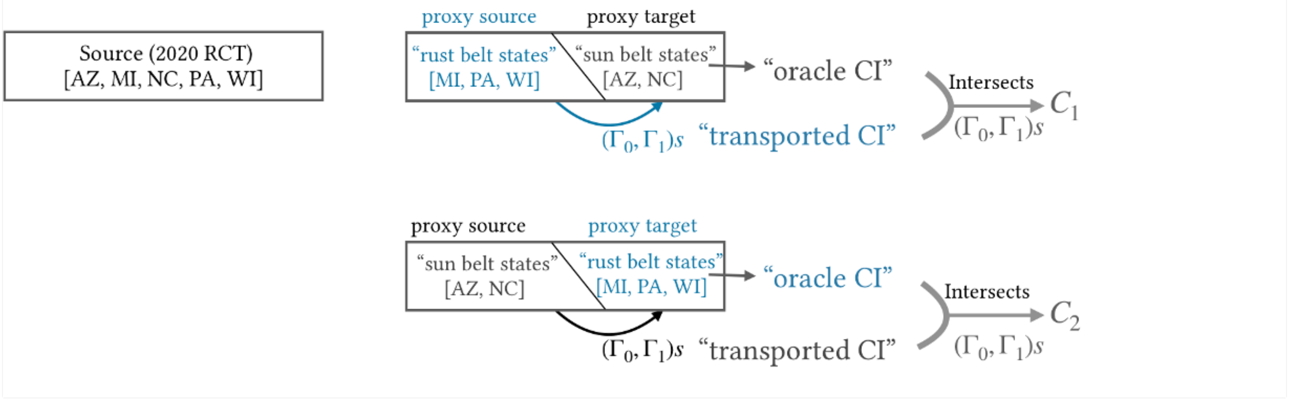
Step II: Find A Plausible Range of Sensitivity Parameters (i.e., Calibration)

Key idea: Find (Γ_0, Γ_1)s that represent the unmeasured differences between two **dissimilar** source subsets.



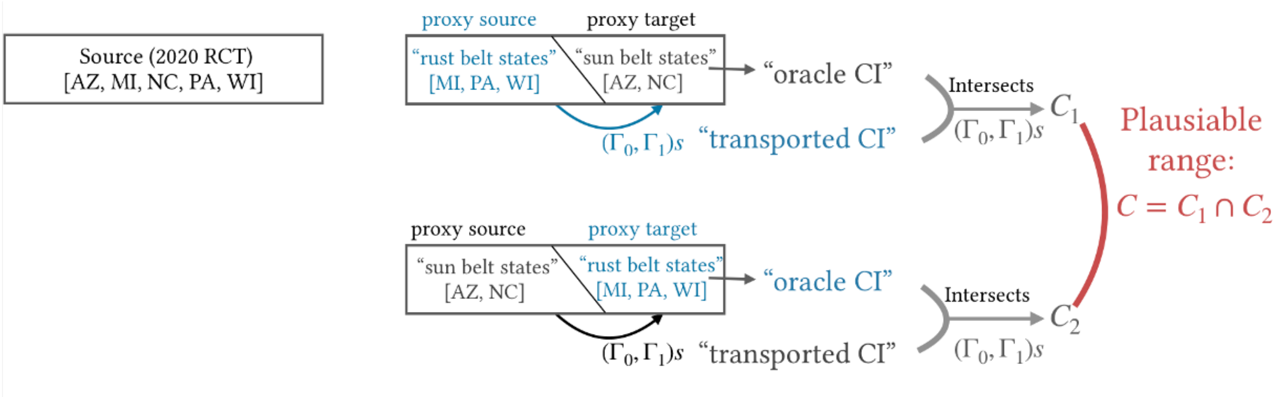
Step II: Find A Plausible Range of Sensitivity Parameters (i.e., Calibration)

Key idea: Find (Γ_0, Γ_1)s that represent the unmeasured differences between two **dissimilar** source subsets.



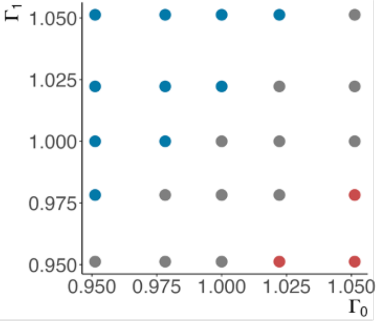
Step II: Find A Plausible Range of Sensitivity Parameters (i.e., Calibration)

Key idea: Find (Γ_0, Γ_1)s that represent the unmeasured differences between two **dissimilar** source subsets.

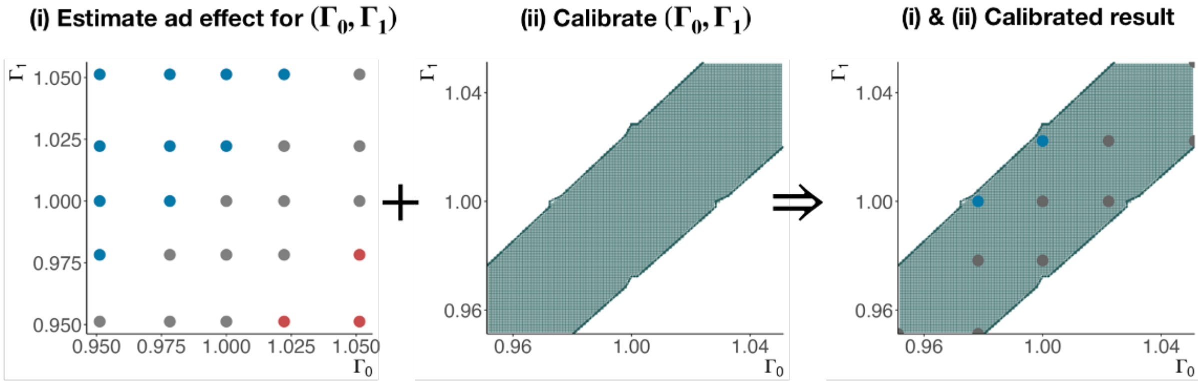


Example: Calibrating Sensitivity Parameters in Philadelphia

(i) Estimate ad effect for (Γ_0, Γ_1)



Example: Calibrating Sensitivity Parameters in Philadelphia



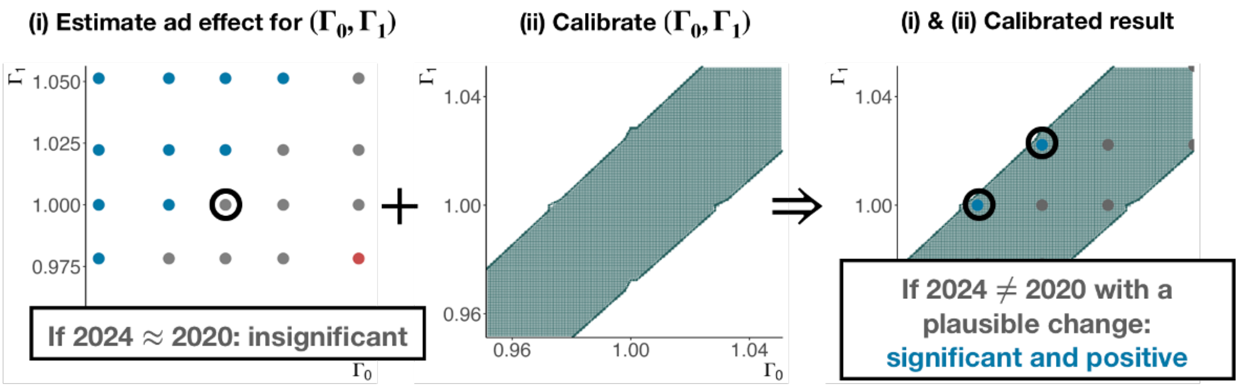
Example: Calibrating Sensitivity Parameters in Philadelphia

We focus on the change of conclusions before and after sensitivity analysis:

Before: $(\Gamma_0, \Gamma_1) = (1, 1)$, i.e., $2024 \approx 2020$, the ad effect is insignificant

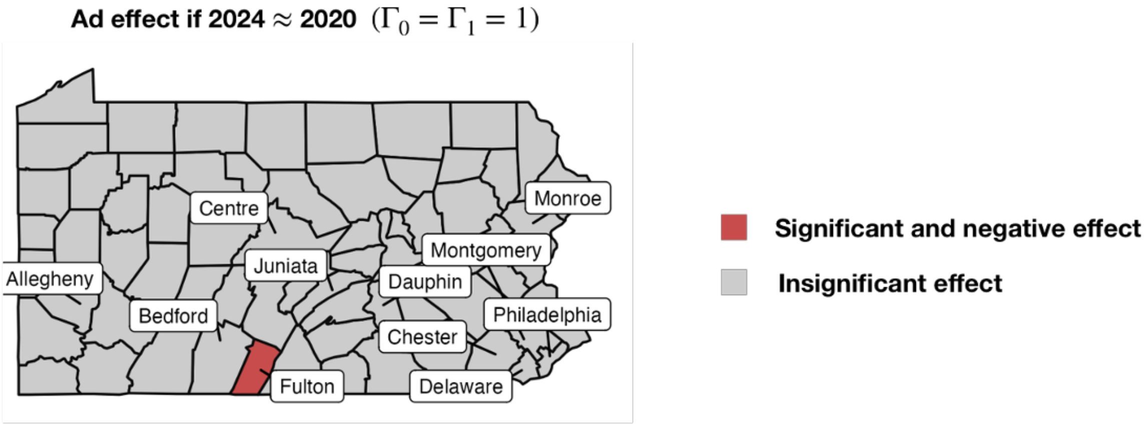
After: $(\Gamma_0, \Gamma_1) \in \mathcal{C}$, i.e., $2024 \neq 2020$ with a plausible change, the ad effect can be significant and positive

Philadelphia is *sensitive* to a significant and positive ad effect.



Data Analysis: Ad Effect Across Counties

- If 2024 $\approx$ 2020, i.e., $(\Gamma_0, \Gamma_1) = (1, 1)$, the digital ads against Trump decreases turnout in Fulton county
  - Trump had 86.03% of votes in Fulton in 2024 (highest margin among counties in PA)

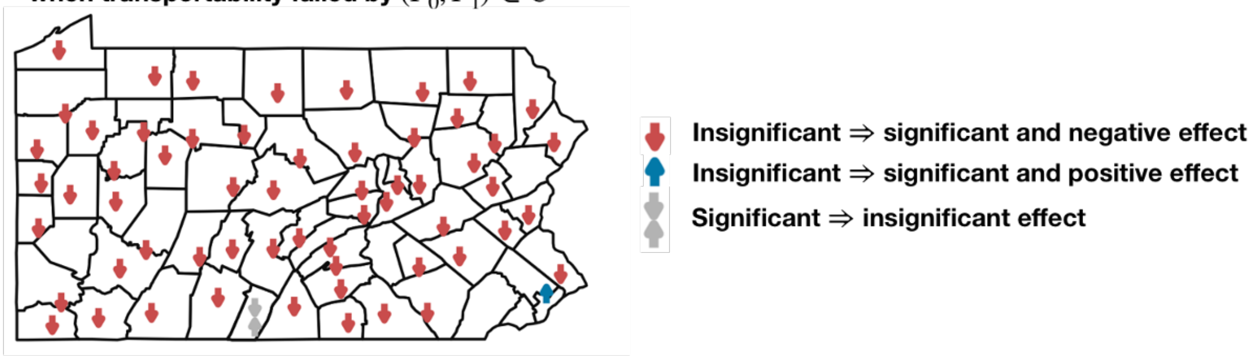


Data Analysis: Ad Effect Across Counties

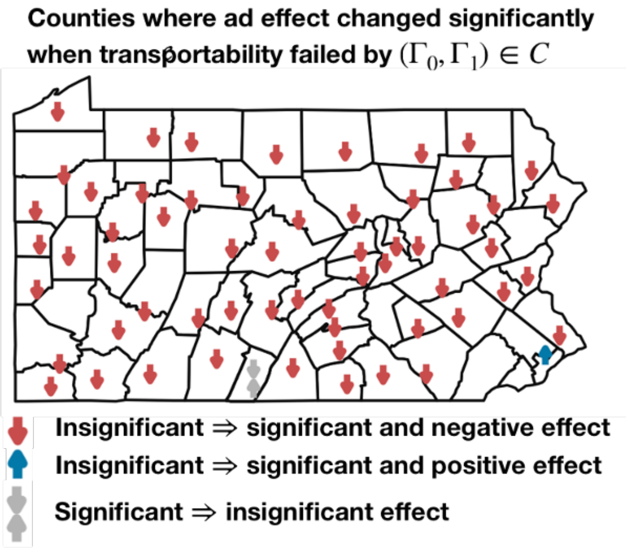
If $2024 \neq 2020$ with a plausible change $(\Gamma_0, \Gamma_1) \in C$:

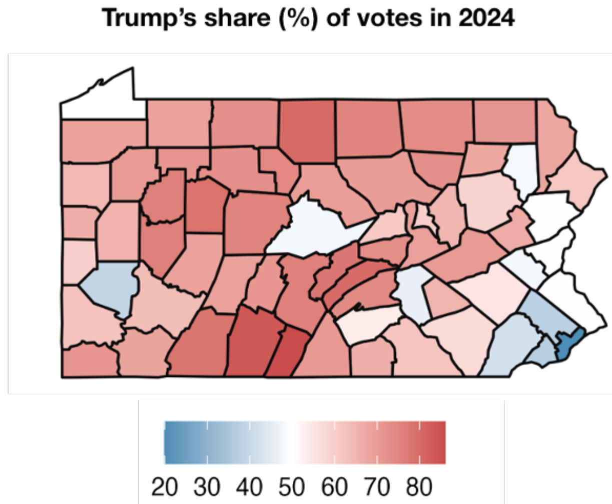
- Some conclusions may change within the calibrated region C (i.e., *sensitive* effects)
  - Philadelphia county is sensitive to a significant and positive effect
  - Fulton county is sensitive to an insignificant effect
  - 59 counties is sensitive to a significant and negative effect
- 6 counties remain to have an insignificant effect (i.e., *insensitive* effects)

Counties where ad effect changed significantly when transportability failed by $(\Gamma_0, \Gamma_1) \in C$



Data Analysis: Ad Effect Across Counties





The direction of the ad effect after calibration roughly corresponds to Trump's share of votes in the 2024 U.S. presidential election.

Summary

- Motivation: From [Aggarwal et al. 2023], would ads against Trump in 2020 remain ineffective in 2024?
- Our approach: transfer learning with sensitivity analysis
  - (a) simple plug-in estimator with bootstrap SE/CIs, (b) efficient influence function based approach (see paper), (c) calibration of sensitivity parameter Γ_a with source data
- Application: would ads against Trump be effective for PA voters in 2024?
  County-by-county analysis:
  - Under $(\Gamma_0, \Gamma_1) = (1, 1)$, the ad effect is significant and negative in Fulton county
  - With the calibrated sensitivity analysis, the direction of sensitive ad effects roughly corresponds to Trump's share of votes in the 2024

  Subgroup analysis: see paper

Preprint: arXiv:2411.01100
Contact: xinran.miao@wisc.edu

Acknowledgements: Xiaobin Zhou (UW-Madison), Adeline Y. Lo (UW-Madison), Sameer Deshpande (UW-Madison), Chan Park (UIUC), Melody Huang (Yale), Ying Jin (Harvard), UW-Madison statistics student seminar participants (April 29, 2024), OCIS participants (April 30, 2024), ACIC participants (May 15, 2024), MEAD participants (February 14, 2025).

References I

- Aggarwal, Minali et al. (2023). **“A 2 million-person, campaign-wide field experiment shows how digital advertising affects voter turnout”.** In: *Nature Human Behaviour* 7.3, pp. 332–341.
- Carroll, Raymond J et al. (1997). **“Generalized partially linear single-index models”.** In: *Journal of the American Statistical Association* 92.438, pp. 477–489.
- Chernozhukov, Victor et al. (2017). **“Double/Debiased/Neyman machine learning of treatment effects”.** In: *American Economic Review* 107.5, pp. 261–265.
- Dahabreh, Issa J, Lucia C Petito, et al. (2020). **“Toward causally interpretable meta-analysis: transporting inferences from multiple randomized trials to a new target population”.** In: *Epidemiology* 31.3, pp. 334–344.
- Dahabreh, Issa J, James M Robins, Sebastien J-PA Haneuse, et al. (2023). **“Sensitivity analysis using bias functions for studies extending inferences from a randomized trial to a target population”.** In: *Statistics in Medicine* 42.13, pp. 2029–2043.
- Dahabreh, Issa J, James M Robins, Sebastien JP Haneuse, and Miguel A Hernán (2019). **“Generalizing causal inferences from randomized trials: counterfactual and graphical identification”.** In: *arXiv preprint arXiv:1906.10792*.
- Dahabreh, Issa J, James M Robins, Sebastien JP Haneuse, Sarah E Robertson, et al. (2022). **“Global sensitivity analysis for studies extending inferences from a randomized trial to a target population”.** In: *arXiv preprint arXiv:2207.09982*.

References II

- Degtiar, Irina and Sherri Rose (2023). **"A review of generalizability and transportability".** In: *Annual Review of Statistics and Its Application* 10, pp. 501–524.
- Egami, Naoki and Erin Hartman (2023). **"Elements of external validity: Framework, design, and analysis".** In: *American Political Science Review* 117.3, pp. 1070–1088.
- Franks, Alexander M, Alexander D'Amour, and Avi Feller (2020). **"Flexible sensitivity analysis for observational studies without observable implications".** In: *Journal of the American Statistical Association* 115.532, pp. 1730–1746.
- Huang, Melody Y (2024). **"Sensitivity analysis for the generalization of experimental results".** In: *Journal of the Royal Statistical Society Series A: Statistics in Society.*
- Jin, Ying and Dominik Rothenhäusler (2024). **"Tailored inference for finite populations: conditional validity and transfer across distributions".** In: *Biometrika* 111.1, pp. 215–233.
- Kennedy, Edward H (2022). **"Semiparametric doubly robust targeted double machine learning: a review".** In: *arXiv preprint arXiv:2203.06469.*
- Nguyen, Trang Quynh et al. (2017). **"Sensitivity analysis for an unobserved moderator in RCT-to-target-population generalization of treatment effects".** In: *The Annals of Applied Statistics* 11.1, pp. 225–247.
- Nie, Xinkun, Guido Imbens, and Stefan Wager (2021). **"Covariate balancing sensitivity analysis for extrapolating randomized trials across locations".** In: *arXiv preprint arXiv:2112.04723.*

References III

- Robins, James M (2002). **“[Covariance Adjustment in Randomized Experiments and Observational Studies]: Comment”.** In: *Statistical Science* 17.3, pp. 309–321.
- Robins, James M, Andrea Rotnitzky, and Daniel O Scharfstein (2000). **“Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models”.** In: *Statistical Models in Epidemiology, the Environment, and Clinical Trials*. Springer, pp. 1–94.
- Rosenbaum, Paul R (2002). **“Covariance adjustment in randomized experiments and observational studies”.** In: *Statistical Science* 17.3, pp. 286–327.
- Scharfstein, Daniel O et al. (2021). **“Semiparametric sensitivity analysis: Unmeasured confounding in observational studies”.** In: *arXiv preprint arXiv:2104.08300*.
- Tipton, Elizabeth and Laura R Peck (2017). **“A design-based approach to improve external validity in welfare policy evaluations”.** In: *Evaluation Review* 41.4, pp. 326–356.
- Zeng, Zhenghao et al. (2023). **“Efficient Generalization and Transportation”.** In: *arXiv preprint arXiv:2302.00092*.

FAQs About Data

- Is the voting data self-reported?
  No. See [Aggarwal et al. 2023] for details.
- Does the data contain which candidate the voter voted for?
  No.
- Why is your voter data discrete?
  We're not sure. Perhaps, this is done to preserve some privacy?
- Is party registration measured accurately?
  Yes and no. [Aggarwal et al. 2023] and current voter registration data documentation discuss some reasons for errors.
- How was the treatment randomized?
  The randomization was stratified within gender, race, and age groups with the intention of increasing the propensity for women, black, and young people. The average treatment probability was 85.6%.
- Why is there a high proportion of treated individuals?
  We're not sure. Perhaps Acronym wanted to deliver the ads against Trump to as many voters as possible?

FAQs About Data

- Was the randomization done through Facebook/Meta?
  No. Our understanding from [Aggarwal et al. 2023] is that the participants were randomized before the advertising company delivered the ads.

- Were ads delivered?
  Yes and no. 60% of the treatment group participants were identified and served ads. The analysis was intention-to-treat. See [Aggarwal et al. 2023] for details. We followed [Aggarwal et al. 2023] and considered intention-to-treat effect.

FAQs About Framework And Assumptions

- Is SUTVA violated?
  Great question! It is possible, especially if
  - Different doses of ads: $Y(700 \text{ ads}) \neq Y(800 \text{ ads}) \neq Y(1)$
  - Different ads in 2020 and 2024: $Y(\text{ad in } 2020) \neq Y(\text{ad in } 2024)$
  - Voters talk to each other due to ads: $Y(\text{my trt, your trt}) \neq Y(\text{my trt})$
  - There are carry-over effects from 2020 ad campaign into 2024

  But, we also picked 2020 and 2024 to minimize SUTVA violations as Trump remains the Republican candidates between the two years.

- Is your population infinite?
  Excellent question! Our framework implicitly assumes that the units in the target and the source data are sampled from an infinite population of voters. But, it may be more appropriate to treat the PA voter registration database as a finite population or a large sample from a finite population. See [Jin and Rothenhäusler 2024] for transfer learning when the target sample is fixed. We are also happy to discuss more.

FAQs About Framework And Assumptions

- Is the data from 2020 independent from 2024?
  Excellent question! Theoretically, this concern is a bit tricky to address, especially if 2024 is a fixed, census-level data. We're happy to talk more about this.
  Application-wise, we repeated the analysis after excluding voters in PA, WI, MI from the source data. Results exhibit similar patterns but are less powerful (see Appendix of the paper). Although this ensures independence between the source and target, it also increases the difference between the source and target. We believe that in general the source and target should be as similar as possible, so we decided to report the results when using all data in [Aggarwal et al. 2023] as the source.

- What about treating this data as longitudinal?
  Excellent idea, but this requires measuring same voter over time. This data is not easy to get.

FAQs About Sensitivity Analysis

- Can you give us some understanding on the odds ratio?
  Yes. The turnout in PA is 76.5% in 2020 and 76.6% in 2024. The odds ratio is roughly 1.0055 (logarithm: 0.0055).

- Can the sensitivity model depend on covariates?
  Yes. While this introduces more complexity in interpreting the sensitivity parameters, it could be useful if there is a priori knowledge about how the ad effects in 2024 and 2020 differ with respect to measured covariates. For example, for $A = 0$, we can define

$$\Gamma_{\text{Democrat?}} I(\text{Democrat?}), \quad \Gamma_{\text{Republican?}} I(\text{Republican?})$$

  if we believe the change between 2020 and 2024 is different for Democrats and Republicans. We call this a **local sensitivity model.**

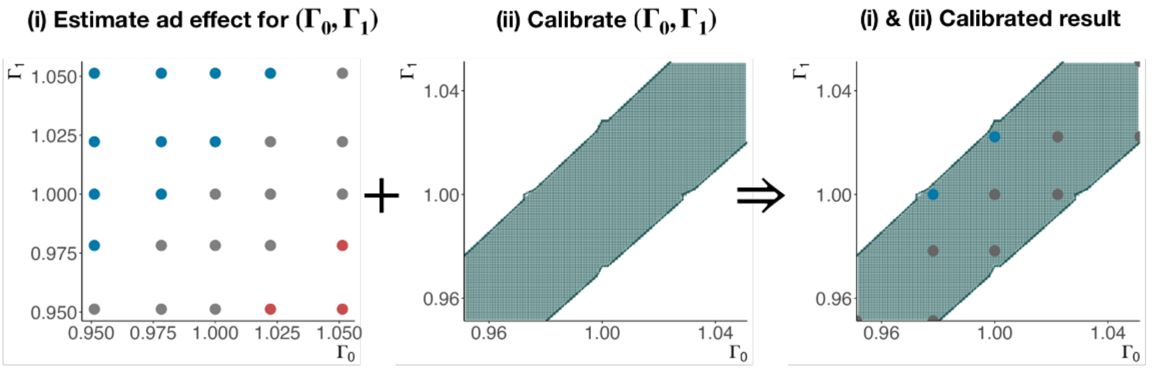
FAQs About Analysis

- Is your conclusion valid after Biden dropped out?
  Great question! Our original analysis plan assumed that President Biden is the Democratic Party's nominee for the presidency. While we believe the interpretations from our analysis about ads will still be plausible since Trump is the nominee for the Republication party and the digital ad campaign consists of negative ads against Trump, we caution readers from over-interpreting the results. Notably, our calibration procedure based on rust belt and sun belt states could under-estimate the dramatic shift in electoral context after Biden dropped out of the race and the consequences of this unprecedented event in American politics.

- Is your conclusion sensitive to data quality from 2024 voter registration data (i.e. target data)?
  Yes. Unfortunately, high quality target data is expensive.

FAQs About Calibration

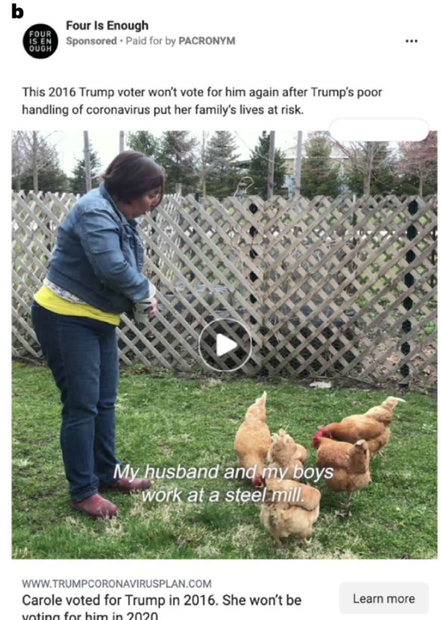
- Can you explain the shape of your calibrated region?
  Sure. The off-diagonal area represents case when $|\Gamma_1 - \Gamma_0|$ is large, i.e., when the change in treated voters differs a lot from the change in control voters from 2020 to 2024.
  - It can be unrealistic when $\Gamma_1 > 1$ is very high and $\Gamma_0 < 1$ is very low
  - The calibration rules out such extreme scenarios



Example Political Ads in Treatment Group

Ads were on Facebook, Instagram, and Outbrain advertising network.

(a): ads promoting news stories



(b): traditional video campaign ads

Some Prior Works

The closest related literature is transportability/generalizability.

We are definitely not the first nor only ones to incorporate sensitivity analysis in transportability/generalizability. A very small, partial list:

- Linear outcome sensitivity model: [Nguyen et al. 2017; Dahabreh, Petito, et al. 2020; Dahabreh, Robins, S. J.-P. Haneuse, et al. 2023; Zeng et al. 2023]
- Exponential tilting sensitivity model: [Dahabreh, Robins, S. J. Haneuse, Robertson, et al. 2022]
- Marginal sensitivity model for domain adaptation: [Nie, Imbens, and Wager 2021]
- Omitted variable bias approach with weighted estimators: [Huang 2024]

Our goal is to tailor these methods to address our key question.

Notation and Causal Assumptions

- Population type: $S \in \{0, 1\}$ where $S = 1$ is source (e.g. 2020) and $S = 0$ is target (e.g. 2024 PA voters)
- Outcome: $Y \in \{0, 1\}$ where $Y = 1$ is voted
- Treatment: $A \in \{0, 1\}$ where $A = 1$ is ad against Trump
- Covariates: $\mathbf{X} \in \mathbb{R}^p$ where
  - Source covariates: $\mathbf{X}$
  - Target covariates: $\mathbf{V} \subset \mathbf{X}$
    $\mathbf{V} =$ (age group, gender, party, part of voting history)
    $\mathbf{X} = \mathbf{V}$ + (race, missing voting history)
- Potential outcomes: $Y(a) \in \{0, 1\}$, $a \in \{0, 1\}$
  - $Y(1)$: voted if, potentially contrary to fact, voter received ads
  - $Y(0)$: voted if, potentially contrary to fact, voter didn't receive ads
- Causal estimand: $\theta = \mathbb{E}[Y(1) - Y(0) \mid S = 0]$
  - Average treatment effect; the ad effect on turnout among 2024 PA voters

Data Table for Our Setup

		$\mathbf{X}$					
	S	$\mathbf{V}$	$\mathbf{X} \setminus \mathbf{V}$	A	$Y(1)$	$Y(0)$	Y
	1	✓	✓	1	✓		✓
	⋮	⋮	⋮	⋮	⋮		⋮
Source RCT (i.e. 2020)	1	✓	✓	1	✓		✓
	1	✓	✓	0		✓	✓
	⋮	⋮	⋮	⋮		⋮	⋮
	1	✓	✓	0		✓	✓
	0	✓					
Target (i.e. 2024 PA voters)	⋮	⋮					
	0	✓					

The goal is to identify and estimate $\theta = \mathbb{E}[Y(1) - Y(0) \mid S = 0]$.

Review Of Transportability In RCT: Assumptions On The Source ($S = 1$)

Causal assumptions on the source (under stratified RCT):

(A1) Stable unit treatment value assumption (SUTVA):
Under $S = 1$, if $A = a$, $Y = Y(a)$

- The observed Y is one realization of the two potential outcomes
- There are no multiple versions of treatment, e.g.,

$$Y(700 \text{ ads}) = Y(800 \text{ ads}) = Y(1)$$

- There is no interference: a voter's voting result cannot be affected by other voters' treatments

(A2) Randomized treatment: $A \perp Y(1), Y(0) \mid \mathbf{X}, S = 1$

(A3) Overlap of treatment: $0 < \pi(\mathbf{x}) = \mathrm{pr}(A = 1 \mid \mathbf{X} = \mathbf{x}, S = 1) < 1$

- The treated individuals and untreated individuals have a common support of $\mathbf{X}$
- The propensity score $\pi(\mathbf{x})$ is known in an RCT

(A2)-(A3) are satisfied with the 2020 RCT by [Aggarwal et al. 2023].

Review Of Transportability In RCT: Assumptions For Transportation

(A4) Overlap between source and target: $0 < \text{pr}(S = 1 \mid \mathbf{V} = \mathbf{v}) < 1$ for all $\mathbf{v}$

- The source voters and target voters have a common support of $\mathbf{V}$
- It can be empirically verified

(A5) Transportability: $Y(1), Y(0) \perp S \mid \mathbf{V}$

- $\underbrace{\text{pr}(Y(a) = 1 \mid \mathbf{V}, S = 0)}_{\text{2024 PA voters}} = \underbrace{\text{pr}(Y(a) = 1 \mid \mathbf{V}, S = 1)}_{\text{2020 RCT}}$
- It is often assumed, but cannot be verified
- We assume it for this short review, but will relax it shortly

See [Tipton and Peck 2017; Dahabreh, Robins, S. J. Haneuse, and Hernán 2019; Egami and Hartman 2023; Degtiar and Rose 2023] for details.

Review: Identification Under (A1)-(A5)

Let $\mu_a = \mathbb{E}[Y \mid \mathbf{X}, A = a, S = 1]$. Under (A1)-(A5), the target ATE θ is identified:

$$\begin{aligned}\theta &= \mathbb{E}[Y(1) - Y(0) \mid S = 0] \\ &= \underbrace{\mathbb{E}[\underbrace{\mathbb{E}[\underbrace{\mu_1(\mathbf{X}) - \mu_0(\mathbf{X})}_{\text{CATE}(\mathbf{X}) \text{ in source}} \mid \mathbf{V}, S = 1]}_{\text{CATE}(\mathbf{V}) \text{ in source}} \mid S = 0]}_{\text{Reweigh CATE}(\mathbf{V}) \text{ to target}}\end{aligned}$$

For reference, when $\mathbf{X} = \mathbf{V}$, we have

$$\theta = \mathbb{E}[\mathbb{E}[\mu_1(\mathbf{X}) - \mu_0(\mathbf{X}) \mid S = 0].$$

In other words, when source and target covariates differ, we have more nuisance parameters (i.e. projection of CATE($\mathbf{X}$) onto $\mathbf{V}$); see recent work by [Zeng et al. 2023].

What If Transportability (A5) Fails? Sensitivity Analysis

Suppose transportability (A5) does not hold:

$$\underbrace{\mathrm{pr}(Y(a) \mid \mathbf{V}, S=0)}_{\text{2024 (i.e., target)}} \neq \underbrace{\mathrm{pr}(Y(a) \mid \mathbf{V}, S=1)}_{\text{2020 (i.e., source)}}$$

For each $Y(a)$, we measure the deviation between the two probabilities via Γ_a:

$$\Gamma_a = \frac{\mathrm{Odd}(Y(a) \mid \mathbf{v}, S=0)}{\mathrm{Odd}(Y(a) \mid \mathbf{v}, S=1)}, \ \Gamma_a > 0 \tag{1}$$

$$\mathrm{Odd}(Y(a) \mid \mathbf{v}, s) = \frac{\mathrm{pr}(Y(a)=1 \mid \mathbf{V}=\mathbf{v}, S=s)}{1-\mathrm{pr}(Y(a)=1 \mid \mathbf{V}=\mathbf{v}, S=s)}, \ s \in \{0,1\} \tag{2}$$

Broadly speaking, the unobservable part ($S=0$; i.e., 2024) differs from the observable part ($S=1$; i.e., 2020) by Γ_a.

- $\Gamma_a = 1$ means $\mathrm{pr}(Y(a) \mid \mathbf{V}, S=0) = \mathrm{pr}(Y(a) \mid \mathbf{V}, S=1)$; i.e., transportability
- In general, a large $|\Gamma_a - 1| \Rightarrow$ a large difference between 2020 and 2024
- Since the red part is unobserved, Γ_a cannot be estimated. Instead, Γ_a is a sensitivity parameter chosen to quantify the difference between the target and the source

Some Remarks On The Sensitivity Model

$$\Gamma_a = \frac{\text{Odd}(Y(a) \mid \mathbf{v}, S = 0)}{\text{Odd}(Y(a) \mid \mathbf{v}, S = 1)}$$

Some remarks:

- The sensitivity model does not place any observable restrictions on the data [Robins, Rotnitzky, and Scharfstein 2000; Franks, D'Amour, and Feller 2020]
- A pseudo-R^2 version of Γ_a is in Proposition 3 of [Franks, D'Amour, and Feller 2020].
- We can also reparametrize the sensitivity model in terms of $P(S = 1 \mid Y(1), Y(0), \mathbf{V} = \mathbf{v})$; see Appendix and [Carroll et al. 1997].
- The sensitivity model can depend on covariates (e.g. $\exp(\Gamma_{\mathbf{v}}^\intercal \mathbf{v} + \ldots)$; "local" sensitivity analysis)
- Some works that use this model: [Robins, Rotnitzky, and Scharfstein 2000; Franks, D'Amour, and Feller 2020; Scharfstein et al. 2021; Dahabreh, Robins, S. J. Haneuse, Robertson, et al. 2022]
- There is a **long and healthy** discussion about what constitutes a "good" model for sensitivity analysis [Robins 2002; Rosenbaum 2002]

Alternate Parametrization of The Sensitivity Model

- The sensitivity model $\Gamma_a = \dfrac{\mathrm{Odd}(Y(a) \mid \mathbf{v}, S=0)}{\mathrm{Odd}(Y(a) \mid \mathbf{v}, S=1)}$ implies the following partially linear logistic regression model [Carroll et al. 1997]:

$$\mathrm{pr}(S = 1 \mid Y(a) = y, \mathbf{V} = \mathbf{v}) = \mathrm{expit}\left(-\log(\Gamma_a)y - \eta_a(\mathbf{v})\right)$$

$$\eta_a(\mathbf{v}) = \log\left(\frac{\mathrm{pr}(S=0)}{\mathrm{pr}(S=1)} \frac{w(\mathbf{v})}{\mathrm{E}\{\Gamma_a \exp(Y(a)) \mid \mathbf{v}, S=1\}}\right)$$

$$w(\mathbf{V}) = p(\mathbf{V} \mid S=0)/p(\mathbf{V} \mid S=1)$$

Identification Under (A1)-(A4) + Sensitivity Model

Again, let $\mu_a(\mathbf{X}) = \mathrm{E}(Y \mid \mathbf{X}, A = a, S = 1)$. Under (A1)-(A4) and the sensitivity model, we have

$$\mathrm{E}[Y(a) \mid S = 0] = \mathrm{E}\left[\frac{\mathrm{E}\{\Gamma_a \mu_a(\mathbf{X}) \mid \mathbf{V}, S = 1\}}{\mathrm{E}\{\Gamma_a \mu_a(\mathbf{X}) + 1 - \mu_a(\mathbf{X}) \mid \mathbf{V}, S = 1\}}\middle| S = 0\right].$$

- It is an exponential tilt of $\rho_a(\mathbf{V}) = \mathbb{E}[\mu_a(\mathbf{X}) \mid \mathbf{V}, S = 1]$
- If $\Gamma_a = 0$ (i.e. transportability (A5) holds), we return to the previous result:

$$\mathbb{E}[Y(a) \mid S = 0] = \mathbb{E}[\mathbb{E}[\mu_a(\mathbf{X}) \mid \mathbf{V}, S = 1] \mid S = 0]$$

Estimation: Simple, Plug-In Estimator

Identification leads to a simple, plug-in estimator:

1. Estimate $\rho_a(\mathbf{V}) = \mathbb{E}[\mu_a(\mathbf{X}) \mid \mathbf{V}, S = 1]$ from source data
   - An example: for $a = 1$, regress $Y/\hat{\pi}(\mathbf{X})$ on $\mathbf{V}$ to get $\hat{\rho}_1(\mathbf{V})$

```
rho1 <- lm(Y ~ AgeGroup + Gender + Party,
           data = source |> filter(a == 1), weights = 1 / pi)

rho0 <- lm(Y ~ AgeGroup + Gender + Party,
           data = source |> filter(a == 0), weights = 1 / (1 - pi))
```

   - Because source is an RCT, $\rho_a(\mathbf{V})$ can be consistently estimated
2. Average the exponentially tilted $\rho_a(\mathbf{V})$ among target sample

$$\widehat{\mathbb{E}}[Y(a) \mid S = 0] = \frac{1}{n_t} \sum_{i \in \text{Target}} \frac{\Gamma_a \widehat{\rho}_a(\mathbf{V}_i)}{\Gamma_a \hat{\rho}_a(\mathbf{V}_i) + 1 - \hat{\rho}_a(\mathbf{V}_i)}$$

$$\widehat{\theta}(\Gamma_1, \Gamma_0) = \widehat{\mathbb{E}}[Y(1) \mid S = 0] - \widehat{\mathbb{E}}[Y(0) \mid S = 0]$$

```
rho1pred <- predict(rho1, data = target)
rho0pred <- predict(rho0, data = target)
theta <- Gamma1 * rho1pred / (Gamma1 * rho1pred + 1 - rho1pred) -
         Gamma0 * rho0pred / (Gamma0 * rho0pred + 1 - rho0pred)
```

Bootstrapping for Transfer Learning

In general, bootstrapping is easy and convenient for estimating SEs or confidence intervals (CIs). Here, we lay out one (theoretically valid) bootstrap for transfer learning with sensitivity analysis. In each bootstrap iteration $b \in \{1, \cdots, B\}$:

1. Resample source data with replacement of size n_s, obtain data $\mathcal{D}_{\mathcal{S}}^*$
2. Resample target data with replacement of size n_t, obtain data $\mathcal{D}_{\mathcal{T}}^*$.
3. With $\mathcal{D}_{\mathcal{S}}^*$ and $\mathcal{D}_{\mathcal{T}}^*$, construct the ATE estimator $\widehat{\theta}_b^*$ from above.

Take $\alpha/2$ and $1 - \alpha/2$ quantiles of $\left\{\widehat{\theta}_b^*\right\}_{b=1}^B$ as a $1 - \alpha$ CI of θ.

Theorem: If $\rho_a(\mathbf{V})$ is smooth enough and Donsker condition holds, the above procedure yields a valid $1 - \alpha$ CI of θ.

The smoothness + Donsker conditions hold for our discrete voter data.

EIF-Based Estimator (Brief Introduction)

For given Γ_a, we have the efficient influence function (EIF) of θ and it directs an EIF-based estimator.

- Four nuisance functions: (i) propensity score $\pi(\mathbf{X})$, (ii) outcome regression $\mu_a(\mathbf{X})$, (iii) projection of outcome regression $\rho_a(\mathbf{V})$, and (iv) population weights $w(\mathbf{V}) = \mathrm{pr}(\mathbf{V} \mid S = 0)/\mathrm{pr}(\mathbf{V} \mid S = 1)$
- The efficient influence function (EIF) is

$$\mathrm{EIF}(\mathbf{O}_i, \theta_a(\Gamma_a)) = \frac{S_i w(\mathbf{V}_i)}{\mathrm{pr}(S_i = 1)} \frac{\Gamma_a}{[\Gamma_a \rho_a(\mathbf{V}_i) + 1 - \rho_a(\mathbf{V}_i)]^2} \left[\left\{ \frac{A_i}{\pi(\mathbf{X}_i)} + \frac{1 - A_i}{1 - \pi(\mathbf{X}_i)} \right\} \{Y_i - \mu_a(\mathbf{X}_i)\} + \mu_a(\mathbf{X}_i) - \rho_a(\mathbf{V}_i) \right] + \frac{1 - S_i}{\mathrm{pr}(S_i = 0)} \left[\frac{\Gamma_a \rho_a(\mathbf{V}_i)}{\Gamma_a \rho_a(\mathbf{V}) + 1 - \rho_a(\mathbf{V}_i)} - \theta_a(\Gamma_a) \right]$$

- Our EIF recovers [Zeng et al. 2023]'s EIF when $\Gamma_a = 1$:

$$\mathrm{EIF}(\mathbf{O}_i, \theta_a(\Gamma_a)) = \frac{S_i w(\mathbf{V}_i)}{\mathrm{pr}(S_i = 1)} \left[\left\{ \frac{A_i}{\pi(\mathbf{X}_i)} + \frac{1 - A_i}{1 - \pi(\mathbf{X}_i)} \right\} \{Y_i - \mu_a(\mathbf{X}_i)\} + \mu_a(\mathbf{X}_i) - \rho_a(\mathbf{V}_i) \right] + \frac{1 - S_i}{\mathrm{pr}(S_i = 0)} \rho_a(\mathbf{V}_i)$$

EIF-Based Estimator (High-Level Remarks)

Some remarks:

- Practically speaking, the estimator is useful if $\mathbf{V}$ is continuous
- To avoid parametric assumptions, we need cross-fitting
- The estimator is not doubly robust when $\Gamma_a \neq 0$; see Appendix D of [Dahabreh, Robins, S. J. Haneuse, Robertson, et al. 2022] for a related comment when $\mathbf{V} = \mathbf{X}$
- EIF-based estimator does not reduce the "plug-in bias" from $\rho_a(\mathbf{V})$ when $\Gamma_a \neq 1$

EIF-Based Estimator (Cross Fitting)

We follow the modern trend in causal inference where we use cross-fitting [Chernozhukov et al. 2017; Kennedy 2022] and the EIF to estimate $\theta_a(\Gamma_a)$.

1. Randomly partition the source and target sample indices $\mathcal{I}_s$ and $\mathcal{I}_t$ into K disjoint sets, $\mathcal{I}_s^{(k)}$ and $\mathcal{I}_t^{(k)}$, respectively, for $k = 1, 2, \cdots, K$ with some pre-specified integer K.
2. For each k, estimate nuisance functions with data in $\mathcal{I} \setminus \mathcal{I}^{(k)}$ and denote them as $\widehat{\pi}^{(k)}(\mathbf{x})$, $\widehat{\mu}_a^{(k)}(\mathbf{x})$, $\widehat{w}^{(k)}(\mathbf{v})$ and $\widehat{\rho}_a^{(k)}(\mathbf{v})$. Plug them into the "uncentered" EIF and evaluate it with the data in $\mathcal{I}^{(k)}$, i.e.,

$$\widehat{\theta}^{(k)}_{\text{EIF,a}}(\Gamma_a) = \frac{1}{|\mathcal{I}_s^{(k)}|} \sum_{i \in \mathcal{I}_s^{(k)}} \frac{\Gamma_a \widehat{w}^{(k)}(\mathbf{V}_i)}{[\Gamma_a \widehat{\rho}_a^{(k)}(\mathbf{V}_i) + 1 - \widehat{\rho}_a^{(k)}(\mathbf{V}_i)]^2} \left[\left\{ \frac{A_i}{\widehat{\pi}^{(k)}(\mathbf{X}_i)} + \frac{1 - A_i}{1 - \widehat{\pi}^{(k)}(\mathbf{X}_i)} \right\} \{ Y_i - \widehat{\mu}_a^{(k)}(\mathbf{X}_i) \} \right.$$
$$\left. + \widehat{\mu}_a^{(k)}(\mathbf{X}_i) - \widehat{\rho}_a^{(k)}(\mathbf{V}_i) \right] + \frac{1}{|\mathcal{I}_t^{(k)}|} \sum_{i \in \mathcal{I}_t^{(k)}} \frac{\Gamma_a \widehat{\rho}_a^{(k)}(\mathbf{V}_i)}{\Gamma_a \widehat{\rho}_a^{(k)}(\mathbf{V}_i) + 1 - \widehat{\rho}_a^{(k)}(\mathbf{V}_i)}.$$

3. Take an average of $\widehat{\theta}^{(k)}_{\text{EIF,a}}(\Gamma_a)$ to arrive at the EIF-based cross-fitting estimator of $\theta_{\text{EIF,a}}(\Gamma_a)$, which we denote as $\widehat{\theta}_{\text{EIF,a}}(\Gamma_a) = \sum_{k=1}^K \widehat{\theta}^{(k)}_{\text{EIF,a}}(\Gamma_a)/K$.

EIF-Based Estimation

Theorem 1

Suppose regularity assumptions hold. Then we have
(i) [Conditional Double Robustness]. Suppose $\widehat{\rho}_a^{(k)}$ *is a consistent estimator of* $\rho_a^{(k)}$. *Then,* $\widehat{\theta}_{\mathrm{EIF,a}}(\Gamma_a) \to_p \theta_a(\Gamma_a)$ *if*

$$\|\widehat{\pi}^{(k)}(\mathbf{X}_i) - \pi^{(k)}(\mathbf{X}_i)\| \cdot \|\widehat{\mu}_a^{(k)}(\mathbf{X}_i) - \mu_a^{(k)}(\mathbf{X}_i)\| = o_p(1), \tag{3}$$

(ii) [Asymptotic normality and Semiparametric Efficiency] Suppose $\widehat{\rho}_a^{(k)}$, $\widehat{\mu}_a^{(k)}$, $\widehat{w}^{(k)}$, *and* $\widehat{\pi}^{(k)}$ *are consistent estimators with the following rates:*

$$\|\widehat{\pi}^{(k)}(\mathbf{X}_i) - \pi^{(k)}(\mathbf{X}_i)\| \cdot \|\widehat{\mu}_a^{(k)}(\mathbf{X}_i) - \mu_a^{(k)}(\mathbf{X}_i)\| = o_p(n^{-1/2}), \tag{4a}$$

$$\|\widehat{w}^{(k)}(\mathbf{V}_i) - w^{(k)}(\mathbf{V}_i)\| \cdot \|\widehat{\rho}_a^{(k)}(\mathbf{V}_i) - \rho_a^{(k)}(\mathbf{V}_i)\| = o_p(n^{-1/2}), \text{ and} \tag{4b}$$

$$\|\widehat{\rho}_a^{(k)}(\mathbf{V}_i) - \rho_a^{(k)}(\mathbf{V}_i)\|^2 = o_p(n^{-1/2}). \tag{4c}$$

Then, $\sqrt{n}\left\{\widehat{\theta}_{\mathrm{EIF,a}}(\Gamma_a) - \theta_a(\Gamma_a)\right\} \to_d N\left(0, \sigma^2_{\mathrm{EIF,a}}(\Gamma_a)\right)$ *where* $\sigma^2_{\mathrm{EIF,a}}(\Gamma_a) = \mathbb{E}[\{\mathrm{EIF}(\mathbf{O}_i, \theta_a(\Gamma_a))\}^2]$.

EIF-Based Estimation

Theorem 2 (EIF-Based Estimation, Continued)

(iii) [Consistent Estimator of Standard Error] Suppose the same assumptions in (ii) hold. Then, $\widehat{\sigma}^2_{\mathsf{EIF,a}}(\Gamma_a) \rightarrow_p \sigma^2_{\mathsf{EIF,a}}(\Gamma_a)$*, where* $\widehat{\sigma}^2_{\mathsf{EIF,a}}(\Gamma_a) = K^{-1} \sum_{k=1}^{K} \frac{1}{|\mathcal{I}^{(k)}|} \sum_{i \in \mathcal{I}^{(k)}} \Big\{ \widehat{\mathsf{EIF}}^{(k)}(\mathbf{O}_i, \widehat{\theta}_{\mathsf{EIF,a}}(\Gamma_a)) \Big\}^2$ *and* $\widehat{\mathsf{EIF}}^{(k)}(\mathbf{O}_i, \widehat{\theta}_{\mathsf{EIF,a}}(\Gamma_a))$ *is the empirical counterpart of* $\mathsf{EIF}^{(k)}(\mathbf{O}_i, \widehat{\theta}_{\mathsf{EIF,a}}(\Gamma_a))$ *with plug-in estimates of the nuisance parameters* $\widehat{\pi}^{(k)}$, $\widehat{\rho}_a^{(k)}$, $\widehat{w}^{(k)}$, *and* $\widehat{\mu}_a^{(k)}$.

Data Analysis: Registered Voters in Pennsylvania (PA)

- Target population: 4,880,730 registered voters (as of April 15, 2024) from 67 counties in PA
- **V**: age group, gender, party, an incomplete voting history
- **X** : **V** + race and missing voting history
- We performed a county-by-county analysis and a subgroup analysis for groups defined through gender, urbanicity, and education attainment in the neighborhood
- For sensitivity parameters, we use $0.951 \leq \Gamma_a \leq 1.051$.
- For inference, we use the simple plug-in estimator with bootstrap

Voter Demographics Between Source and Target Population

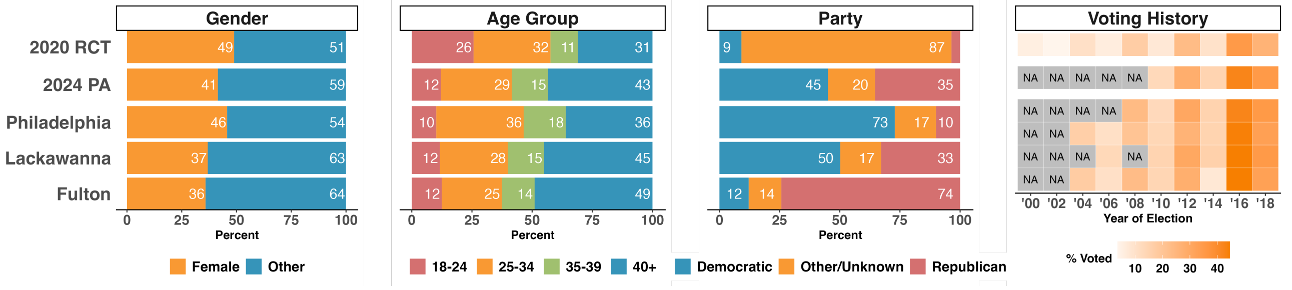
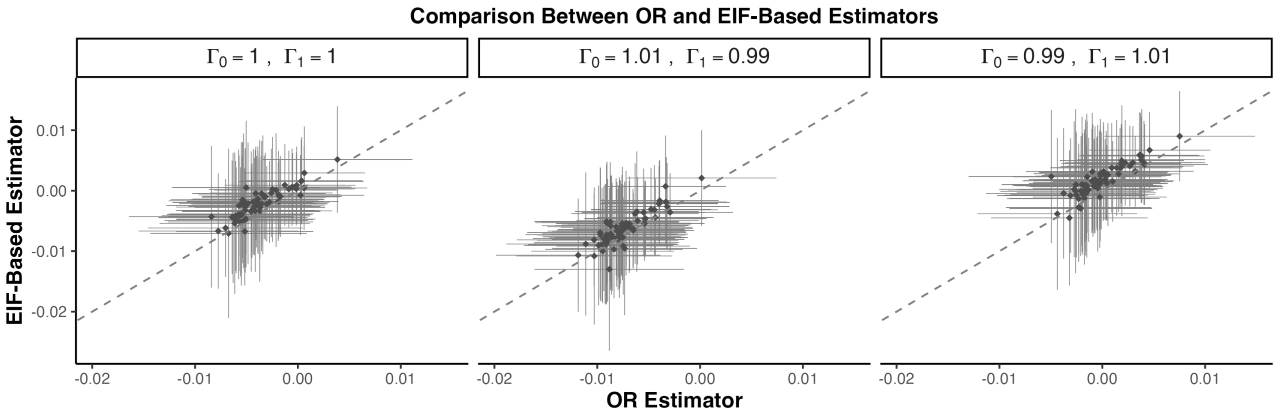


Figure. Registered voter demographics.

Comparison Between Estimators

Comparison between the OR and EIF-based estimators for estimating ad effects for every county in PA.



Data Analysis: Ad Effect In Subgroups

- We consider 20 subgroups defined by a three-way interaction between
  - gender (female & male)
  - urbanicity of the census tract (rural & urban, obtained from U.S. Census)
  - % of Bachelor's degree or higher in the same zip-code area (five levels, obtained from U.S. Census)