Paper review:

Transfer Learning Between U.S. Presidential Elections: How Should We Learn From A 2020 Ad Campaign To Inform 2024 Ad Campaigns?

Xinran Miao, Jiwei Zhao, and Hyunseung Kang (2024)

Suehyun Kim

2 June 2025

Causal Inference Lab.
Seoul National University

Table of contents

Table of contents

Introduction

The aim of this paper is to answer the following empirical question:

> Using RCT data from the 2020 U.S. presidential election,
> how can we learn about the *causal effect of negative digital ads against Trump on voter turnout* in the 2024 presidential election?

- **Methodology**: Transfer learning with sensitivity analysis
- **Practical guideline**: Calibration of sensitivity parameters
- **Results**: Data analysis & interpretation, subgroup analysis

Transfer learning

Transfer learning is a machine learning paradigm that aims to leverage knowledge from one domain (source) to improve performance in another domain (target).

- Generalizability/Transportability
- Enhancing model performance when train and test distributions differ
- Reduces data and computation costs by reusing pre-trained models

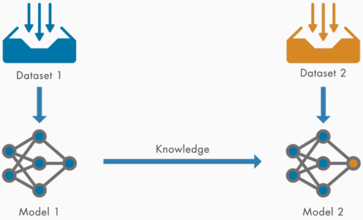


Figure reference: https://www.mathworks.com/discovery/transfer-learning.html

Transfer learning

Different types of dataset shifts are possible:

Shift Type	Explanation
Covariate shift	$P_{tr}(x) \neq P_{te}(x)$, but $P_{tr}(y \mid x) = P_{te}(y \mid x)$
Concept drift	The mechanisms $P_{tr}(y \mid x)$ and $P_{te}(y \mid x)$ differ
Prior probability shift	Change in $P(y)$

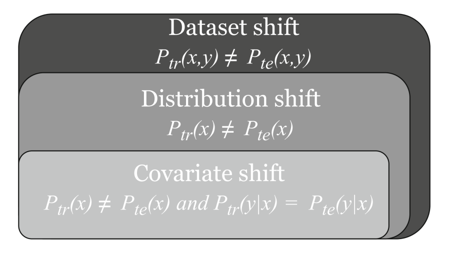


Figure reference: Bayram and Ahmed (2023)

Transfer learning and causal inference

In causal inference, the concept of transfer learning can be applied to generalization/transportation from an RCT to an observational study.

RCT (source)	**Observational Study** (target)
Internal validity	External validity
More control over covariates	Unobserved confounders
Expensive and time-consuming	Cheaper and faster, abundant data

Question. $\mathbb{E}[Y(1) - Y(0) \mid \text{source}, X] = \mathbb{E}[Y(1) - Y(0) \mid \text{target}, X]$?

Answer 1. Only if transportability holds, i.e.

$$P_{source}(Y(a) \mid X) = P_{target}(Y(a) \mid X)$$

which we cannot guarantee in practice. (cf. CATE estimation)

Transfer learning and causal inference

In causal inference, the concept of transfer learning can be applied to generalization/transportation from an RCT to an observational study.

RCT (source)	**Observational Study** (target)
Internal validity	External validity
More control over covariates	Unobserved confounders
Expensive and time-consuming	Cheaper and faster, abundant data

Question. $\mathbb{E}[Y(1) - Y(0) \mid \text{source}, X] = \mathbb{E}[Y(1) - Y(0) \mid \text{target}, X]$?

Answer 2. However, it is sensible to assume

$$P_{source}(Y(a) \mid X, U) = P_{target}(Y(a) \mid X, U).$$

This is where **sensitivity analysis** comes into play.

Main empirical question

2020 Digital Ad Campaign (Aggarwal et al. 2023):
How would online ads against Trump affect voter turnout in five battleground states: AZ, MI, NC, PA, and WI?

- Large-scale RCT with 1,999,282 registered voters
  - Treatment: An average of 754 negative ads by Acronym
  - Outcome: Voter turnout in 2020 U.S. election
- Total cost: \$8.9 million
- Results: $\widehat{\text{ATE}} = -0.06\%$, $\widehat{\text{SE}} = 0.12\%\text{p}$

2024 Rematch: Would negative ads remain ineffective in 2024 in Pennsylvania, a key "tipping point" state?

Main empirical question

However, transportability may not hold.

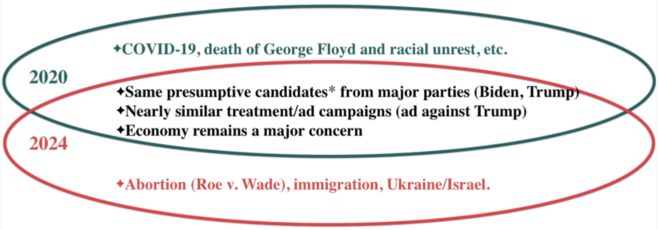


Figure reference: X. Miao's slides, Pew Research Center

- **Source**: 2020 RCT data from Aggarwal et al. (2023)
- **Target**: 2024 voters in Pennsylvania
  - 2024 covariates $\subset$ 2020 covariates
- **Estimand**: Ad effect (ATE) in the target population

Table of contents

Setup: Observed data

For $i \in \mathcal{I}_s = \{1, \ldots, n_s\}$ and $i \in \mathcal{I}_t = \{n_s + 1, \ldots, n_s + n_t = n\}$,

we observe i.i.d. samples of the following variables:

- $\mathbf{X}_i \in \mathcal{X}$: Covariates in the source population
- $\mathbf{V}_i \in \mathcal{V} \subseteq \mathcal{X}$: Shared covariates in the source and target populations
- A_i: Binary treatment indicator ($A_i = 1$ if assigned to ad campaign)
- Y_i: Binary outcome ($Y_i = 1$ if voted)
- S_i: Binary source indicator ($S_i = 1$ if unit is from source)

Source data: $\{\mathbf{O}_i = (\mathbf{X}_i, A_i, Y_i, S_i = 1),\ i \in \mathcal{I}_s\}$

Target data: $\{\mathbf{O}_i = (\mathbf{V}_i, S_i = 0),\ i \in \mathcal{I}_t\}$

Setup: Causal estimands and nuisance functions

Target estimand (θ): ATE in the target population (TATE)

$$\theta = \theta_1 - \theta_0 = \mathbb{E}[Y_i^{(1)} - Y_i^{(0)} \mid S_i = 0]$$

- $Y_i^{(a)}$: Counterfactual outcomes
- $\theta_a = \mathbb{E}[Y_i^{(a)} \mid S_i = 0]$

Setup: Causal estimands and nuisance functions

Nuisance functions:

- Propensity score in the *source population*:

$$\pi(\mathbf{x}) = P(A_i = 1 \mid \mathbf{X}_i = \mathbf{x}, S_i = 1)$$

- The outcome models in the *source population* for each level of treatment:
    - With all covariates $\mathbf{X}_i$: $\mu_a(\mathbf{x}) = \mathbb{E}[Y_i \mid \mathbf{X}_i = \mathbf{x}, A_i = a, S_i = 1]$
    - With shared covariates $\mathbf{V}_i$: $\rho_a(\mathbf{v}) = \mathbb{E}[\mu_a(\mathbf{X}_i) \mid \mathbf{V}_i = \mathbf{v}, S_i = 1]$
- The ratio of probability densities of $\mathbf{V}_i$ *between the two populations*:

$$w(\mathbf{v}) = \frac{p_{\mathbf{V_i}|\mathbf{S_i=0}}(\mathbf{v} \mid S_i = 0)}{p_{\mathbf{V_i}|\mathbf{S_i=1}}(\mathbf{v} \mid S_i = 1)} = \frac{\text{target density of } \mathbf{v}}{\text{source density of } \mathbf{v}}$$

Causal identification of TATE

To identify the TATE, we require the following assumptions:

Assumption 1 (Identification of the ATE in the Source Population)

1. (SUTVA in the source population) $Y_i = Y_i^{(A_i)}$ if $S_i = 1$.
2. (Strong ignorability) $Y_i^{(1)}, Y_i^{(0)} \perp\!\!\!\perp A_i \mid \mathbf{X}_i, S_i = 1$ and $0 < \pi(\mathbf{x}) < 1$.

Assumption 2 (Positivity of S_i)

$P(S_i = 1 \mid \mathbf{V}_i = \mathbf{v}) > 0$ for $\mathbf{v} \in \mathcal{V}$; $P(S_i = 0) > 0$.

Assumption 3 (Transportability)

$Y_i^{(1)}, Y_i^{(0)} \perp\!\!\!\perp S_i \mid \mathbf{V}_i$.

Causal identification of TATE

Under Assumptions 1-3, the TATE can be identified (Zeng et al., 2023):

$$\theta = \mathbb{E}[\, \mathbb{E}[\mu_1(\mathbf{X}_i) - \mu_0(\mathbf{X}_i) \mid \mathbf{V}_i, S_i = 1] \mid S_i = 0]. \tag{1}$$

However, the *transportability assumption* requires caution.

- Assumption 1 (identification of ATE in the source population) is satisfied by the RCT design.
- Assumption 2 (positivity of S_i) can be checked with data.
- Assumption 3 (transportability) depends on counterfactual quantities, thus not verifiable. $\rightarrow$ *Sensitivity analysis!*

Table of contents

Sensitivity analysis of transportability

Suppose transportability no longer holds, even after conditioning on $\mathbf{V}_i$.
We measure its violation by the **sensitivity parameter** $\Gamma_a \in (0, \infty)$:

$$\Gamma_a = \frac{\text{ODD}_a(\mathbf{v}, 0)}{\text{ODD}_a(\mathbf{v}, 1)}, \quad \text{ODD}_a(\mathbf{v}, s) = \frac{P(Y_i^{(a)} = 1 \mid \mathbf{V}_i = \mathbf{v}, S_i = s)}{P(Y_i^{(a)} = 0 \mid \mathbf{V}_i = \mathbf{v}, S_i = s)}. \quad (2)$$

- In words, Γ_a is the *odds ratio of counterfactual outcomes*.
- When $\Gamma_a = 1$, transportability holds; as Γ_a moves away from 1, the violation increases.
- (Example) $\Gamma_1 = 1.05$: The counterfactual odd of voting in 2024 is 1.05 times that in 2020 when a registered voter, possibly contrary to fact, is exposed to the ad campaign.

Sensitivity analysis of transportability

For a given Γ_a, the TATE can be identified via the following lemma:

Lemma (Identification of TATE Under Sensitivity Model)

Suppose Assumptions 1-2 hold. For a given Γ_a, the expected counterfactual outcome under treatment level $a \in \{0, 1\}$ is

$$\mathbb{E}[Y_i^{(a)} \mid S_i = 0] = \mathbb{E}\left[\frac{\Gamma_a \rho_a(\mathbf{V}_i)}{\Gamma_a \rho_a(\mathbf{V}_i) + 1 - \rho_a(\mathbf{V}_i)} \mid S_i = 0\right] = \theta_a(\Gamma_a). \quad (3)$$

Denote the TATE by $\theta_1(\Gamma_1) - \theta_0(\Gamma_0)$ to highlight the inclusion of sensitivity analysis.

Remarks.

(i) This model was first proposed by Robins et al. (2000), and later called an exponential tilting model:

$$p_{Y^{(a)}|\mathbf{V},S=0}(y_a \mid \mathbf{v}, S_i = 0) \propto \exp(\gamma_a y_a) \cdot p_{Y^{(a)}|\mathbf{V},S=1}(y_a \mid \mathbf{v}, S_i = 1).$$

(ii) The model has its advantages in *(a) positing no testable assumptions on the data*, *(b) making statistical inference tractable*, and *(c) allowing for a simple, odds ratio representation*.

(iii) The model can be extended to more general settings:

- To handle continuous outcomes,
- To allow Γ_a to depend on $\mathbf{V}_i$ and $Y_i^{(a)}$, at the expense of more sensitivity parameters.

Table of contents

Inference pipeline

The authors propose two different approaches to estimate the TATE and conduct inference, assuming the covariates $\mathbf{X}_i$ are discrete:

(i) **Plug-in estimator**: Outcome regression + percentile bootstrap.
   - Nonparametric and efficient; authors' recommendation

(ii) **EIF-based estimator**: Messy and not doubly robust, but allows for broader applications.

Plug-in estimator

Recall the identification formula:

$$\theta_a(\Gamma_a) = \mathbb{E}\left[\frac{\Gamma_a \rho_a(\mathbf{V}_i)}{\Gamma_a \rho_a(\mathbf{V}_i) + 1 - \rho_a(\mathbf{V}_i)} \mid S_i = 0\right].$$

A natural estimator of $\theta_a(\Gamma_a)$ would be a plug-in estimator, which is referred to as the *outcome regression (OR) estimator*:

$$\hat{\theta}_{OR,a}(\Gamma_a) = \frac{1}{n_t} \sum_{i \in \mathcal{I}_t} \frac{\Gamma_a \hat{\rho}_a(\mathbf{V}_i)}{\Gamma_a \hat{\rho}_a(\mathbf{V}_i) + 1 - \hat{\rho}_a(\mathbf{V}_i)}. \tag{4}$$

Plug-in estimator

For inference, the authors recommend a *nonparametric, percentile bootstrap CI*

$$\widehat{\mathrm{CI}}_{OR,a}(\Gamma_a; 1-\alpha) = \left[\hat{L}_a(\Gamma_a; 1-\alpha), \hat{U}_a(\Gamma_a; 1-\alpha)\right],$$

where $\hat{L}_a(\Gamma_a; 1-\alpha), \hat{U}_a(\Gamma_a; 1-\alpha)$ are quantiles of the bootstrapped estimates of $\hat{\theta}_{OR,a}(\Gamma_a)$.

Plug-in estimator

The following theorem shows that under regularity conditions, the OR estimator is consistent and the percentile bootstrap leads to a valid CI.

Theorem (OR estimator and bootstrapped CI)

Suppose Assumptions 1-2 hold and $\theta_a(\Gamma_a) \in \Theta$, where Θ is open and compact. Also, suppose $\rho_a(\mathbf{v}; \eta_a)$ is twice differentiable in η_a.

If $\hat{\eta}_a$ is an asymptotically linear estimate of η_a and $n_s \asymp n_t$, then $\hat{\theta}_{OR,a}(\Gamma_a) \xrightarrow{p} \theta_a(\Gamma_a)$.

Furthermore, under appropriate regularity conditions, the bootstrap interval satisfies $P(\theta_a(\Gamma_a) \in \widehat{\mathrm{CI}}_{OR,a}(\Gamma_a; 1-\alpha)) \to 1-\alpha$.

EIF-based estimator

Recap. Efficiency theory

1. RAL (Regular and Asymptotically Linear) estimators

- An estimator $\hat{\theta}$ of θ is *asymptotically linear* if it can be expressed as

$$\sqrt{n}(\hat{\theta} - \theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(\mathbf{O}_i) + o_p(1),$$

  where ψ is the (unique) influence function.
- An estimator $\hat{\theta}$ of θ is *regular* if the limiting distribution of $\sqrt{n}(\hat{\theta} - \theta)$ does not depend on the local data generating process.
- **RAL** = asymptotically linear + regular
  - Regularity means no weirdness
  - Asymptotic linearity guarantees efficiency (Hájek-Le Cam)

EIF-based estimator

2. Building estimators using the EIF

- **Step 1.** Figure out the set of all possible influence functions

$$\mathbf{\Psi} = \{\psi : \psi \text{ is an IF for some RAL estimator}\}.$$

- **Step 2.** Find ψ corresponding to the estimator with the smallest variance (i.e. the EIF).
- The resulting estimator exhibits asymptotic normality & achieves semiparametric efficiency.

EIF-based estimator

Instead of the plug-in estimator, we can construct an *EIF-based estimator* of $\theta_a(\Gamma_a)$. However, it is more complex and requires estimation of nuisance functions.

Theorem (EIF of $\theta_a(\Gamma_a)$)

Under Assumptions 1-2, the EIF of $\theta_a(\Gamma_a)$ is

$$\begin{aligned}\mathrm{EIF}(\mathbf{O}_i, \theta_a(\Gamma_a)) = & \frac{S_i w(\mathbf{V}_i)}{\mathbb{P}(S_i = 1)} \cdot \frac{\Gamma_a}{[\Gamma_a \rho_a(\mathbf{V}_i) + 1 - \rho_a(\mathbf{V}_i)]^2} \\ & \cdot \left[\left\{ \frac{A_i}{\pi(\mathbf{X}_i)} + \frac{1 - A_i}{1 - \pi(\mathbf{X}_i)} \right\} (Y_i - \mu_a(\mathbf{X}_i)) + \mu_a(\mathbf{X}_i) - \rho_a(\mathbf{V}_i) \right] \\ & + \frac{1 - S_i}{\mathbb{P}(S_i = 0)} \left[\frac{\Gamma_a \rho_a(\mathbf{V}_i)}{\Gamma_a \rho_a(\mathbf{V}_i) + 1 - \rho_a(\mathbf{V}_i)} - \theta_a(\Gamma_a) \right].\end{aligned}$$

EIF-based estimator

Using the EIF, we can construct an estimator with *cross-fitting*. In the k-th fold, evaluate the uncentered EIF at the observed data:

$$\hat{\theta}^{(k)}_{\mathrm{EIF},a}(\Gamma_a) = \frac{1}{|\mathcal{I}_s^{(k)}|} \sum_{i \in \mathcal{I}_s^{(k)}} \frac{\Gamma_a \hat{w}^{(k)}(\mathbf{V}_i)}{[\Gamma_a \hat{\rho}_a^{(k)}(\mathbf{V}_i) + 1 - \hat{\rho}_a^{(k)}(\mathbf{V}_i)]^2}$$

$$\times \left[\left\{ \frac{A_i}{\hat{\pi}^{(k)}(\mathbf{X}_i)} + \frac{1 - A_i}{1 - \hat{\pi}^{(k)}(\mathbf{X}_i)} \right\} \left(Y_i - \hat{\mu}_a^{(k)}(\mathbf{X}_i) \right) + \hat{\mu}_a^{(k)}(\mathbf{X}_i) - \hat{\rho}_a^{(k)}(\mathbf{V}_i) \right]$$

$$+ \frac{1}{|\mathcal{I}_t^{(k)}|} \sum_{i \in \mathcal{I}_t^{(k)}} \frac{\Gamma_a \hat{\rho}_a^{(k)}(\mathbf{V}_i)}{\Gamma_a \hat{\rho}_a^{(k)}(\mathbf{V}_i) + 1 - \hat{\rho}_a^{(k)}(\mathbf{V}_i)}.$$

The EIF-based estimator is obtained by averaging over all folds:

$$\hat{\theta}_{EIF,a}(\Gamma_a) = \frac{1}{K} \sum_{k=1}^{K} \hat{\theta}^{(k)}_{\mathrm{EIF},a}(\Gamma_a).$$

EIF-based estimator

The following theorem shows that under appropriate conditions, $\hat{\theta}_{EIF,a}(\Gamma_a)$ is consistent, asymptotically normal, and semiparametrically efficient.

Theorem (Theoretical properties of the EIF-based estimator)

Suppose Assumptions 1-2 hold and there exist c, $C > 0$ such that $c < \hat{\pi}^{(k)}(\mathbf{x}), \hat{w}^{(k)}(\mathbf{v}) < C$ and $\hat{\rho}_a^{(k)}(\mathbf{v}) \in [0,1]$ for $\mathbf{v} \in \mathcal{V}$ and $\mathbf{x} \in \mathcal{X}$.

(i) (Conditional double robustness) Suppose $\hat{\rho}_a^{(k)}$ is a consistent estimator of $\rho_a^{(k)}$. Then, $\hat{\theta}_{EIF,a}(\Gamma_a) \xrightarrow{p} \theta_a(\Gamma_a)$ if

$$\left\|\hat{\pi}^{(k)}(\mathbf{X}_i) - \pi^{(k)}(\mathbf{X}_i)\right\| \cdot \left\|\hat{\mu}_a^{(k)}(\mathbf{X}_i) - \mu_a^{(k)}(\mathbf{X}_i)\right\| = o_p(1). \tag{5}$$

EIF-based estimator

Theorem (continued)

(ii) (Asymptotic normality and semiparametric efficiency) Suppose $\hat{\rho}_a^{(k)}$, $\hat{\mu}_a^{(k)}$, $\hat{w}^{(k)}$ and $\hat{\pi}^{(k)}$ are consistent estimators with the following rates:

$$\left\|\hat{\pi}^{(k)}(\mathbf{X}_i) - \pi^{(k)}(\mathbf{X}_i)\right\| \cdot \left\|\hat{\mu}_a^{(k)}(\mathbf{X}_i) - \mu_a^{(k)}(\mathbf{X}_i)\right\| = o_p(n^{-1/2}), \tag{6}$$

$$\left\|\hat{w}^{(k)}(\mathbf{V}_i) - w^{(k)}(\mathbf{V}_i)\right\| \cdot \left\|\hat{\rho}_a^{(k)}(\mathbf{V}_i) - \rho_a^{(k)}(\mathbf{V}_i)\right\| = o_p(n^{-1/2}), \tag{7}$$

$$\left\|\hat{\rho}_a^{(k)}(\mathbf{V}_i) - \rho_a^{(k)}(\mathbf{V}_i)\right\|^2 = o_p(n^{-1/2}). \tag{8}$$

Then,

$$\sqrt{n}\left\{\hat{\theta}_{\mathrm{EIF},a}(\Gamma_a) - \theta_a(\Gamma_a)\right\} \xrightarrow{d} N\left(0, \sigma^2_{\mathrm{EIF},a}(\Gamma_a)\right),$$

where $\sigma^2_{\mathrm{EIF},a}(\Gamma_a) = \mathbb{E}\left[\{\mathrm{EIF}(\mathbf{O}_i, \theta_a(\Gamma_a))\}^2\right]$.

EIF-based estimator

Theorem (continued)

(iii) (Consistent estimator of standard error) Suppose the same assumptions as in (ii) hold.

Then, $\hat{\sigma}^2_{\mathrm{EIF},a}(\Gamma_a) \xrightarrow{p} \sigma^2_{\mathrm{EIF},a}(\Gamma_a)$, where

$$\hat{\sigma}^2_{\mathrm{EIF},a}(\Gamma_a) = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{|\mathcal{I}^{(k)}|}\sum_{i\in\mathcal{I}^{(k)}}\left\{\widehat{\mathrm{EIF}}^{(k)}\left(\mathbf{O}_i, \hat{\theta}_{\mathrm{EIF},a}(\Gamma_a)\right)\right\}^2,$$

and $\widehat{\mathrm{EIF}}^{(k)}\left(\mathbf{O}_i, \hat{\theta}_{\mathrm{EIF},a}(\Gamma_a)\right)$ is the empirical version of $\mathrm{EIF}^{(k)}\left(\mathbf{O}_i, \hat{\theta}_{\mathrm{EIF},a}(\Gamma_a)\right)$ with plug-in estimates of the nuisance parameters $\hat{\rho}_a^{(k)}$, $\hat{\mu}_a^{(k)}$, $\hat{w}^{(k)}$ and $\hat{\pi}^{(k)}$.

EIF-based estimator

Remarks.

- Part (i): $\hat{\theta}_{\mathsf{EIF},a}(\Gamma_a)$ is *conditionally doubly robust* in that if $\hat{\rho}_a^{(k)}$ *is consistent*, $\hat{\theta}_{\mathsf{EIF},a}(\Gamma_a)$ is consistent when either $\hat{\pi}^{(k)}$ or $\hat{\mu}_a^{(k)}$, but not necessarily both, are consistent.
- Part (ii): If nuisance functions are estimated consistently at sufficiently fast rates, $\hat{\theta}_{\mathsf{EIF},a}(\Gamma_a)$ is *asymptotically normal and achieves semiparametric efficiency*.
  - Equation (8) can be viewed as the *cost of violating transportability*.
- Part (iii): The normality-based CI $\hat{\theta}_{\mathsf{EIF},a}(\Gamma_a) \pm z_{1-\alpha/2}\sqrt{\hat{\sigma}^2_{\mathsf{EIF},a}(\Gamma_a)}$ is asymptotically valid.

Estimation of nuisance functions

In the OR and EIF-based estimators, we need to estimate the following nuisance functions:

(i) *Classical nuisance functions* $(\pi(\mathbf{x}), \mu_a(\mathbf{x}))$: Analyst's choice of classification and regression models.

(ii) *Projection of outcome model* $\rho_a(\mathbf{v})$:

$$\rho_a(\mathbf{v}) = \mathbb{E}[\mu_a(\mathbf{X}_i) \mid \mathbf{V}_i = \mathbf{v}, S_i = 1]$$

$\rightarrow$ Regress the predicted outcome $\hat{\mu}_a(\mathbf{X}_i)$ on shared covariates $\mathbf{V}_i$.

(iii) *Density ratio* $w(\mathbf{v})$: Entropy balancing

$$\underset{w_i}{\text{argmin}} \sum_{i \in \mathcal{I}_s} w_i \log(w_i), \quad \text{s.t.} \quad \frac{1}{n_s} \sum_{i \in \mathcal{I}_s} w_i \mathbf{V}_i = \frac{1}{n_t} \sum_{i \in \mathcal{I}_t} \mathbf{V}_i.$$

Table of contents

Motivation for calibration

Specifying a *reasonable reference set of sensitivity parameters* have been a long-standing question in the literature, which is referred to as **calibration**.

Before we proceed, we formally define the notion of *sensitivity to transportability*.

Definition (Sensitivity to Transportability)

Consider a significance level α and a set $\mathcal{C} \subset \mathbb{R}^+ \times \mathbb{R}^+ \neq \{(1,1)\}$. The TATE is sensitive to transportability in $\mathcal{C}$ if there exists $(\Gamma_0, \Gamma_1) \in \mathcal{C}$ such that either of the following holds:

(i) *from significant to insignificant:* $0 \notin \widehat{\mathrm{CI}}(1,1;1-\alpha)$ and $0 \in \widehat{\mathrm{CI}}(\Gamma_0, \Gamma_1; 1-\alpha)$;

(ii) *from insignificant to significant:* $0 \in \widehat{\mathrm{CI}}(1,1;1-\alpha)$ and $0 \notin \widehat{\mathrm{CI}}(\Gamma_0, \Gamma_1; 1-\alpha)$.

Data-driven calibration procedure

Step 1. Partition the source data (Aggarwal's 2020 RCT) into rust belt states ($\mathcal{I}_{s_1}$; PA, MI, WI) and sun belt states ($\mathcal{I}_{s_2}$; AZ, NC).

Step 2. Temporarily treat $\mathcal{I}_{s_2}$ (sun belt states) as the proxy target population, and construct two CIs for its TATE:

- (Transfer learning approach) Use the proposed method treating $\mathcal{I}_{s_1}$ as the source population; $\widehat{\text{CI}}_{s_1 \to s_2}(\Gamma_0, \Gamma_1; 1-\alpha)$.
- (Standard approach) Using data from $\mathcal{I}_{s_2}$ only, estimate the TATE and construct a CI; $\widehat{\text{CI}}_{s_2}(1-\alpha)$.

Step 3. Keep the values of (Γ_0, Γ_1) where the two CIs overlap, i.e. $\mathcal{C}_1 = \{(\Gamma_0, \Gamma_1) : \widehat{\text{CI}}_{s_1 \to s_2}(\Gamma_0, \Gamma_1; 1-\alpha) \cap \widehat{\text{CI}}_{s_2}(1-\alpha) \neq \emptyset\}$.

Step 4. Repeat Steps 2-3 with roles of $\mathcal{I}_{s_1}$ and $\mathcal{I}_{s_2}$ reversed to obtain $\mathcal{C}_2$.

Step 5. The final calibration set is given by $\mathcal{C} = \mathcal{C}_1 \cap \mathcal{C}_2$.

Rationale behind the calibration procedure

- In **Step 1**, dissimilar partitions of the source sample are created in order to quantify the possible magnitude of unobservable differences.
  - Other partitions are possible, but some may be more useful/interpretable.
- The proposed calibration procedure is a data-driven approach to understand and interpret the sensitivity parameters based on the observed data.
  - It does *not* represent the true unmeasurable differences between the original source and target populations.
  - For instance, if the TATE is sensitive to the calibrated set $\mathcal{C}$, it suggests that *the unmeasured differences that are as large as those between the sun belt states and the rust belt states in the 2020 election can overturn the conclusion about the TATE in 2024*.

Table of contents

Setup

Data collection:

- **Source**: 2020 RCT data (PA, MI, WI, AZ, NC; $n_s = 1,999,282$)
- **Target**: 2024 registered voters in PA ($n_t = 4,880,729$)
- Shared covariates $\mathbf{V}_i$: Gender, age groups, party, subset of voting history
- Covariates $\mathbf{X}_i$: $\mathbf{V}_i$ & race, richer voting history
- All covariates are discrete

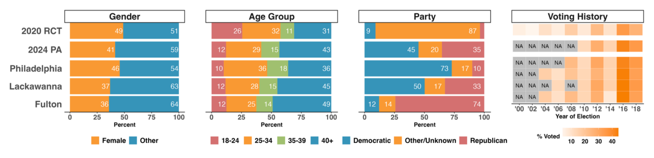


Figure 1: Covariate distributions in the source and target populations.

Setup

Analysis pipeline:

(i) **Estimation and inference:** OR estimator + percentile bootstrap CI
  - Nuisance functions: $\rho_a(\mathbf{v})$ (regression), $\mu_a(\mathbf{x})$ (IPW)

(ii) **Sensitivity analysis and calibration:**
  - Under transportability,
  - Under prespecified (Γ_0, Γ_1),
  - With calibrated set $\mathcal{C}$.

(iii) **Analysis:** County-by-county & subgroup analysis.

Ad effect by counties

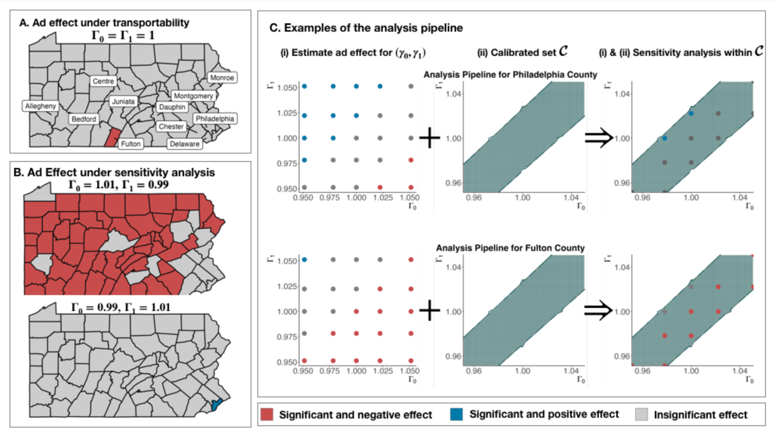


Figure 2: County-by-county analysis for 2024 PA voters.
Panel A: Under transportability. Panel B: Under prespecified (Γ_0, Γ_1).
Panel C: Analysis pipeline for Philadelphia county and Fulton county.

Ad effect by counties

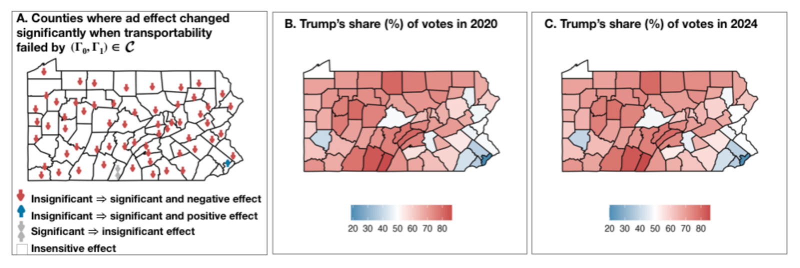


Figure 3: County-by-county analysis results.
Panel A: Results of sensitivity analysis under the calibrated set $\mathcal{C}$.
Panel B & C: Trump's vote share in 2020 and 2024.

Ad effect by counties

Key findings:

Sensitive for a positive effect	1 (Philadelphia)
Sensitive for a negative effect	59
Sensitive for an insignificant effect	1 (Fulton)
Insensitive	6
Total	67

- *"Small effects are sensitive to small [unmeasured] biases,"* (Rosenbaum, 2010).
- Negative ads against Trump will generally decrease voter turnout in 2024.
- The direction of the ad effect on voter turnout does not equate to whether the adds will help or hurt Trump, but we can make certain conjectures (Figure 3).

Ad effect by subgroups

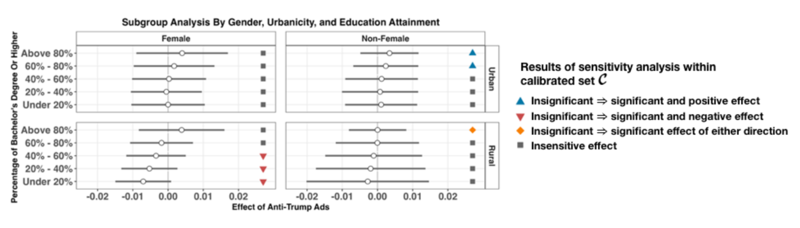


Figure 4: Subgroup analysis by gender, urbanicity, and education attainment.

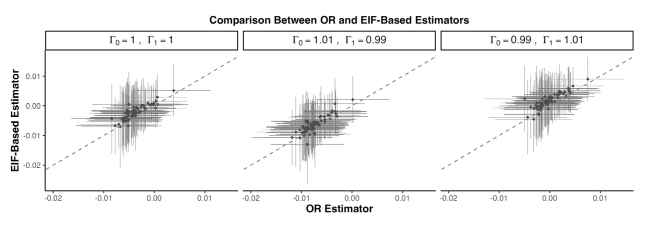
Key findings:

- None of the estimates are statistically significant.
- Urban areas / Non-female / High educational attainment: Positive ad effects, sensitive for a positive effect.
- Rural areas / Female / Moderate - Low educational attainment: Negative ad effects, sensitive for a negative effect.

Remarks and robustness checks

(i) OR vs EIF-based estimator

- Analyses based on the OR estimator and the EIF-based estimator yield similar results, although not identical.
- The widths of the CIs does not uniformly dominate one another.



(ii) Violation of i.i.d assumption

- The statistical framework assumes that the target and source samples are independent and there are no overlapping voters beetween the two samples.
- However, this assumption is implausible and unverifiable.
- Original vs restricted analysis (source with NC and AZ only):
  - Restricted analysis: Loss of sample size, less statistically significant results.
  - Original analysis: Transportability is more plausible *(recommended)*.

Table of contents

Discussion

Contributions of the paper:

(i) Practical & statistically valid framework for transportability of causal effects.

(ii) Addressing violation of transportability when $\mathcal{X} \neq \mathcal{V}$.

(iii) Data-driven calibration procedure.

Future work & disclaimers:

(i) Relaxing assumptions on the target population: Infinite to finite.

(ii) Accounting for withdrawal of Biden in July 2024.

Reference

Bayram, F., & Ahmed, B. S. (2023). *A domain-region based evaluation of ML performance robustness to covariate shift*. Neural Computing and Applications.

Miao, X. (2024). *Transfer Learning Between U.S. Presidential Elections: How Should We Learn From A 2020 Ad Campaign To Inform 2024 Ad Campaigns?* [ACIC 2024 Slides].

Schuler, A., & van der Laan, M. *Introduction to Modern Causal Inference*. Retrieved June 1, 2025, from `https://alejandroschuler.github.io/mci/index.html`

Tsiatis, A. A. (2006). *Semiparametric Theory and Missing Data*. Springer.