Adjusting for Unmeasured Confounding in Marginal Structural Models with Propensity-Score Fixed Effects

Matthew Blackwell Soichiro Yamauchi

Harvard University

Applied Statistics Workshop
February 17, 2021

Motivation

- Estimating effects of time-varying treatments
  - Many applications in epidemiology, sociology, and political science, etc
  - Presence of time-varying confounders

- Marginal structural model adjusts for time-varying confounders
  - Flexibly adjusts for potentially post-treatment confounders
  - Models treatment history $\rightsquigarrow$ Complex causal quantities

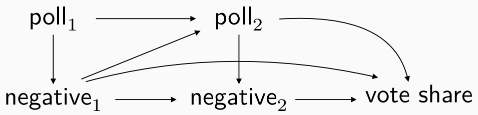
- Presence of unmeasured confounders in observational studies
  - MSM assumes away unmeasured confounding $\rightsquigarrow$ Biased causal estimates

- Propose a method based on propensity score fixed effect
  - Estimate propensity score with unit fixed effects
    $\rightsquigarrow$ Accounts for unmeasured time-invariant confounder
  - Consider an asymptotic regime where n and T grows
    $\rightsquigarrow$ Address the incidental parameter problem

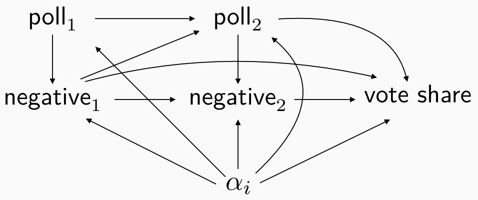
- Simulation evidence to demonstrate finite sample performances
- Application to the effect of negative ads on electoral outcomes in the US

The Impact of Negative Ads in the Election Cycle

- Effect of negative ads on the vote share (Blackwell, 2013)
- Binary treatment (negative/positive) sequence for 8 $\sim$ 40 weeks
- Outcome is observed only after the election
- Original DAG without unobserved confounder



- DAG with unobserved time-invariant confounder



$\rightsquigarrow$ District characteristics, candidate characteristics, etc.

Contributions

- Causal inference with time-varying treatments

- Linear fixed effect model addresses time-invariant unobserved confounders
  - Assumes either past treatments do not directly affect the current outcome, or treatment is unaffected by time-varying confounders (Imai and Kim, 2019; Sobel, 2012)

    $\rightsquigarrow$ Rule out dynamical aspect of time-varying treatments

- Marginal structural model allows for complex dependences
  - Sequential ignorability assumption (Robins, 1999)

    $\rightsquigarrow$ Need to assume away unobserved confounders

- Propose IPTW estimator based on propensity score fixed effects
  - Includes fixed effects in the propensity score estimation

    $\rightsquigarrow$ Allows for unobserved time-invariant confounder
  - Estimated weights are used in the marginal structural model

    $\rightsquigarrow$ Allows for complex dependences between treatments and the outcome
  - Address the incidental parameter problem (# params. grows with n) with a large-T approximation (Hahn and Newey, 2004; Fernandez-Val and Weidner, 2018)

Unit-specific Randomized Experiment: Setup

Setup

- Binary treatment: D_{it} for $t = 1, \ldots, T$ and $i = 1, \ldots, n$
- Outcome is observed at the end of the experiment: Y_i
- Consider estimating the effect of the final treatment: $Y_i(d_T)$
- Assume that the treatment is randomized within an individual

$$\mathbb{P}(D_{it} = 1 \mid \alpha_i) = \pi(\alpha_i) \equiv \pi_i$$

- Assume conditional independence: $Y_i(d_T) \perp\!\!\!\perp D_{iT} \mid \alpha_i$

Infeasible Estimator

- Consider $\tau_1 = \mathbb{E}[Y_i(1)]$ as our quantity of interest (for now)
- If we knew the true propensity score π_i for all i, we could estimate τ_1 via

$$\widetilde{\tau}_1 = \frac{\sum_{i=1}^{n} D_{iT} Y_i / \pi_i}{\sum_{i=1}^{n} D_{iT} / \pi_i}$$

- Under some regularity conditions, $\widetilde{\tau}_1$ is asymptotically normal

Unit-specific Randomized Experiment: Result

Estimator with Estimated Propensity Score

- Consider the following feasible IPTW estimator

$$\widehat{\tau}_1 = \frac{\sum_{i=1}^n D_{iT} Y_i / \widehat{\pi}_i}{\sum_{i=1}^n D_{iT} / \widehat{\pi}_i}$$

  where $\widehat{\pi}_i = \sum_{t=1}^T D_i / T$

- π_i is an incidental parameter $\rightsquigarrow$ Nonlinearly enters the estimator
- $\widehat{\tau}_1$ is not consistent under fixed-T regime (incidental parameter problem)

Proposed Strategy

- Consider a large-T regime: As $n, T \to \infty$ with $n/T \to \rho$ where $0 < \rho < \infty$
- Then, $\widehat{\tau}_1$ is consistent and asymptotically normally distributed

$$\sqrt{n}(\widehat{\tau}_1 - \tau_1) \xrightarrow{d} \mathcal{N}(0, \sigma^2_{\tau_1})$$

Unit-specific Randomized Experiment: Analysis

- How do we avoid the incidental parameter problem?
- We can re-express the feasible estimator as a solution to the following

$$\frac{1}{n}\sum_{i=1}^{n}\underbrace{\frac{D_{iT}}{\widehat{\pi}_i}(Y_i - \widehat{\tau}_1)}_{\equiv U_i(\widehat{\pi}_i, \widehat{\tau}_1)} = 0$$

- As $n, T \to \infty$ we have

$$\sqrt{n}(\widehat{\tau}_1 - \tau_1) = \underbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_i(\pi_i, \tau_1)}_{\text{Term (I)}} - \underbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\widehat{\pi}_i - \pi_i}{\pi_i} U_i(\pi_i, \tau_1)}_{\text{Term (II)}} + \text{(high-order)}$$

- Term (I) is what we get if we knew the true propensity score
- Term (II) is due to the noise in estimated propensity score
  - We show that Term (II) is $O_p(1/\sqrt{T}) \rightsquigarrow$ Vanishes as $T \to \infty$

Unit-specific Randomized Experiment: Remarks

- Time horizon T grows with the number of units n

  $\rightsquigarrow$ Reasonable approximation when we have data with large T

- $n/T \to \rho$ ensures that the cross-sectional dimension does not dominate T

  $\rightsquigarrow$ n and T are of similar order (e.g., time-series and cross-sectional data)

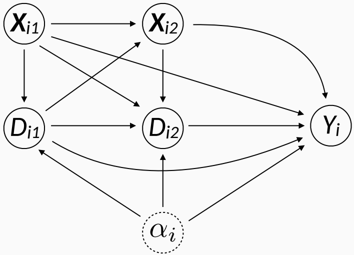
  $\rightsquigarrow$ Value of ρ affects the finite sample performance

- The estimator is scaled by $\sqrt{n}$ not by $\sqrt{nT}$

  $\rightsquigarrow$ We make inference on some finite number of outcomes in time dimension

Marginal Structural Models

- We generalize the result in previous slides to a general setting
- In addition to treatments, we have time-varying covariates $\boldsymbol{X}_{it}$



- Treatment effect defined as contrast between two treatment histories

$$\tau(\underline{d}_k, \underline{d}'_k) = \mathbb{E}\left[Y_i(\underline{d}_k) - Y_i(\underline{d}'_k)\right], \quad \underline{d}_k = (d_{T-k}, \ldots, d_T)$$

- In our example: $\mathbb{E}[Y_i(\texttt{negative}_2, \texttt{negative}_1) - Y_i(\texttt{negative}_2, \texttt{positive}_1)]$

Assumptions

- Relax assumptions regularly employed in MSM
- Unit-specific sequential ignorability: Treatment is independent of $Y_i(\underline{d}_{T-k}) \equiv Y_i(\overline{D}_{i,T-k-1}, \underline{d}_k)$ conditional on information up to time t and unit fixed effect

$$Y_i(\underline{d}_{T-k}) \perp\!\!\!\perp D_{it} \mid \overline{\boldsymbol{X}}_{it}, \overline{D}_{i,t-1}, \alpha_i$$

$\rightsquigarrow$ Allow for unobserved time-invariant confounder via α_i

- Propensity score model: Parametric model of treatment assignment

$$\mathbb{P}(D_{it} = 1 \mid \boldsymbol{V}_{it} = \boldsymbol{v}, \alpha_i) = F(\alpha_i + \boldsymbol{\beta}^\top \boldsymbol{v}) \equiv \pi_{it}(\alpha_i, \boldsymbol{\beta})$$

where $\boldsymbol{V}_{it} = (\overline{\boldsymbol{X}}_{it}, \overline{D}_{i,t-1})$ is the history up to time t

- Marginal structural model: For $\underline{d}_k = (d_{T-k}, \ldots, d_T)$ with fixed k,

$$\mathbb{E}[Y_i(\underline{d}_k)] = g(\underline{d}_k; \gamma)$$

$\rightsquigarrow$ Reduce dimensionality of $Y_i(\underline{d}_k)$

$$\tau(\underline{d}_k, \underline{d}'_k) = g(\underline{d}_k; \gamma) - g(\underline{d}'_k; \gamma)$$

The Proposed Estimator

- Propensity score: Prob. of observing a particular treatment history $\underline{d}_k$

$$W_i(\underline{d}_k; \alpha_i, \boldsymbol{\beta}) = \prod_{j=0}^{k} \left\{\pi_{i,T-j}(\alpha_i, \boldsymbol{\beta})\right\}^{d_{i,T-j}} \left\{1 - \pi_{i,T-j}(\alpha_i, \boldsymbol{\beta})\right\}^{1-d_{i,T-j}}$$

- Estimate propensity score via MLE $\rightsquigarrow$ Logistic regression with unit indicators

$$\widehat{\boldsymbol{\beta}} = \operatorname*{argmax}_{\boldsymbol{\beta}} \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \ell_{it}(\boldsymbol{\beta}, \widehat{\alpha}_i(\boldsymbol{\beta})), \quad \widehat{\alpha}_i = \operatorname*{argmax}_{\alpha} \frac{1}{T} \sum_{t=1}^{T} \ell_{it}(\boldsymbol{\beta}, \alpha)$$

- Estimate parameters in MSM by solving the estimating equation:

$$\mathbb{E}\left[\frac{h(\underline{D}_i)(Y_i - g(\underline{D}_i; \widehat{\gamma}))}{W_i(\underline{d}_k; \widehat{\alpha}_i, \widehat{\boldsymbol{\beta}})}\right] = \mathbf{0}$$

$\rightsquigarrow$ Weighted least square estimator with $1/\widehat{W}_i(\underline{d}_k)$ as weights

- **Theorem 1**: Under regularity conditions, we have

$$\sqrt{n}(\widehat{\gamma} - \gamma) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{V}_\gamma)$$

as $n, T \to \infty$ with $n/T \to \rho$.

Simulation Study

Setup

- Number of units: $n \in \{200, 500, 1000, 3000\}$
- Unit-time ratio: $n/T = \rho \in \{5, 50\}$
- Treatment assignment depends on FE, past-treatment & covariates

$$D_{it} \sim \text{Bern}(\text{expit}(\alpha_i + \phi D_{i,t-1} + \boldsymbol{\beta}^\top \mathbf{X}_{it}))$$

- Unobserved heterogeneity: $\alpha_i \sim \text{Uniform}[-a, a]$ with $a \in \{1, 2\}$
- Outcome model:

$$Y_i = \alpha_i + \underbrace{\tau_F D_{iT}}_{\text{contemporaneous effect}} + \underbrace{\tau_C \sum_{t=T-1}^{T-3} D_{it}}_{\text{cumulative effect}} + \boldsymbol{\gamma}^\top \overline{\mathbf{X}}_i + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, 1)$$

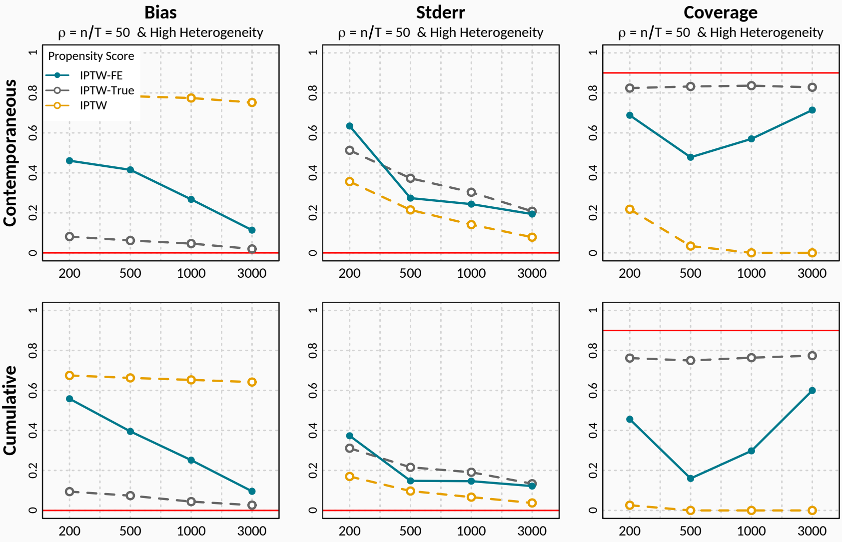
Estimators

- Weighted least square under correct specification

$$(\widehat{\tau}_F, \widehat{\tau}_C) = \text{argmin} \sum_{i=1}^{n} \widehat{W}_i \left\{ Y_i - \alpha - \tau_F D_{iT} - \tau_C \sum_{t=T-1}^{T-3} D_{it} \right\}^2$$

- (1) FE-PS, (2) PS, and (3) true PS

Results


Bias
ρ = n/T = 50 & High Heterogeneity
Stderr
ρ = n/T = 50 & High Heterogeneity
Coverage
ρ = n/T = 50 & High Heterogeneity
Contemporaneous
Cumulative
Propensity Score
IPTW-FE
IPTW-True
IPTW
0
0.2
0.4
0.6
0.8
1
200
500
1000
3000

Empirical Study

Marginal structural model

- Treatment: Democratic candidate going negative ($D_{it} = 1$) or not ($D_{it} = 0$)
- Effect of additional negative ads in the last 5 weeks

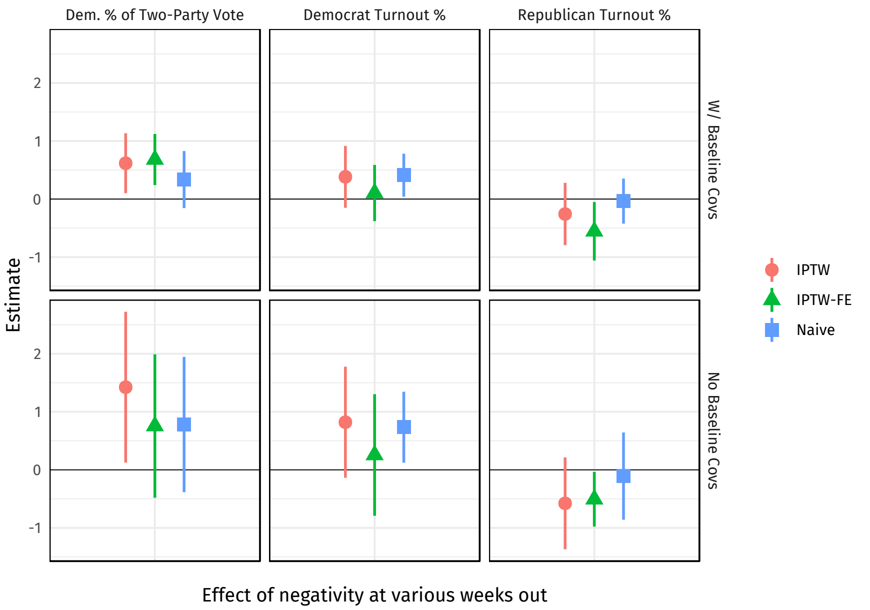
$$\mathbb{E}[Y_i(\underline{d})] = \gamma + \tau \sum_{k=0}^{4} d_{T-k}$$

- Two outcomes:
  1. Dem. two-party vote share (electoral outcome)
  2. Democrat / Republican turnout (mobilization effect)
- Focus on US Senate & Gubernatorial elections between 2000 and 2008
- $n = 201$ unique races

Covariates

- Time-varying covariate: Opinion polls, time-trend, opponent's ad, etc
- Time-invariant baseline covariates: Predicted competitiveness of the race, incumbency status, measures of challenger quality, etc.

Results



Dem. % of Two-Party Vote
Democrat Turnout %
Republican Turnout %
W/ Baseline Covs
No Baseline Covs
Estimate
2
1
0
-1
IPTW
IPTW-FE
Naive
Effect of negativity at various weeks out

Concluding Remarks

- Causal inference with time-varying treatments
  - Dynamical relationships between treatments and the outcome
  - Presence of time-varying confounders

- Existing methods either assumes away some dynamics (linear FE) or unmeasured confounding (MSM)

- Propose a method to incorporate fixed effect in propensity score estimation
  - Accounts for time-invariant unmeasured confounders
  - Consider a large-T approximation to address incidental parameter problem

Appendix

Details of Weighting Estimator

- Stabilized weights: Let $\overline{\pi}_{it} = \mathbb{P}(D_{it} = 1 \mid \overline{D}_{i,t-1})$, and

$$\widehat{W}_i = \prod_{j=0}^{k} \left(\frac{\overline{\pi}_{i,T-j}}{\widehat{\pi}_{i,T-j}} \right)^{D_{i,T-j}} \left(\frac{1 - \overline{\pi}_{i,T-j}}{1 - \widehat{\pi}_{i,T-j}} \right)^{1-D_{i,T-j}}$$

- Trimming weights: When MLE is unbounded for $\widehat{\alpha}_i$, we could either replace $\widehat{\alpha}_i$ with a constant, or drop the observations

Additional Simulation Results