Augmented Balancing Estimators of the Average Treatment Effect on the Treated in cross-sectional and panel designs

Apoorva Lal

SME, work done at Stanford

August 18, 2023

Overview

Introduction

- Explosion of methods in observational causal inference methods in the last decade that aim to weaken identification assumptions, relax functional form assumptions, and estimate new quantities
  - Double Machine Learning (Chernozhukov, Chetverikov, et al. 2018) is now well known and hinges on *selection on observables*: the treatment is as good as randomly assigned conditional on observed covariates
  - With repeated measurements, we can relax this and allow for *selection on unobservables* using Difference-in-Differences, or Synthetic Control (and friends)

Introduction

- Explosion of methods in observational causal inference methods in the last decade that aim to weaken identification assumptions, relax functional form assumptions, and estimate new quantities
  - Double Machine Learning (Chernozhukov, Chetverikov, et al. 2018) is now well known and hinges on *selection on observables*: the treatment is as good as randomly assigned conditional on observed covariates
  - With repeated measurements, we can relax this and allow for *selection on unobservables* using Difference-in-Differences, or Synthetic Control (and friends)
- **This Paper**: Express popular research designs in a shared augmented balancing form, extensive simulation studies to guide empirical practice, software `abal`
- Complementary practitioner's guide to common framework for combining flexible models for causal problems : Ben-Michael, Feller, and Rothstein (2021), Shen et al. (2022), and Bruns-Smith et al. (2023)

Overview

The Estimand

- $(Y_i, W_i, \mathbf{X}_i)_{i=1}^N \in \mathbb{R} \times \{0,1\} \times \mathcal{X} \subseteq \mathbb{R}^d$. Corresponding covariate distributions for treatment $\mathcal{T}$ and control $\mathcal{C}$.
- ATE ($\mathbb{E}\left[Y^{(1)} - Y^{(0)}\right]$) and ATT ($\mathbb{E}\left[Y^{(1)} - Y^{(0)} \mid W = 1\right]$) are both substantively meaningful estimands, and require related but distinct identification assumptions

The Estimand

- $(Y_i, W_i, \mathbf{X}_i)_{i=1}^N \in \mathbb{R} \times \{0, 1\} \times \mathcal{X} \subseteq \mathbb{R}^d$. Corresponding covariate distributions for treatment $\mathcal{T}$ and control $\mathcal{C}$.
- ATE ($\mathbb{E}\left[Y^{(1)} - Y^{(0)}\right]$) and ATT ($\mathbb{E}\left[Y^{(1)} - Y^{(0)} \mid W = 1\right]$) are both substantively meaningful estimands, and require related but distinct identification assumptions
  - ATE requires positive treatment probability for *all* units. In many observational settings where units self-select into treatment, this is simply not credible.
  - ATT is a compromise: condition on realised treatment assignment $\mathcal{T}$, and construct potential outcome in the absence of treatment for treated units.

The Estimand

- $(Y_i, W_i, \mathbf{X}_i)_{i=1}^N \in \mathbb{R} \times \{0,1\} \times \mathcal{X} \subseteq \mathbb{R}^d$. Corresponding covariate distributions for treatment $\mathcal{T}$ and control $\mathcal{C}$.
- ATE ($\mathbb{E}\left[Y^{(1)} - Y^{(0)}\right]$) and ATT ($\mathbb{E}\left[Y^{(1)} - Y^{(0)} \mid W = 1\right]$) are both substantively meaningful estimands, and require related but distinct identification assumptions
  - ATE requires positive treatment probability for *all* units. In many observational settings where units self-select into treatment, this is simply not credible.
  - ATT is a compromise: condition on realised treatment assignment $\mathcal{T}$, and construct potential outcome in the absence of treatment for treated units.
  - Sample mean of treatment outcomes is consistent for $\mathbb{E}\left[Y^{(1)} \mid W = 1\right]$
  - Need to construct estimator for $\mathbb{E}\left[Y^{(0)} \mid W = 1\right] \equiv \widehat{\mathbb{E}}_{\mathcal{T}}[Y^{(0)}] =: \widehat{\xi}$

The Estimand

- $(Y_i, W_i, \mathbf{X}_i)_{i=1}^N \in \mathbb{R} \times \{0,1\} \times \mathcal{X} \subseteq \mathbb{R}^d$. Corresponding covariate distributions for treatment $\mathcal{T}$ and control $\mathcal{C}$.
- ATE ($\mathbb{E}\left[Y^{(1)} - Y^{(0)}\right]$) and ATT ($\mathbb{E}\left[Y^{(1)} - Y^{(0)} \mid W = 1\right]$) are both substantively meaningful estimands, and require related but distinct identification assumptions
  - ATE requires positive treatment probability for *all* units. In many observational settings where units self-select into treatment, this is simply not credible.
  - ATT is a compromise: condition on realised treatment assignment $\mathcal{T}$, and construct potential outcome in the absence of treatment for treated units.
  - Sample mean of treatment outcomes is consistent for $\mathbb{E}\left[Y^{(1)} \mid W = 1\right]$
  - Need to construct estimator for $\mathbb{E}\left[Y^{(0)} \mid W = 1\right] \equiv \widehat{\mathbb{E}}_{\mathcal{T}}[Y^{(0)}] =: \widehat{\xi}$
- Analogous 'canonicalization' problem: generalise from A/B test to a target distribution: Estimate $\mathbb{E}\left[Y^{(w)} \mid S = 0\right]$ (bridgerton)

Propensity Scores vs Balancing Weights

- One way to compute $\widehat{\xi}$ is through reweighting $\widehat{\mathbb{E}}_{\mathcal{C}}\left(\frac{d\mathcal{T}}{d\mathcal{C}}(X)Y\right)$
- Density ratio $\frac{d\mathcal{T}}{d\mathcal{C}}(X)$ is challenging to estimate using plug-in estimation
- Standard practice: fit model $\pi(\mathbf{X}) = \mathbb{E}\left[W = 1 \mid \mathbf{X}\right]$, plug in to construct inverse-pscore weight $\frac{\pi(\mathbf{X})}{1-\pi(\mathbf{X})}$

Propensity Scores vs Balancing Weights

- One way to compute $\widehat{\xi}$ is through reweighting $\widehat{\mathbb{E}}_{\mathcal{C}}\left(\frac{d\mathcal{T}}{d\mathcal{C}}(X)Y\right)$
- Density ratio $\frac{d\mathcal{T}}{d\mathcal{C}}(X)$ is challenging to estimate using plug-in estimation
- Standard practice: fit model $\pi(\mathbf{X}) = \mathbb{E}\left[W = 1 \mid \mathbf{X}\right]$, plug in to construct inverse-pscore weight $\frac{\pi(\mathbf{X})}{1-\pi(\mathbf{X})}$
- This inversion step can inflate errors
- Alternative: directly estimate weights to minimize covariate imbalance
  - 'Automatic' estimation of the Riesz Representer (Hirshberg and Wager 2021; Chernozhukov, Newey, and Singh 2022)

Propensity Scores vs Balancing Weights

- One way to compute $\widehat{\xi}$ is through reweighting $\widehat{\mathbb{E}}_{\mathcal{C}}\left(\frac{d\mathcal{T}}{d\mathcal{C}}(X)Y\right)$
- Density ratio $\frac{d\mathcal{T}}{d\mathcal{C}}(X)$ is challenging to estimate using plug-in estimation
- Standard practice: fit model $\pi(\mathbf{X}) = \mathbb{E}\left[W = 1 \mid \mathbf{X}\right]$, plug in to construct inverse-pscore weight $\frac{\pi(\mathbf{X})}{1-\pi(\mathbf{X})}$
- This inversion step can inflate errors
- Alternative: directly estimate weights to minimize covariate imbalance
  - 'Automatic' estimation of the Riesz Representer (Hirshberg and Wager 2021; Chernozhukov, Newey, and Singh 2022)

$$\min_{\gamma} \overbrace{h_{\zeta}(\mathbf{X}_1 - \mathbf{X}_0'\boldsymbol{\gamma})}^{\text{Balance}} + \sum_{i \in \mathcal{C}} \overbrace{f(\gamma_i)}^{\text{Dispersion}}$$

$$\text{s.t.} \sum_{i \in \mathcal{C}} \gamma_i = 1$$

Propensity Scores vs Balancing Weights

- One way to compute $\widehat{\xi}$ is through reweighting $\widehat{\mathbb{E}}_{\mathcal{C}}\left(\frac{d\mathcal{T}}{d\mathcal{C}}(X)Y\right)$
- Density ratio $\frac{d\mathcal{T}}{d\mathcal{C}}(X)$ is challenging to estimate using plug-in estimation
- Standard practice: fit model $\pi(\mathbf{X}) = \mathbb{E}[W = 1 \mid \mathbf{X}]$, plug in to construct inverse-pscore weight $\frac{\pi(\mathbf{X})}{1-\pi(\mathbf{X})}$
- This inversion step can inflate errors
- Alternative: directly estimate weights to minimize covariate imbalance
    - 'Automatic' estimation of the Riesz Representer (Hirshberg and Wager 2021; Chernozhukov, Newey, and Singh 2022)

$$\min_{\gamma} \overbrace{h_{\zeta}(\mathbf{X}_1 - \mathbf{X}_0'\boldsymbol{\gamma})}^{\text{Balance}} + \sum_{i \in \mathcal{C}} \overbrace{f(\gamma_i)}^{\text{Dispersion}}$$

$$\text{s.t. } \sum_{i \in \mathcal{C}} \gamma_i = 1$$

For convex $f(\cdot)$, dual is easy to solve as regularized propensity score

$$\min_{\alpha, \beta} \sum_{i \in \mathcal{C}} f^*(\alpha + \boldsymbol{\beta}'\mathbf{X}_{i \cdot}) - (\alpha + \boldsymbol{\beta}'\mathbf{X}_1) + h_{\zeta}^*(\boldsymbol{\beta})$$

$$\widehat{\gamma}^* = f^{*\prime}(\widehat{\alpha} + \widehat{\boldsymbol{\beta}}'\mathbf{X}_i)$$

rsw implementation with ADMM

Cross Sectional: Identification and Estimation

Identification Assumptions

- SUTVA:
  $Y_i = W_i Y^{(1)} + (1 - W_i) Y^{(0)}$
- Unconfoundedness: $Y^{(0)} \perp\!\!\!\perp W | \mathbf{X}_i$
- Overlap: $\mathbf{Pr}\left(W = 1 | \mathbf{X}\right) < 1$

Share of treated observations

$\widehat{\rho} := \mathbf{Pr}\left(W = 1\right)$

Cross Sectional: Identification and Estimation

Estimators

Identification Assumptions

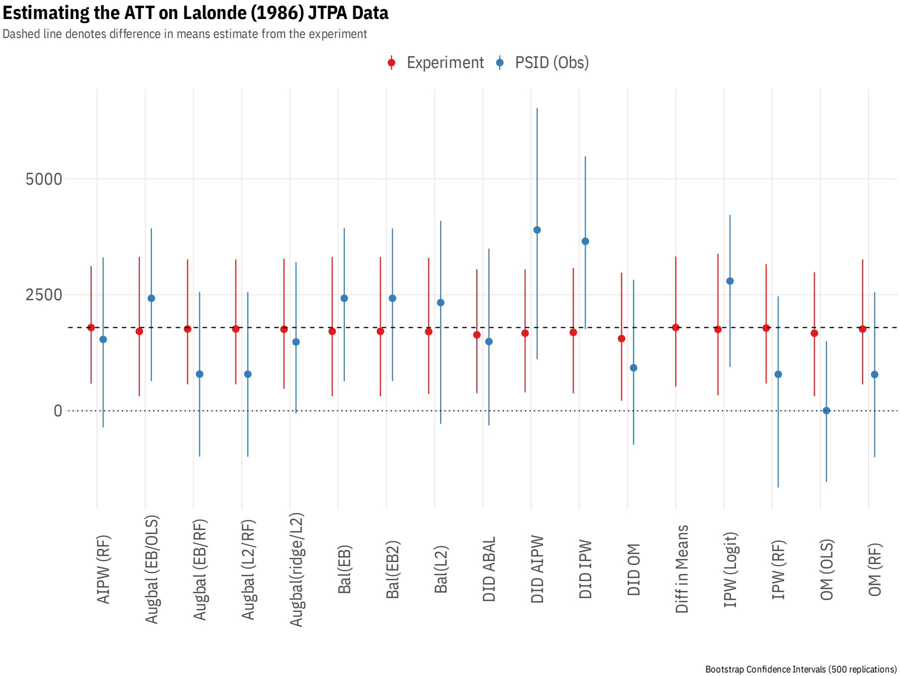
- SUTVA:
  $Y_i = W_i Y^{(1)} + (1 - W_i) Y^{(0)}$
- Unconfoundedness: $Y^{(0)} \perp\!\!\!\perp W | \mathbf{X}_i$
- Overlap: $\mathbf{Pr}\left(W = 1 | \mathbf{X}\right) < 1$

Share of treated observations

$\widehat{\rho} := \mathbf{Pr}\left(W = 1\right)$

- **Outcome Modelling**
  $\widehat{\xi}^{\text{OM}} := \frac{1}{n_t} \sum_{i \in \mathcal{T}} \widehat{\mu}^{(0)}(\mathbf{X}_i)$
- **Reweighting** $\widehat{\xi}^{\text{wt}} = \sum_{i \in \mathcal{C}} \gamma_i Y_i$
- **Augmented Balancing**

$$\widehat{\xi}^{\text{AUGBAL}} = \underbrace{\frac{1}{\widehat{\rho}} \sum_{i \in \mathcal{T}} \widehat{\mu}^{(0)}(\mathbf{X}_i)}_{\text{Reg}} + \underbrace{\frac{1}{\widehat{\rho}} \frac{1}{n} \sum_{i \in \mathcal{C}} \gamma_i \left\{ Y_i - \widehat{\mu}^{(0)}(\mathbf{X}_i) \right\}}_{\text{Reweighted Residuals}}$$



Estimating the ATT on Lalonde (1986) JTPA Data
Dashed line denotes difference in means estimate from the experiment
Experiment
PSID (Obs)
5000
2500
0
AIPW (RF)
Augbal (EB/OLS)
Augbal (EB/RF)
Augbal (L2/RF)
Augbal(ridge/L2)
Bal(EB)
Bal(EB2)
Bal(L2)
DID ABAL
DID AIPW
DID IPW
DID OM
Diff in Means
IPW (Logit)
IPW (RF)
OM (OLS)
OM (RF)
Bootstrap Confidence Intervals (500 replications)

Formal Properties: The role of augmentation

Following Ben-Michael, Feller, Hirshberg, et al. (2021), we can decompose errors

$$\widehat{\xi}^{\text{WT}} - \xi = \underbrace{\frac{1}{n}\sum_i (1-W_i)\widehat{\gamma}_i \mu^{(0)}(\mathbf{x}_i) - \frac{1}{n}\sum_i W_i \mu^{(0)}(\mathbf{x}_i)}_{\text{Bias from Imbalance}} - \underbrace{\frac{1}{n}\sum_{i=1}^{n}(1-W_i)\widehat{\gamma}_i \varepsilon_i}_{\text{Noise}} + \underbrace{\frac{1}{n}\sum_{i=1}^{n} W_i \mu^{(0)}(\mathbf{x}_i) - \xi}_{\text{Sampling}}$$

$$\widehat{\xi}^{\text{AUGBAL}} - \xi = \underbrace{\frac{1}{n}\sum_i W_i \underbrace{\widetilde{\mu}^{(0)}(\mathbf{x}_i)}_{=:\ \widehat{\mu}^{(0)} - \mu^{(0)}} - \frac{1}{n}\sum_{i=1}^{n}(1-W_i)\widehat{\gamma}_i \widetilde{\mu}^{(0)}(\mathbf{x}_i)}_{\text{Bias from Imbalance}} + \underbrace{\frac{1}{n}\sum_{i=1}^{n}(1-W_i)\widehat{\gamma}_i \varepsilon_i}_{\text{Noise}} + \underbrace{\frac{1}{n}\sum_{i=1}^{n} W_i \mu^{(0)}(\mathbf{x}_i) - \xi}_{\text{Sampling}}$$

Formal Properties: The role of augmentation

Following Ben-Michael, Feller, Hirshberg, et al. (2021), we can decompose errors

$$\widehat{\xi}^{\text{WT}} - \xi = \overbrace{\frac{1}{n}\sum_i (1-W_i)\widehat{\gamma}_i \mu^{(0)}(\mathbf{x}_i) - \frac{1}{n}\sum_i W_i \mu^{(0)}(\mathbf{x}_i)}^{\text{Bias from Imbalance}} - \overbrace{\frac{1}{n}\sum_{i=1}^{n}(1-W_i)\widehat{\gamma}_i \varepsilon_i}^{\text{Noise}} + \overbrace{\frac{1}{n}\sum_{i=1}^{n} W_i \mu^{(0)}(\mathbf{x}_i) - \xi}^{\text{Sampling}}$$

$$\widehat{\xi}^{\text{AUGBAL}} - \xi = \underbrace{\frac{1}{n}\sum_i W_i \underbrace{\widetilde{\mu}^{(0)}(\mathbf{x}_i)}_{=:\,\widehat{\mu}^{(0)} - \mu^{(0)}} - \frac{1}{n}\sum_{i=1}^{n}(1-W_i)\widehat{\gamma}_i \widetilde{\mu}^{(0)}(\mathbf{x}_i)}_{\text{Bias from Imbalance}} + \underbrace{\frac{1}{n}\sum_{i=1}^{n}(1-W_i)\widehat{\gamma}_i \varepsilon_i}_{\text{Noise}} + \underbrace{\frac{1}{n}\sum_{i=1}^{n} W_i \mu^{(0)}(\mathbf{x}_i) - \xi}_{\text{Sampling}}$$

If our outcome model isn't totally useless, regression error will be easier to balance than the unknown regression. $\widehat{\mu}^{(0)}$ and $\widehat{\gamma}_i$ play complementary roles: regression could soak up strong signals and weights pick up higher order ones.

Formal Properties: The role of augmentation

Following Ben-Michael, Feller, Hirshberg, et al. (2021), we can decompose errors

$$\widehat{\xi}^{\text{WT}} - \xi = \underbrace{\frac{1}{n}\sum_i (1-W_i)\widehat{\gamma}_i \mu^{(0)}(\mathbf{x}_i) - \frac{1}{n}\sum_i W_i \mu^{(0)}(\mathbf{x}_i)}_{\text{Bias from Imbalance}} - \underbrace{\frac{1}{n}\sum_{i=1}^n (1-W_i)\widehat{\gamma}_i \varepsilon_i}_{\text{Noise}} + \underbrace{\frac{1}{n}\sum_{i=1}^n W_i \mu^{(0)}(\mathbf{x}_i) - \xi}_{\text{Sampling}}$$

$$\widehat{\xi}^{\text{AUGBAL}} - \xi = \underbrace{\frac{1}{n}\sum_i W_i \underbrace{\widetilde{\mu}^{(0)}(\mathbf{x}_i)}_{=:\,\widehat{\mu}^{(0)} - \mu^{(0)}} - \frac{1}{n}\sum_{i=1}^n (1-W_i)\widehat{\gamma}_i \widetilde{\mu}^{(0)}(\mathbf{x}_i)}_{\text{Bias from Imbalance}} + \underbrace{\frac{1}{n}\sum_{i=1}^n (1-W_i)\widehat{\gamma}_i \varepsilon_i}_{\text{Noise}} + \underbrace{\frac{1}{n}\sum_{i=1}^n W_i \mu^{(0)}(\mathbf{x}_i) - \xi}_{\text{Sampling}}$$

If our outcome model isn't totally useless, regression error will be easier to balance than the unknown regression. $\widehat{\mu}^{(0)}$ and $\widehat{\gamma}_i$ play complementary roles: regression could soak up strong signals and weights pick up higher order ones.

Bruns-Smith et al. (2023) show that augmented balancing with linear weights collapses to a single ridge regression.

Formal Properties: The role of augmentation

Following Ben-Michael, Feller, Hirshberg, et al. (2021), we can decompose errors

$$\hat{\xi}^{\text{WT}} - \xi = \overbrace{\frac{1}{n}\sum_i (1-W_i)\hat{\gamma}_i \mu^{(0)}(\mathbf{x}_i) - \frac{1}{n}\sum_i W_i \mu^{(0)}(\mathbf{x}_i)}^{\text{Bias from Imbalance}} - \overbrace{\frac{1}{n}\sum_{i=1}^n (1-W_i)\hat{\gamma}_i \varepsilon_i}^{\text{Noise}} + \overbrace{\frac{1}{n}\sum_{i=1}^n W_i \mu^{(0)}(\mathbf{x}_i) - \xi}^{\text{Sampling}}$$

$$\hat{\xi}^{\text{AUGBAL}} - \xi = \underbrace{\frac{1}{n}\sum_i W_i \underbrace{\tilde{\mu}^{(0)}(\mathbf{x}_i)}_{=:\ \hat{\mu}^{(0)} - \mu^{(0)}} - \frac{1}{n}\sum_{i=1}^n (1-W_i)\hat{\gamma}_i \tilde{\mu}^{(0)}(\mathbf{x}_i)}_{\text{Bias from Imbalance}} + \underbrace{\frac{1}{n}\sum_{i=1}^n (1-W_i)\hat{\gamma}_i \varepsilon_i}_{\text{Noise}} + \underbrace{\frac{1}{n}\sum_{i=1}^n W_i \mu^{(0)}(\mathbf{x}_i) - \xi}_{\text{Sampling}}$$

If our outcome model isn't totally useless, regression error will be easier to balance than the unknown regression. $\hat{\mu}^{(0)}$ and $\hat{\gamma}_i$ play complementary roles: regression could soak up strong signals and weights pick up higher order ones.
Bruns-Smith et al. (2023) show that augmented balancing with linear weights collapses to a single ridge regression. One possible implication: Balancing score and outcome model should be in different bases.

Formal Properties: The role of augmentation

Following Ben-Michael, Feller, Hirshberg, et al. (2021), we can decompose errors

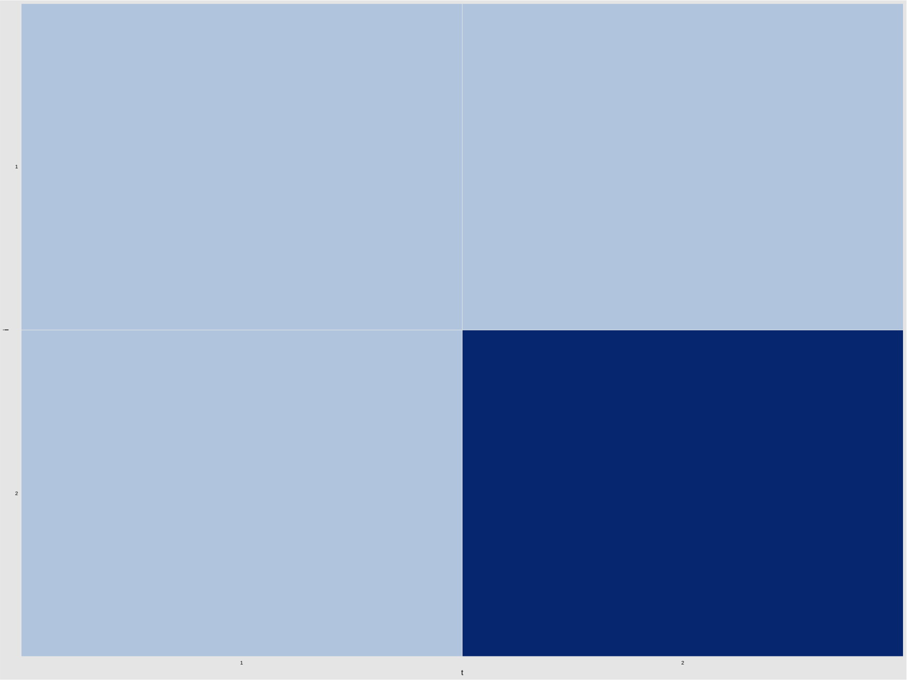
$$\hat{\xi}^{\text{WT}} - \xi = \overbrace{\frac{1}{n}\sum_i (1-W_i)\hat{\gamma}_i \mu^{(0)}(\mathbf{x}_i) - \frac{1}{n}\sum_i W_i \mu^{(0)}(\mathbf{x}_i)}^{\text{Bias from Imbalance}} - \overbrace{\frac{1}{n}\sum_{i=1}^{n}(1-W_i)\hat{\gamma}_i \varepsilon_i}^{\text{Noise}} + \overbrace{\frac{1}{n}\sum_{i=1}^{n} W_i \mu^{(0)}(\mathbf{x}_i) - \xi}^{\text{Sampling}}$$

$$\hat{\xi}^{\text{AUGBAL}} - \xi = \underbrace{\frac{1}{n}\sum_i W_i \underbrace{\tilde{\mu}^{(0)}(\mathbf{x}_i)}_{=:\hat{\mu}^{(0)} - \mu^{(0)}} - \frac{1}{n}\sum_{i=1}^{n}(1-W_i)\hat{\gamma}_i \tilde{\mu}^{(0)}(\mathbf{x}_i)}_{\text{Bias from Imbalance}} + \underbrace{\frac{1}{n}\sum_{i=1}^{n}(1-W_i)\hat{\gamma}_i \varepsilon_i}_{\text{Noise}} + \underbrace{\frac{1}{n}\sum_{i=1}^{n} W_i \mu^{(0)}(\mathbf{x}_i) - \xi}_{\text{Sampling}}$$

If our outcome model isn't totally useless, regression error will be easier to balance than the unknown regression. $\hat{\mu}^{(0)}$ and $\hat{\gamma}_i$ play complementary roles: regression could soak up strong signals and weights pick up higher order ones.

Bruns-Smith et al. (2023) show that augmented balancing with linear weights collapses to a single ridge regression. One possible implication: Balancing score and outcome model should be in different bases.

Consistent and Asymptotically Normal, Semiparametrically efficient, admits to standard variance formula (Ben-Michael, Feller, Hirshberg, et al. 2021)[Appdx A]



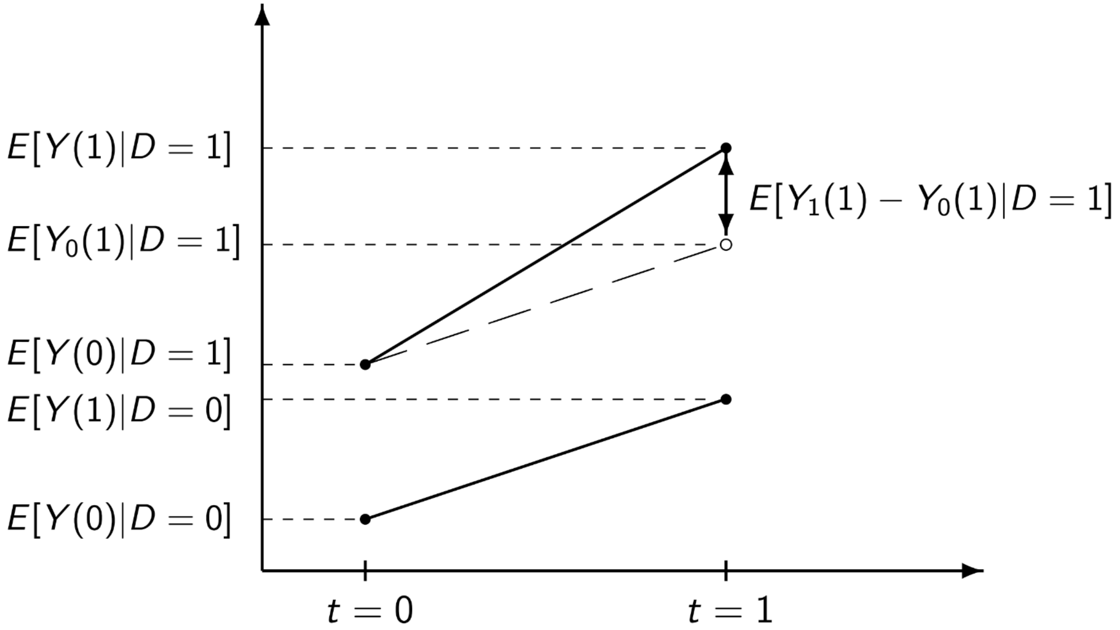
i
1
2
1
2
t

Difference in Differences: Identification

- Unconfoundedness is often not credible. We want to allow for level differences in $Y^{(0)}$ across treatment and control due to unobserved factors
- Two periods, $(Y_{i1}, Y_{i0}, W_i, \mathbf{X}_i)_{i=1}^N$. Treatment W applies in second period
- Estimator $\widehat{\xi} := \widehat{\mathbb{E}}\left[Y_{i1}^{(0)} \mid W = 1\right]$

Difference in Differences: Identification

- Unconfoundedness is often not credible. We want to allow for level differences in $Y^{(0)}$ across treatment and control due to unobserved factors
- Two periods, $(Y_{i1}, Y_{i0}, W_i, \mathbf{X}_i)_{i=1}^N$. Treatment W applies in second period
- Estimator $\widehat{\xi} := \widehat{\mathbb{E}}\left[Y_{i1}^{(0)} \mid W = 1\right]$
- Identification Assumptions
  1. No anticipation $\mathbb{E}\left[Y_{i0} \mid W_i = 1\right] = \mathbb{E}\left[Y_{i0}^{(0)} \mid W_i = 0\right]$
  2. Conditional Parallel Trends
     $\mathbb{E}\left[Y_{i1}^{(0)} - Y_{i0}^{(0)} \mid W = 1, \mathbf{X}\right] = \mathbb{E}\left[Y_{i1}^{(0)} - Y_{i0}^{(0)} \mid W = 0, \mathbf{X}\right]$



$E[Y(1)|D=1]$
$E[Y_1(1) - Y_0(1)|D=1]$
$E[Y_0(1)|D=1]$
$E[Y(0)|D=1]$
$E[Y(1)|D=0]$
$E[Y(0)|D=0]$
$t=0$
$t=1$

Difference in Differences: Estimation

- Outcome Modelling

$$\widehat{\xi}^{\text{DID}} = \underbrace{\frac{1}{|\mathcal{T}|}\sum_{i\in\mathcal{T}} Y_{i0}}_{\text{Baseline outcome for treated}} + \underbrace{\frac{1}{|\mathcal{C}|}\sum_{i\in\mathcal{C}}\left(\overbrace{Y_{i1} - Y_{0i}}^{=:\Delta_i}\right)}_{\text{Trend for control}}$$

Difference in Differences: Estimation

- Outcome Modelling

$$\widehat{\xi}^{\text{DID}} = \underbrace{\frac{1}{|\mathcal{T}|}\sum_{i\in\mathcal{T}} Y_{i0}}_{\text{Baseline outcome for treated}} + \underbrace{\frac{1}{|\mathcal{C}|}\sum_{i\in\mathcal{C}} \left(\overbrace{Y_{i1} - Y_{0i}}^{=:\Delta_i}\right)}_{\text{Trend for control}}$$

- Reweighting (Abadie (2005) proposes IPW with $\gamma_i = \pi(\mathbf{X}_i)/(1-\pi(\mathbf{X}_i))$

$$\widehat{\xi}^{\text{wt}} = \frac{1}{|\mathcal{T}|}\sum_{i\in\mathcal{T}} Y_{i0} + \frac{1}{|\mathcal{C}|}\sum_{i\in\mathcal{C}} \gamma_i (Y_{i1} - Y_{i0})$$

Difference in Differences: Estimation

- Outcome Modelling

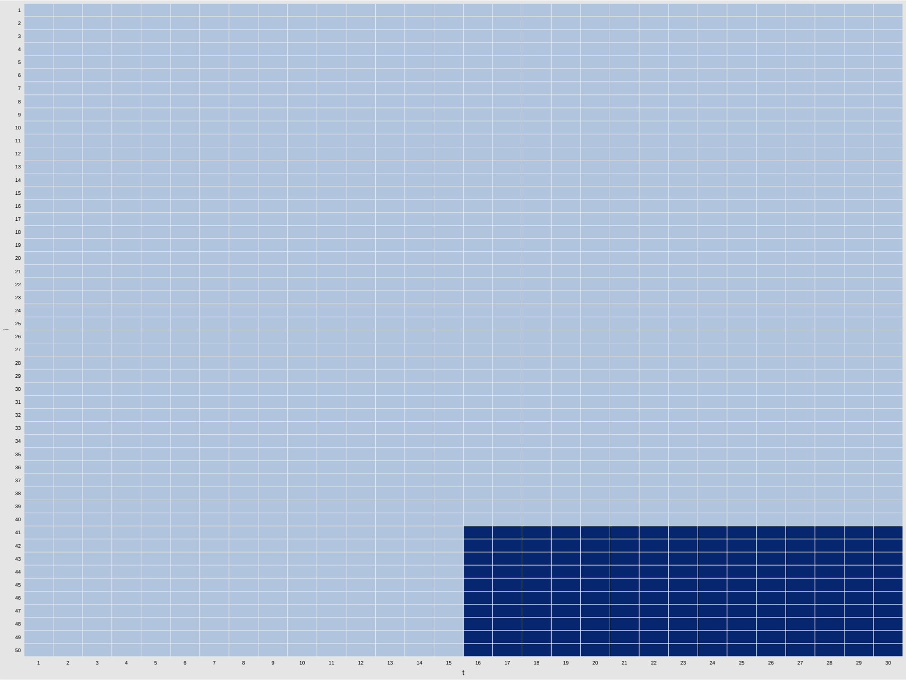
$$\widehat{\xi}^{\text{DID}} = \underbrace{\frac{1}{|\mathcal{T}|}\sum_{i\in\mathcal{T}} Y_{i0}}_{\text{Baseline outcome for treated}} + \underbrace{\frac{1}{|\mathcal{C}|}\sum_{i\in\mathcal{C}}\left(\overbrace{Y_{i1} - Y_{0i}}^{=:\Delta_i}\right)}_{\text{Trend for control}}$$

- Reweighting (Abadie (2005) proposes IPW with $\gamma_i = \pi(\mathbf{X}_i)/(1-\pi(\mathbf{X}_i))$

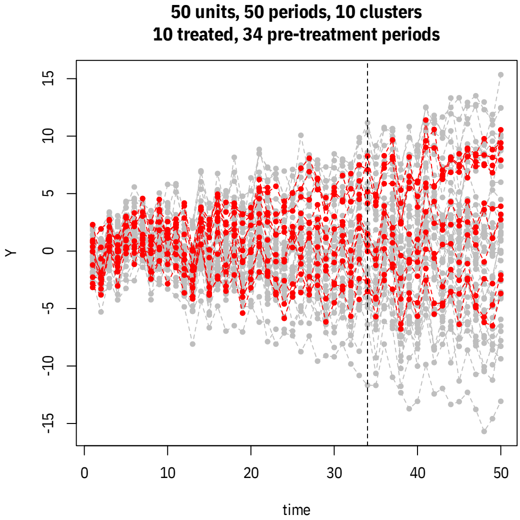
$$\widehat{\xi}^{\text{wt}} = \frac{1}{|\mathcal{T}|}\sum_{i\in\mathcal{T}} Y_{i0} + \frac{1}{|\mathcal{C}|}\sum_{i\in\mathcal{C}} \gamma_i (Y_{i1} - Y_{i0})$$

- Augmented Balancing (with $\widehat{\mu}^0(\mathbf{X}_i) = \mathbb{E}\left[\Delta_i \mid \mathbf{X}_i = \mathbf{x}_i, W_i = 0\right]$)

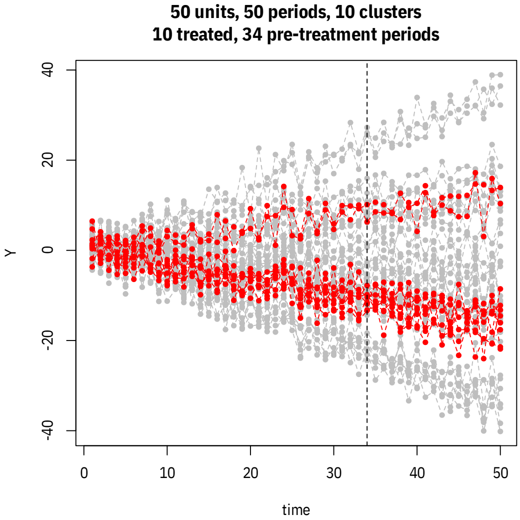
$$\widehat{\xi}^{\text{AUGBAL DID}} = \frac{1}{|\mathcal{T}|}\sum_{i\in\mathcal{T}} Y_{i0} - \widehat{\mu}^0(\mathbf{X}_i) + \frac{1}{|\mathcal{C}|}\sum_{i\in\mathcal{C}} \gamma_i\left(\Delta_i - \widehat{\mu}^0(\mathbf{X}_i)\right)$$



i
t
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50



50 units, 50 periods, 10 clusters
10 treated, 34 pre-treatment periods
Y
time
15
10
5
0
-5
-10
-15
0
10
20
30
40
50



50 units, 50 periods, 10 clusters
10 treated, 34 pre-treatment periods
Y
time
40
20
0
-20
-40
0
10
20
30
40
50

Panel Data: Identification

- Data: $(\mathbf{Y}_{it}, \mathbf{W}_{it})_{i=1}^{N}, t \in [T]$. Absorbing treatment, one-shot adoption by N_1 units at time $T_0 + 1$
- Commonly used under analogous assumptions to 2-period DID
  - 'Long' Parallel Trends
    $\mathbb{E}\left[Y_{it}^{(0)} - Y_{it'}^{(0)} | W_i = 1\right] = \mathbb{E}\left[Y_{it}^{(0)} - Y_{it'}^{(0)} | W_i = 0\right] \; \forall \, t \neq t'$
  - Frequently paired with corresponding representation for untreated PO
    $Y_{it}^{(0)} = \alpha_i + \gamma_t + \varepsilon_{it}$ (Liu, Wang, and Xu 2021; Borusyak, Jaravel, and Spiess 2022)

Panel Data: Identification

- Data: $(\mathbf{Y}_{it}, \mathbf{W}_{it})_{i=1}^{N}, t \in [T]$. Absorbing treatment, one-shot adoption by N_1 units at time $T_0 + 1$
- Commonly used under analogous assumptions to 2-period DID
  - 'Long' Parallel Trends
    $\mathbb{E}\left[Y_{it}^{(0)} - Y_{it'}^{(0)} | W_i = 1\right] = \mathbb{E}\left[Y_{it}^{(0)} - Y_{it'}^{(0)} | W_i = 0\right] \; \forall \, t \neq t'$
  - Frequently paired with corresponding representation for untreated PO $Y_{it}^{(0)} = \alpha_i + \gamma_t + \varepsilon_{it}$ (Liu, Wang, and Xu 2021; Borusyak, Jaravel, and Spiess 2022)
- Alternate identification assumptions
  1. Latent Factor Model: $Y_{it}^{(0)} = \sum_{j=1}^{J} \phi_{ij} \mu_{jt} + \varepsilon_{it}$ with unknown time-varying factors $\boldsymbol{\mu}_t = \{\mu_{jt}\} \in \mathbb{R}^T, j = 1, \ldots, J$ and unknown unit loadings $\boldsymbol{\phi}_i \in \mathbb{R}^J$ (Abadie, Diamond, and Hainmueller 2010; Xu 2017)
  2. Unconfoundedness given history: $Y_{it}^{(0)} \perp\!\!\!\perp W_i | \mathbf{Y}_{i,1:T_0} \; \forall t > T_0$ (Ben-Michael, Feller, and Rothstein 2021)

Panel Data: Estimation inference

$$
\left(\begin{array}{cccc|c}
Y_{1,1} & Y_{1,2} & \dots & Y_{1,T_0} & Y_{1,T} \\
Y_{2,1} & Y_{2,2} & \dots & Y_{2,T_0} & Y_{2,T} \\
\vdots & & & & \vdots \\
Y_{N_0,1} & Y_{N_0,2} & \dots & Y_{N_0,T_0} & Y_{N_0,T} \\
\hline
\vdots & & & & ? \\
Y_{N,1} & Y_{N,2} & \dots & Y_{N,T_0} & ?
\end{array}\right)
=: \left(\begin{array}{c|c}
\mathbf{X}^0 & \mathbf{y}^n \\
\hline
\mathbf{X}^1 & ?
\end{array}\right)
$$

(Athey et al. 2021) formalism:
SC fit $\mathbf{X}^1 \sim \mathbf{X}^0$ (Vertical Regression)
Autoregressive models fit $\mathbf{y}^n \sim \mathbf{X}^0$ (Horizontal Regression)
Some outcome models (DFM, MC) fit both.

Panel Data: Estimation

inference

$$\begin{pmatrix} Y_{1,1} & Y_{1,2} & \dots & Y_{1,T_0} & Y_{1,T} \\ Y_{2,1} & Y_{2,2} & \dots & Y_{2,T_0} & Y_{2,T} \\ \vdots & & & & \vdots \\ Y_{N_0,1} & Y_{N_0,2} & \dots & Y_{N_0,T_0} & Y_{N_0,T} \\ \hline \vdots & & & & ? \\ Y_{N,1} & Y_{N,2} & \dots & Y_{N,T_0} & ? \end{pmatrix} =: \begin{pmatrix} \mathbf{X}^0 & \mathbf{y}^n \\ \hline \mathbf{X}^1 & ? \end{pmatrix}$$

(Athey et al. 2021) formalism:
SC fit $\mathbf{X}^1 \sim \mathbf{X}^0$ (Vertical Regression)
Autoregressive models fit $\mathbf{y}^n \sim \mathbf{X}^0$ (Horizontal Regression)
Some outcome models (DFM, MC) fit both.

Outcome Modelling : $\widehat{\xi}^{\text{HR}} = \widehat{\mu}^0(\mathbf{X}^1)$

Balancing

$\widehat{\xi}^{\text{VR}} = \langle \widehat{\gamma}, \mathbf{y}^n \rangle$ where

$$\widehat{\gamma} = \underset{\gamma \in \Delta_{|C|-1}}{\operatorname{argmin}} \; h(\gamma) \text{ s.t. } \langle \gamma, \mathbf{X}^0 \rangle \approx \overline{\mathbf{X}}^1 + \mu$$

Augmented Balancing

$$\widehat{\xi}^{\text{AugBal}} = \widehat{\mu}^0(\mathbf{X}, \mathbf{y}_n) + \sum_{i \in \mathcal{C}, t > T_0} \widehat{\gamma}_i (Y_{it} - \widehat{\mu}(\mathbf{X}, \mathbf{y}_n))$$

Panel Data: Estimation inference

$$\begin{pmatrix} Y_{1,1} & Y_{1,2} & \dots & Y_{1,T_0} & Y_{1,T} \\ Y_{2,1} & Y_{2,2} & \dots & Y_{2,T_0} & Y_{2,T} \\ \vdots & & & & \vdots \\ Y_{N_0,1} & Y_{N_0,2} & \dots & Y_{N_0,T_0} & Y_{N_0,T} \\ \hline \vdots & & & & ? \\ Y_{N,1} & Y_{N,2} & \dots & Y_{N,T_0} & ? \end{pmatrix} =: \begin{pmatrix} \mathbf{X}^0 & \mathbf{y}^n \\ \hline \mathbf{X}^1 & ? \end{pmatrix}$$

(Athey et al. 2021) formalism:
SC fit $\mathbf{X}^1 \sim \mathbf{X}^0$ (Vertical Regression)
Autoregressive models fit $\mathbf{y}^n \sim \mathbf{X}^0$ (Horizontal Regression)
Some outcome models (DFM, MC) fit both.

Outcome Modelling : $\widehat{\xi}^{\text{HR}} = \widehat{\mu}^0(\mathbf{X}^1)$
Balancing

$$\widehat{\xi}^{\text{VR}} = \langle \widehat{\gamma}, \mathbf{y}^n \rangle \text{ where}$$

$$\widehat{\gamma} = \underset{\gamma \in \Delta_{|C|-1}}{\text{argmin}}\ h(\gamma) \text{ s.t. } \langle \gamma, \mathbf{X}^0 \rangle \approx \overline{\mathbf{X}}^1 + \mu$$

Augmented Balancing

$$\widehat{\xi}^{\text{AugBal}} = \widehat{\mu}^0(\mathbf{X}, \mathbf{y}_n) + \sum_{i \in \mathcal{C}, t > T_0} \widehat{\gamma}_i (Y_{it} - \widehat{\mu}(\mathbf{X}, \mathbf{y}_n))$$

SDID: simplex regression for both
Augsynth: ridge regression for both

Panel Data: Estimation

inference

$$\begin{pmatrix} Y_{1,1} & Y_{1,2} & \dots & Y_{1,T_0} & Y_{1,T} \\ Y_{2,1} & Y_{2,2} & \dots & Y_{2,T_0} & Y_{2,T} \\ \vdots & & & & \vdots \\ Y_{N_0,1} & Y_{N_0,2} & \dots & Y_{N_0,T_0} & Y_{N_0,T} \\ \hline \vdots & & & & ? \\ Y_{N,1} & Y_{N,2} & \dots & Y_{N,T_0} & ? \end{pmatrix} =: \begin{pmatrix} \mathbf{X}^0 & \mathbf{y}^n \\ \hline \mathbf{X}^1 & ? \end{pmatrix}$$

(Athey et al. 2021) formalism:

SC fit $\mathbf{X}^1 \sim \mathbf{X}^0$ (Vertical Regression)

Autoregressive models fit $\mathbf{y}^n \sim \mathbf{X}^0$ (Horizontal Regression)

Some outcome models (DFM, MC) fit both.

Outcome Modelling : $\widehat{\xi}^{\text{HR}} = \widehat{\mu}^0(\mathbf{X}^1)$

Balancing

$\widehat{\xi}^{\text{VR}} = \langle \widehat{\gamma}, \mathbf{y}^n \rangle$ where

$$\widehat{\gamma} = \underset{\gamma \in \Delta_{|C|-1}}{\operatorname{argmin}} \; h(\gamma) \text{ s.t. } \langle \gamma, \mathbf{X}^0 \rangle \approx \overline{\mathbf{X}}^1 + \mu$$

Augmented Balancing

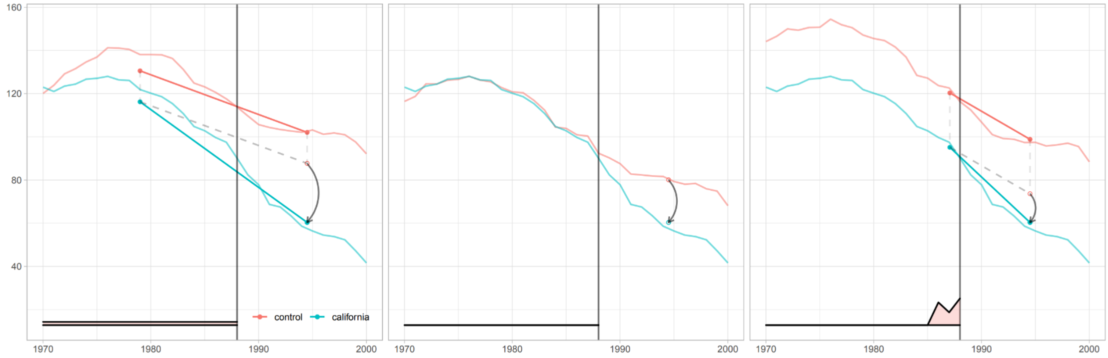
$$\widehat{\xi}^{\text{AugBal}} = \widehat{\mu}^0(\mathbf{X}, \mathbf{y}_n) + \sum_{i \in \mathcal{C}, t > T_0} \widehat{\gamma}_i (Y_{it} - \widehat{\mu}(\mathbf{X}, \mathbf{y}_n))$$

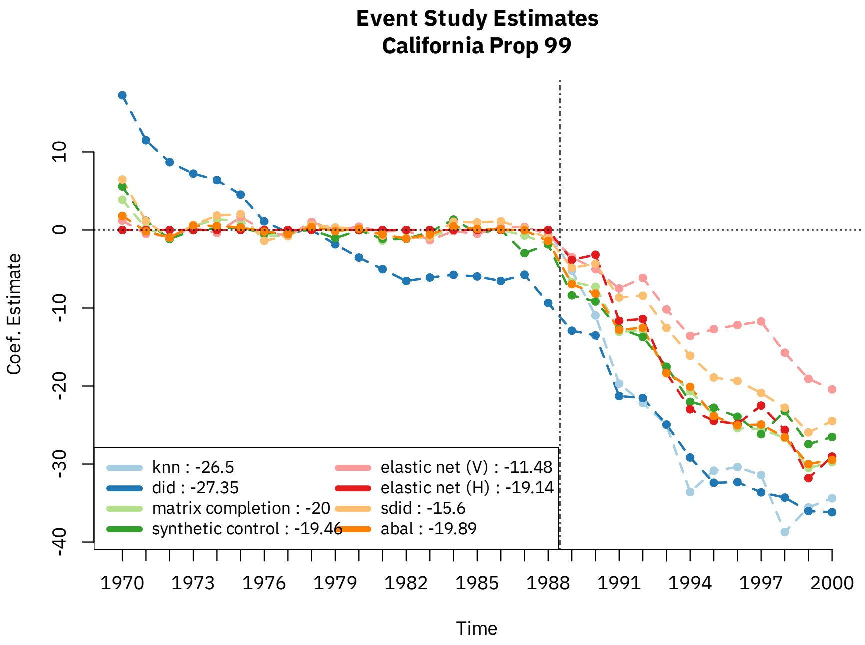
SDID: simplex regression for both

Augsynth: ridge regression for both

Alternative: Matrix Completion + Entropy

Diff. in Differences | Synthetic Control | Synth. Diff. in Diff.





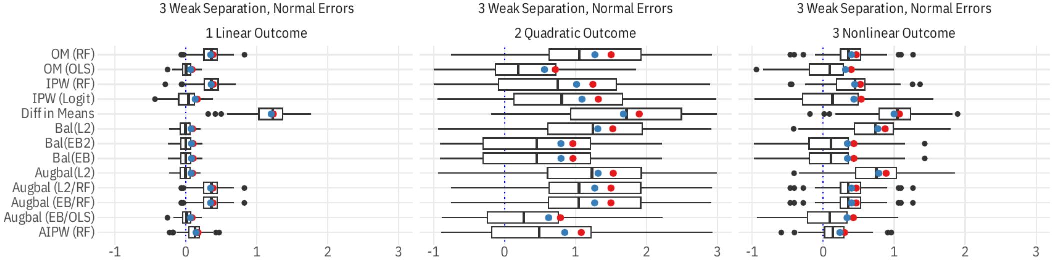
Event Study Estimates
California Prop 99
Coef. Estimate
10
0
-10
-20
-30
-40
1970
1973
1976
1979
1982
1985
1988
1991
1994
1997
2000
Time
knn : -26.5
did : -27.35
matrix completion : -20
synthetic control : -19.46
elastic net (V) : -11.48
elastic net (H) : -19.14
sdid : -15.6
abal : -19.89

Inference

- For cross-sectional and two-period estimators, we have a conventional score function that can be used to construct confidence intervals
  - With flexible nuisance models, cross-fitting required for valid inference
  - With a restricted class of models (Donsker or 'simple-enough' (leave-out stability (Chen, Syrgkanis, and Austern 2022)), can use full data
- For panel data, analogous techniques aren't available. Bootstrap or Jackknife shown to work well (Arkhangelsky et al. 2020)
  - With single treated unit, inference procedure is non-standard: use permutation tests or conformal methods

Overview

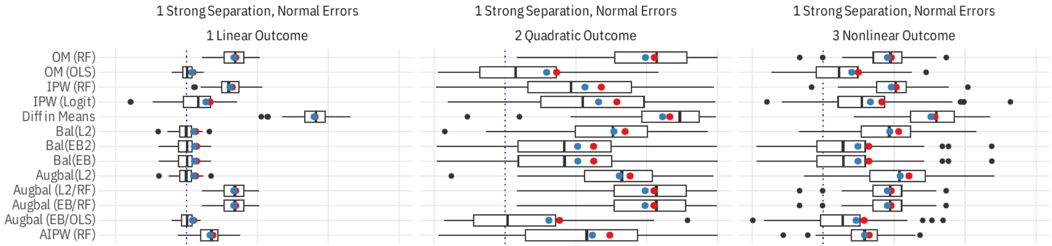
Cross-sectional Simulation: Good overlap



True effect of 0.

Extension of Froelich(2007), Hainmueller (2012)

Cross-sectional Simulation: Poor overlap



Cross-Sectional : ACIC 2016 DGP

ACIC (2016) DGPs (Dorie et al. 2019):
4802 observations and 58 covariates.
100 replications of 77 simulation settings that vary

- **Treatment model** $\in$ { Linear, polynomial, step }
- **Response model** $\in$ { Linear, exponential, step }
- **Treatment/Response Alignment** $\in$ {None, Low, High }
- **Heterogeneity** $\in$ { None, Low, High }
- **Overlap** $\in$ { Full, Penalty }
- **Treated %**

Cross-Sectional : ACIC 2016 DGP

ACIC (2016) DGPs (Dorie et al. 2019):
4802 observations and 58 covariates.
100 replications of 77 simulation settings that vary

- **Treatment model** $\in$ { Linear, polynomial, step }
- **Response model** $\in$ { Linear, exponential, step }
- **Treatment/Response Alignment** $\in$ {None, Low, High }
- **Heterogeneity** $\in$ { None, Low, High }
- **Overlap** $\in$ { Full, Penalty }
- **Treated %**

	Bias	RMSE
OLS	0.6146	0.7435
IPW	0.6302	2.3247
AIPW	0.1516	0.2070
EB1	0.4951	0.6461
EB2	0.2578	0.3689
HBAL	3.2490	3.8954
balHD	0.4585	0.5904
AugBalE	0.2001	0.3344

Previously, both L2 and ebal only succesfully computed $\approx$ 60% (Cousineau et al (2022)).

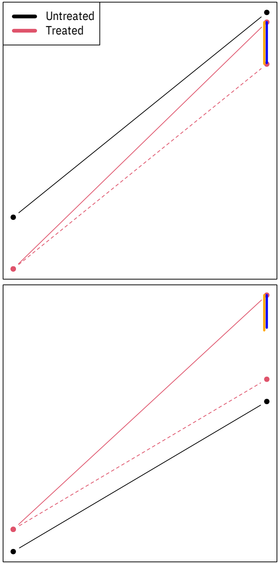
ebal performance in high dimensions

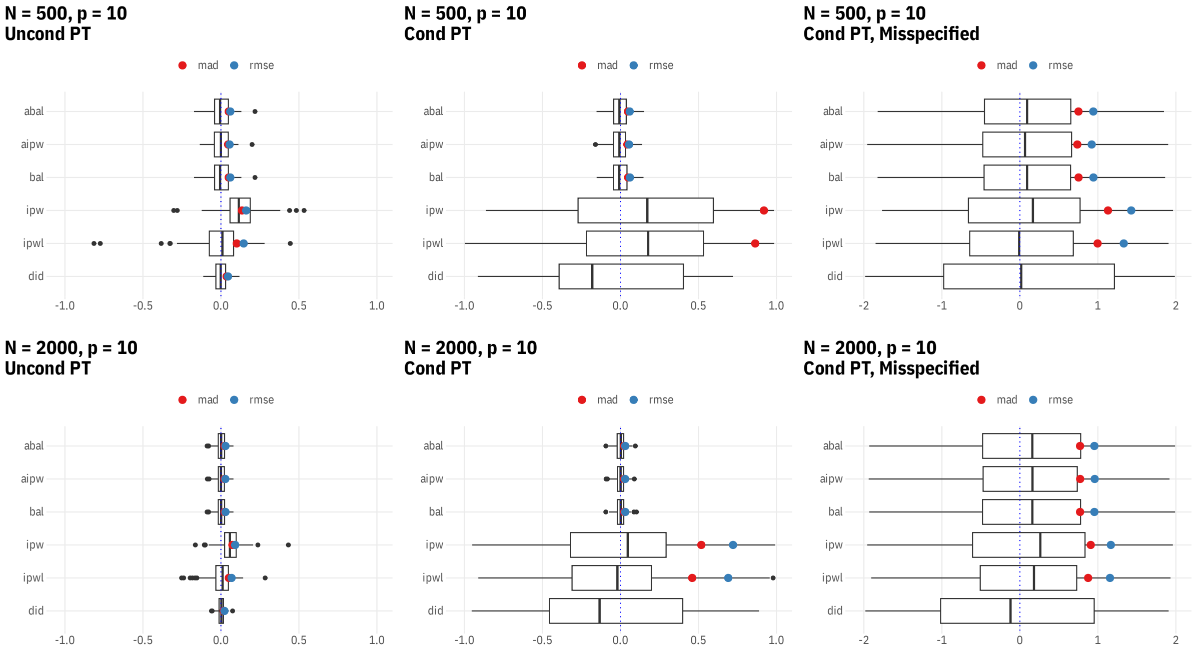
DiD simulation setup

- p-vector $\mathbf{X}_i \sim \mathcal{N}(0, \mathbf{\Sigma})$ where $\mathbf{\Sigma}$ follows Toeplitz form with entries $0.5^{0:(p-1)}$ (correlated covariates)
- $W_i \sim$ Bern $(\Lambda(\mathbf{X}_i'\boldsymbol{\gamma}))$, $\boldsymbol{\gamma}$ sparse U $[-1, 1]$
- Baseline outcomes $Y_{(w)i}(0)$ generated $\mathbf{X}_i'\boldsymbol{\beta}^{(w)} + \varepsilon_i$ with $\boldsymbol{\beta}^{(w)}$ sparse
- Trend $Y_{0i}(1) - Y_{0i}(0)$ generated $\mathbf{X}_i'\boldsymbol{\beta}^{\Delta} + \varepsilon_i$ (where $\boldsymbol{\beta}^{\Delta} = 0$ for PT)
- Estimand: ATT in the 2nd period

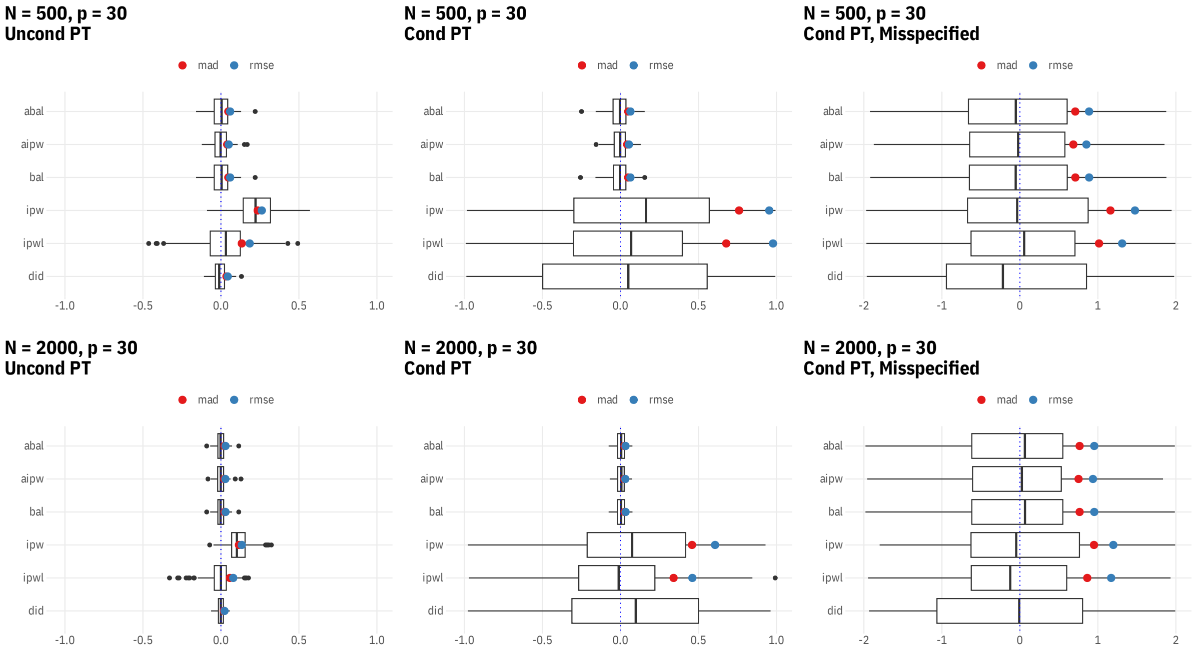
DiD simulation setup

- p-vector $\mathbf{X}_i \sim \mathcal{N}(0, \mathbf{\Sigma})$ where $\mathbf{\Sigma}$ follows Toeplitz form with entries $0.5^{0:(p-1)}$ (correlated covariates)
- $W_i \sim$ Bern $(\Lambda(\mathbf{X}_i'\boldsymbol{\gamma}))$, $\boldsymbol{\gamma}$ sparse U $[-1, 1]$
- Baseline outcomes $Y_{(w)i}(0)$ generated $\mathbf{X}_i'\boldsymbol{\beta}^{(w)} + \varepsilon_i$ with $\boldsymbol{\beta}^{(w)}$ sparse
- Trend $Y_{0i}(1) - Y_{0i}(0)$ generated $\mathbf{X}_i'\boldsymbol{\beta}^{\Delta} + \varepsilon_i$ (where $\boldsymbol{\beta}^{\Delta} = 0$ for PT)
- Estimand: ATT in the 2nd period





N = 500, p = 10
Uncond PT
N = 500, p = 10
Cond PT
N = 500, p = 10
Cond PT, Misspecified
N = 2000, p = 10
Uncond PT
N = 2000, p = 10
Cond PT
N = 2000, p = 10
Cond PT, Misspecified
mad
rmse
abal
aipw
bal
ipw
ipwl
did
-1.0
-0.5
0.0
0.5
1.0
-2
-1
0
1
2



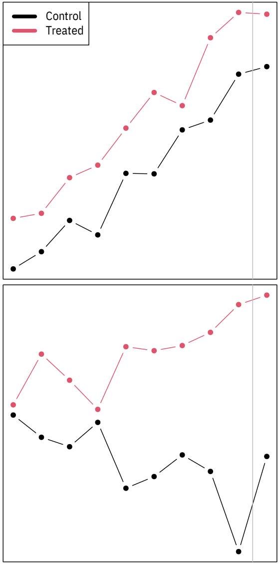
N = 500, p = 30
Uncond PT
N = 500, p = 30
Cond PT
N = 500, p = 30
Cond PT, Misspecified
N = 2000, p = 30
Uncond PT
N = 2000, p = 30
Cond PT
N = 2000, p = 30
Cond PT, Misspecified
mad
rmse
abal
aipw
bal
ipw
ipwl
did
-1.0
-0.5
0.0
0.5
1.0
-2
-1
0
1
2

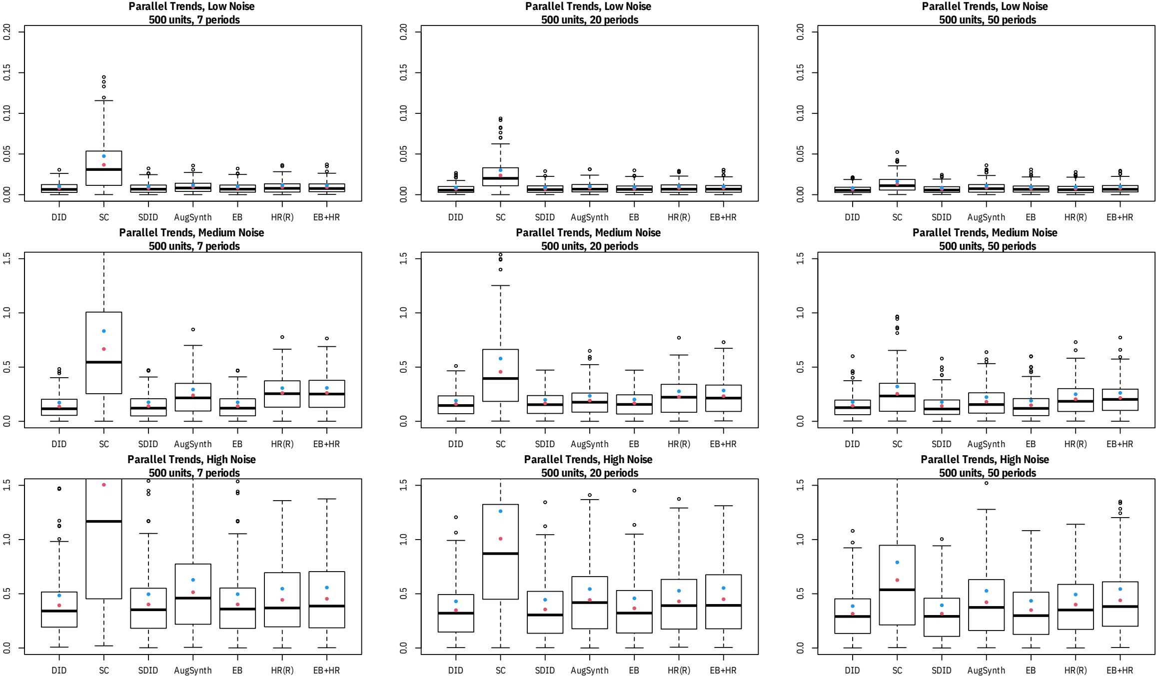
Panel Simulation Setup

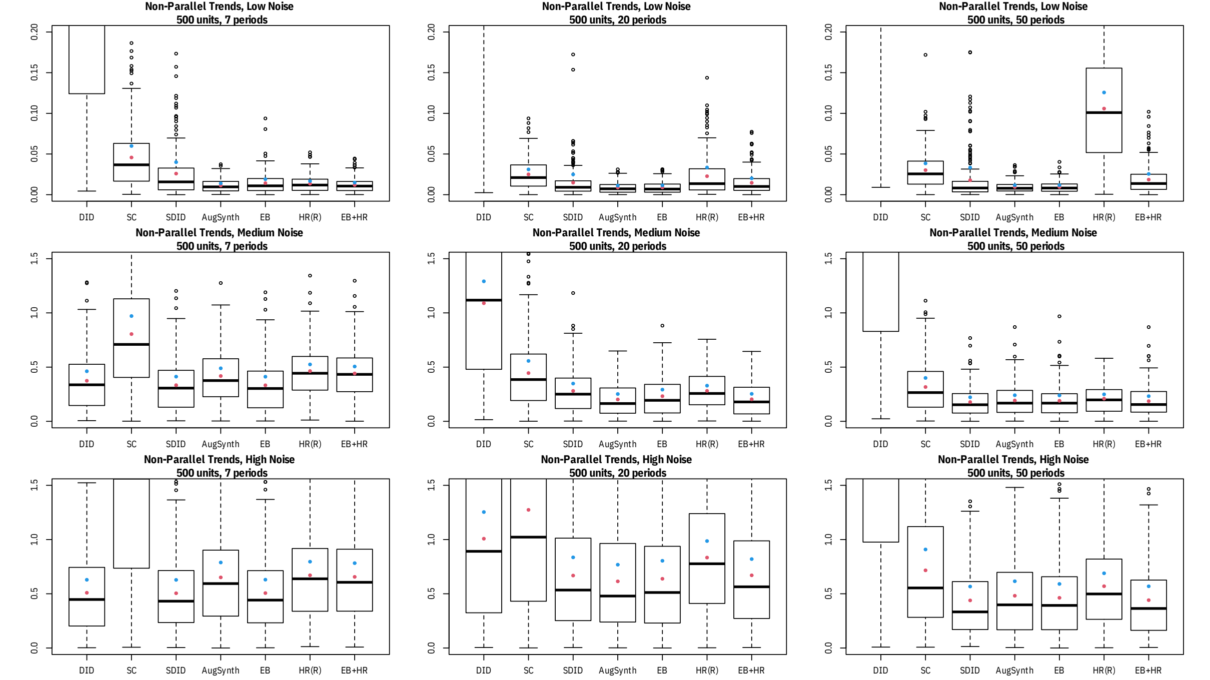
- N units, T periods
- Single unknown factor: $\mu_i \sim \mathcal{N}\left(i/N - 0.5, 0.5\right)$
- Treatment: $W_i \sim \text{Bern}\left(\Lambda(\mu_i)\right)$
- Outcome:
  - **parallel trends**: $Y_{it} = \mu_i + 0.1t + \varepsilon_i, \varepsilon_i \sim \mathcal{N}\left(0, \sigma\right)$
  - **time trends**: $Y_{it} = \mu_i \alpha_t t + \varepsilon_i, \varepsilon_i \sim \mathcal{N}\left(0, \sigma\right), \alpha_t \sim \text{U}\left[l, u\right]$
  - Later: ARIMA with dynamics in both Y_{it}, ε_{it}
- Estimand: ATT in the last period (true effect is 0)

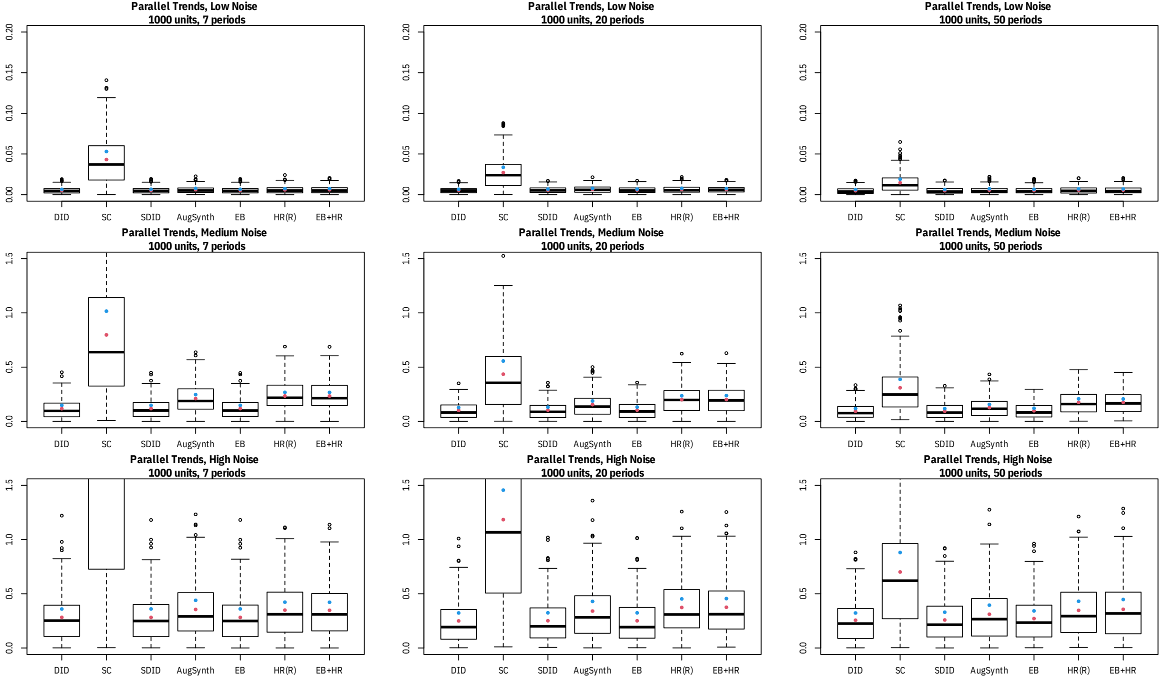
Panel Simulation Setup

- N units, T periods
- Single unknown factor: $\mu_i \sim \mathcal{N}(i/N - 0.5, 0.5)$
- Treatment: $W_i \sim$ Bern $(\Lambda(\mu_i))$
- Outcome:
  - **parallel trends**: $Y_{it} = \mu_i + 0.1t + \varepsilon_i, \varepsilon_i \sim \mathcal{N}(0, \sigma)$
  - **time trends**: $Y_{it} = \mu_i \alpha_t t + \varepsilon_i, \varepsilon_i \sim \mathcal{N}(0, \sigma), \alpha_t \sim$ U $[l, u]$
  - Later: ARIMA with dynamics in both Y_{it}, ε_{it}
- Estimand: ATT in the last period (true effect is 0)

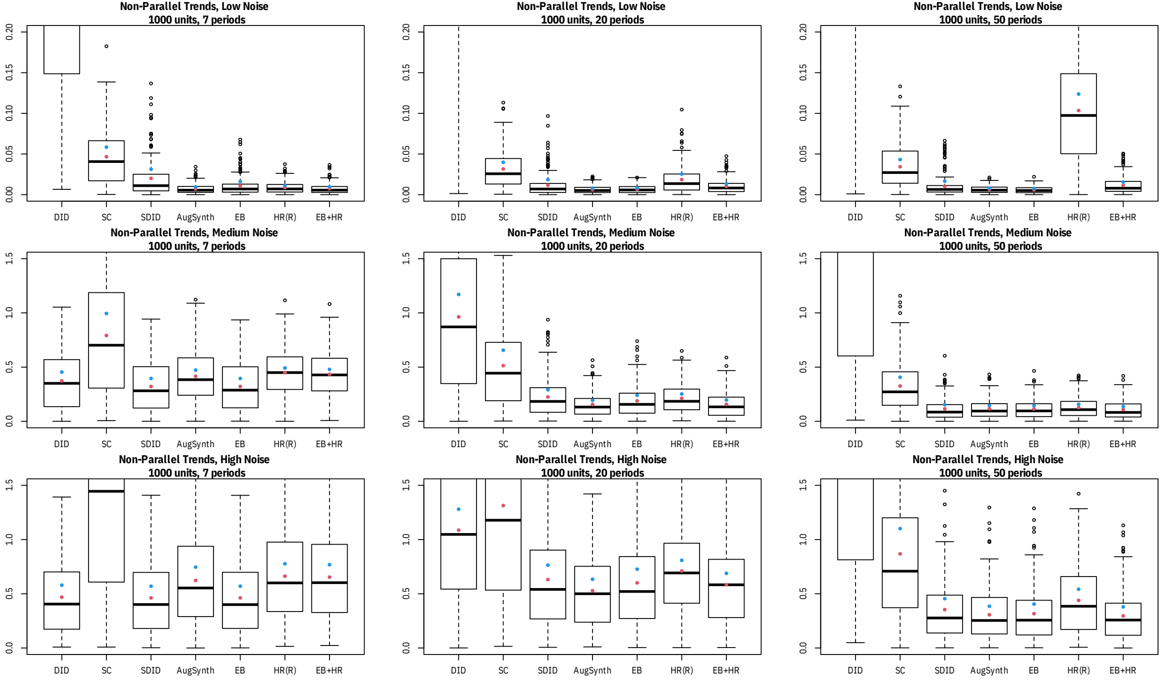



Parallel Trends, Low Noise
500 units, 7 periods
Parallel Trends, Low Noise
500 units, 20 periods
Parallel Trends, Low Noise
500 units, 50 periods
Parallel Trends, Medium Noise
500 units, 7 periods
Parallel Trends, Medium Noise
500 units, 20 periods
Parallel Trends, Medium Noise
500 units, 50 periods
Parallel Trends, High Noise
500 units, 7 periods
Parallel Trends, High Noise
500 units, 20 periods
Parallel Trends, High Noise
500 units, 50 periods
0.00 0.05 0.10 0.15 0.20
0.0 0.5 1.0 1.5
DID SC SDID AugSynth EB HR(R) EB+HR


Non-Parallel Trends, Low Noise
500 units, 7 periods
Non-Parallel Trends, Low Noise
500 units, 20 periods
Non-Parallel Trends, Low Noise
500 units, 50 periods
Non-Parallel Trends, Medium Noise
500 units, 7 periods
Non-Parallel Trends, Medium Noise
500 units, 20 periods
Non-Parallel Trends, Medium Noise
500 units, 50 periods
Non-Parallel Trends, High Noise
500 units, 7 periods
Non-Parallel Trends, High Noise
500 units, 20 periods
Non-Parallel Trends, High Noise
500 units, 50 periods
DID
SC
SDID
AugSynth
EB
HR(R)
EB+HR



Parallel Trends, Low Noise
1000 units, 7 periods
Parallel Trends, Low Noise
1000 units, 20 periods
Parallel Trends, Low Noise
1000 units, 50 periods
Parallel Trends, Medium Noise
1000 units, 7 periods
Parallel Trends, Medium Noise
1000 units, 20 periods
Parallel Trends, Medium Noise
1000 units, 50 periods
Parallel Trends, High Noise
1000 units, 7 periods
Parallel Trends, High Noise
1000 units, 20 periods
Parallel Trends, High Noise
1000 units, 50 periods
DID
SC
SDID
AugSynth
EB
HR(R)
EB+HR


Non-Parallel Trends, Low Noise
1000 units, 7 periods
Non-Parallel Trends, Low Noise
1000 units, 20 periods
Non-Parallel Trends, Low Noise
1000 units, 50 periods
Non-Parallel Trends, Medium Noise
1000 units, 7 periods
Non-Parallel Trends, Medium Noise
1000 units, 20 periods
Non-Parallel Trends, Medium Noise
1000 units, 50 periods
Non-Parallel Trends, High Noise
1000 units, 7 periods
Non-Parallel Trends, High Noise
1000 units, 20 periods
Non-Parallel Trends, High Noise
1000 units, 50 periods
0.00
0.05
0.10
0.15
0.20
0.0
0.5
1.0
1.5
DID
SC
SDID
AugSynth
EB
HR(R)
EB+HR

The choice of loss function for Panel Balancing

$$\min_{\gamma \in \Delta} \overbrace{h_\zeta(\mathbf{X}_1 - \mathbf{X}_0'\gamma)}^{\text{Balance}} + \sum_{i \in \mathcal{C}} \overbrace{f(\gamma_i)}^{\text{Dispersion}}$$

$h_\zeta(\cdot)$ fixed to L_2
Penalty for dispersion $f(\cdot)$ consequential

- SC has no penalisation
- SDID has theoretically motivated penalisation
- EB penalises *deviation from uniform weights*: interpolates between DiD and balancing

DGP with factor structure, $N = 500, T = 10$; perfect fit

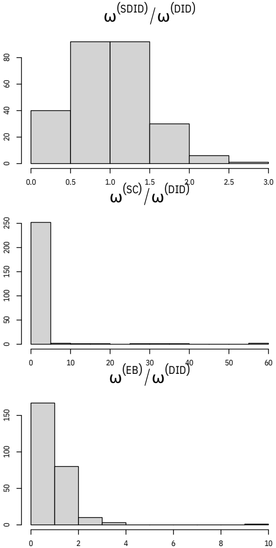
The choice of loss function for Panel Balancing

$$\min_{\gamma \in \Delta} \overbrace{h_\zeta(\mathbf{X}_1 - \mathbf{X}_0'\gamma)}^{\text{Balance}} + \sum_{i \in \mathcal{C}} \overbrace{f(\gamma_i)}^{\text{Dispersion}}$$

$h_\zeta(\cdot)$ fixed to L_2
Penalty for dispersion $f(\cdot)$ consequential

- SC has no penalisation
- SDID has theoretically motivated penalisation
- EB penalises *deviation from uniform weights*: interpolates between DiD and balancing

DGP with factor structure,
$N = 500, T = 10$; perfect fit



The choice of loss function for Panel Balancing

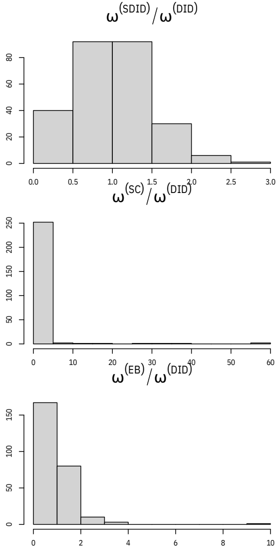
$$\min_{\gamma\in\Delta} \overbrace{h_\zeta(\mathbf{X}_1 - \mathbf{X}_0'\gamma)}^{\text{Balance}} + \sum_{i\in\mathcal{C}} \overbrace{f(\gamma_i)}^{\text{Dispersion}}$$

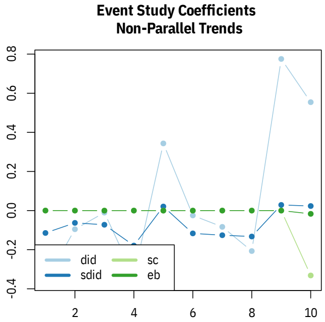
$h_\zeta(\cdot)$ fixed to L_2

Penalty for dispersion $f(\cdot)$ consequential

- SC has no penalisation
- SDID has theoretically motivated penalisation
- EB penalises *deviation from uniform weights*: interpolates between DiD and balancing

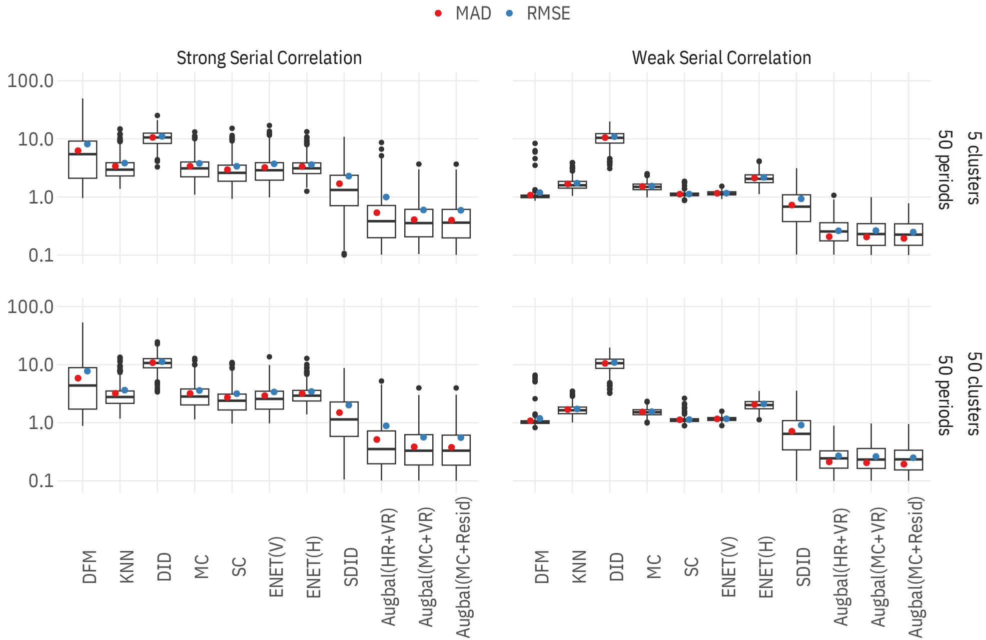
DGP with factor structure, $N = 500, T = 10$; perfect fit





dilution of weights behaviour from Ferman (2021)

factor recovery | If parallel trends holds


MAD
RMSE
Strong Serial Correlation
Weak Serial Correlation
100.0
10.0
1.0
0.1
5 clusters
50 periods
50 clusters
50 periods
DFM
KNN
DID
MC
SC
ENET(V)
ENET(H)
SDID
Augbal(HR+VR)
Augbal(MC+VR)
Augbal(MC+Resid)

Conclusion

- Flexible models can be judiciously used with reweighting methods to improve the robustness of our estimates to misspecification
- Increasing consensus on adopting a hybrid structure of combining a performant outcome model with weights that explicitly target sample balance
  - No feedback is a very strong assumption in panel settings (reversals are common)
  - Double robustness is heuristic in this setting, since assignment mechanism isn't directly modelled
  - Progress: Arkhangelsky and Imbens(2022, 2023), Arkhangelsky et al (2023)

Conclusion

- Flexible models can be judiciously used with reweighting methods to improve the robustness of our estimates to misspecification
- Increasing consensus on adopting a hybrid structure of combining a performant outcome model with weights that explicitly target sample balance
  - No feedback is a very strong assumption in panel settings (reversals are common)
  - Double robustness is heuristic in this setting, since assignment mechanism isn't directly modelled
  - Progress: Arkhangelsky and Imbens(2022, 2023), Arkhangelsky et al (2023)
- We propose a common framework for these 'augmented balancing' estimators in three popular designs and perform extensive simulation studies to show that they weakly outperform standard estimators (including AIPW), and provide heuristic understanding of when gains are likely to be particularly large
- Forthcoming R package `abal` that uses analogously modular construction to pair flexible outcome models with a fast and numerically stable estimation procedure for balancing weights

[1] Alberto Abadie. "Semiparametric Difference-in-Differences Estimators". en. In: *The Review of economic studies* 72.1 (Jan. 2005), pp. 1–19. URL: https://academic.oup.com/restud/article-abstract/72/1/1/1581053.

[2] Alberto Abadie, Alexis Diamond, and Jens Hainmueller. "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program". In: *Journal of the American statistical Association* 105.490 (2010), pp. 493–505.

[3] Dmitry Arkhangelsky et al. "Synthetic Difference in Differences". In: *American Economic Review* (2020).

[4] Susan Athey et al. "Matrix Completion Methods for Causal Panel Data Models". In: *Journal of the American Statistical Association* (2021).

[5] Eli Ben-Michael, Avi Feller, David A Hirshberg, et al. "The Balancing Act in Causal Inference". In: (Oct. 2021). arXiv: 2110.14831 [stat.ME]. URL: http://arxiv.org/abs/2110.14831.

[6] Eli Ben-Michael, Avi Feller, and Jesse Rothstein. "The augmented synthetic control method". In: *Journal of the American Statistical Association* 116.536 (2021), pp. 1789–1803.

[7] Kirill Borusyak, Xavier Jaravel, and Jann Spiess. "Revisiting Event Study Designs: Robust and Efficient Estimation". In: *Available at SSRN 2826228* (2022).

[8] David Bruns-Smith et al. "Augmented balancing weights as undersmoothed regressions". In: (2023).

[9] Qizhao Chen, Vasilis Syrgkanis, and Morgane Austern. "Debiased Machine Learning without Sample-Splitting for Stable Estimators". In: (June 2022). arXiv: 2206.01825 [econ.EM]. URL: http://arxiv.org/abs/2206.01825.

[10] Victor Chernozhukov, Denis Chetverikov, et al. "Double/debiased machine learning for treatment and structural parameters". In: *The econometrics journal* 21.1 (Feb. 2018). URL: http://doi.wiley.com/10.1111/ectj.12097.

[11] Victor Chernozhukov, Whitney K Newey, and Rahul Singh. "Automatic Debiased Machine Learning of Causal and Structural Effects". In: *Econometrica: journal of the Econometric Society* 90.3 (May 2022), pp. 967–1027. URL: https://doi.org/10.3982/ECTA18515.