The Augmented Synthetic Control Method

Eli Ben-Michael, Avi Feller, and Jesse Rothstein

UC Berkeley

Econometric Society European Winter Meeting 2018
December 5, 2018

Synthetic Control Method (SCM)

> "Arguably the most important innovation in the policy evaluation literature in the last 15 years"
>
> [Athey and Imbens, 2017]

Synthetic Control Method (SCM)

> "Arguably the most important innovation in
> the policy evaluation literature in the last 15 years"
>
> [Athey and Imbens, 2017]

Abadie and Gardeazabal [2003]; Abadie et al. [2010, 2015] $\rightarrow$ **4,000+ citations**

Synthetic Control Method (SCM)

> "Arguably the most important innovation in
> the policy evaluation literature in the last 15 years"
>
> [Athey and Imbens, 2017]

Abadie and Gardeazabal [2003]; Abadie et al. [2010, 2015] $\rightarrow$ **4,000+ citations**

Allows estimation where strong research designs have not been available:

- Panels with relatively few units
- Only a single treated unit, potentially with strong selection
- No explicit model of the time series or of the selection process

Synthetic Control Method (SCM)

> “Arguably the most important innovation in
> the policy evaluation literature in the last 15 years”
>
> [Athey and Imbens, 2017]

Abadie and Gardeazabal [2003]; Abadie et al. [2010, 2015] $\rightarrow$ **4,000+ citations**

Allows estimation where strong research designs have not been available:

- Panels with relatively few units
- Only a single treated unit, potentially with strong selection
- No explicit model of the time series or of the selection process

Poorly understood, gap between theory and practice

Our Paper

SCM is Biased

- Curse of dimensionality $\Rightarrow$ SCM is biased in practical settings

Our Paper

SCM is Biased

- Curse of dimensionality $\Rightarrow$ SCM is biased in practical settings

Augmenting SCM

- With an outcome model, estimate the bias due to covariate imbalance and adjust

Our Paper

SCM is Biased

- Curse of dimensionality $\Rightarrow$ SCM is biased in practical settings

Augmenting SCM

- With an outcome model, estimate the bias due to covariate imbalance and adjust

SCM is Inverse Propensity Weighting (IPW)

- SCM implicitly fits a regularized propensity score model

Synthetic Controls

But First, Notation...

- Observe N units over T time periods
- Unit $i = 1$ is treated at time $t = T_0 = T - 1$ *
- Units $i = 2, \ldots, N$ are never treated
- W_i is the treatment indicator

* Unit $i = 1$ can be an *average* of N_1 treated units

But First, Notation...

- Observe N units over T time periods
- Unit $i = 1$ is treated at time $t = T_0 = T - 1$ *
- Units $i = 2, \ldots, N$ are never treated
- W_i is the treatment indicator

$$\begin{pmatrix} Y_{11} & \ldots & Y_{1T_0} & \color{red}{Y_{1T}(1)} \\ Y_{21} & \ldots & Y_{2T_0} & Y_{2T}(0) \\ \vdots & & & \vdots \\ \underbrace{Y_{N1} \quad \ldots \quad Y_{NT_0}}_{\text{pre-treatment outcomes}} & & & Y_{NT}(0) \end{pmatrix}$$

* Unit $i = 1$ can be an *average* of N_1 treated units

But First, Notation...

- Observe N units over T time periods
- Unit $i = 1$ is treated at time $t = T_0 = T - 1$ *
- Units $i = 2, \ldots, N$ are never treated
- W_i is the treatment indicator

$$\left(\begin{array}{cccc} Y_{11} & \ldots & Y_{1T_0} & Y_{1T}(1) \\ Y_{21} & \ldots & Y_{2T_0} & Y_{2T}(0) \\ \vdots & & & \vdots \\ \underbrace{Y_{N1} \quad \ldots \quad Y_{NT_0}}_{\text{pre-treatment outcomes}} & & & Y_{NT}(0) \end{array}\right) \equiv \left(\begin{array}{ccc|c} X_{11} & \ldots & X_{1T_0} & Y_1 \\ \hline X_{21} & \ldots & X_{2T_0} & Y_2 \\ \vdots & & & \vdots \\ \underbrace{X_{N1} \quad \ldots \quad X_{NT_0}}_{\text{pre-treatment outcomes}} & & & Y_N \end{array}\right) \equiv \left(\begin{array}{c|c} X_{1\cdot} & Y_1 \\ \hline X_{0\cdot} & Y_0 \end{array}\right)$$

* Unit $i = 1$ can be an *average* of N_1 treated units

Synthetic Control Method

SCM weights $\hat{\gamma}^{\text{scm}}$ minimize L^2 imbalance with treated unit

$$\begin{aligned} \min_{\gamma} \quad & \|X_{1\cdot} - X_{0\cdot}'\gamma\|_2^2 \\ \text{subject to} \quad & \sum_{i=2}^{N} \gamma_i = 1 \\ & \gamma_i \geq 0 \end{aligned}$$

$$\hat{Y}_1^{\text{scm}}(0) = Y_0'\hat{\gamma}^{\text{scm}}$$

$$\hat{\tau}^{\text{scm}} = Y_1 - Y_0'\hat{\gamma}^{\text{scm}}$$

Synthetic Control Method

SCM weights $\hat{\gamma}^{\text{scm}}$ minimize L^2 imbalance with treated unit

$$\begin{aligned} \min_{\gamma} \quad & \|X_{1\cdot} - X_{0\cdot}'\gamma\|_2^2 \\ \text{subject to} \quad & \sum_{i=2}^{N} \gamma_i = 1 \\ & \gamma_i \geq 0 \end{aligned}$$

$$\hat{Y}_1^{\text{scm}}(0) = Y_0'\hat{\gamma}^{\text{scm}}$$

$$\hat{\tau}^{\text{scm}} = Y_1 - Y_0'\hat{\gamma}^{\text{scm}}$$

Suppressing some details:

- *Constrained regression* formulation [Doudchenko and Imbens, 2017]

SCM in Theory and Practice

In theory [Abadie et al., 2010]

- Assume Y_{it} follows a factor model: $Y_{it} = \sum_{j=1}^{J} \phi_{ij}\mu_{jt} + \varepsilon_{it}$
- Assume $\hat{\gamma}^{\text{scm}}$ achieves exact balance even as T_0 grows
- Then SCM bias decreases as T_0 increases
- Intuition: For large T_0, can only balance observed X by balancing latent ϕ.

SCM in Theory and Practice

In theory [Abadie et al., 2010]

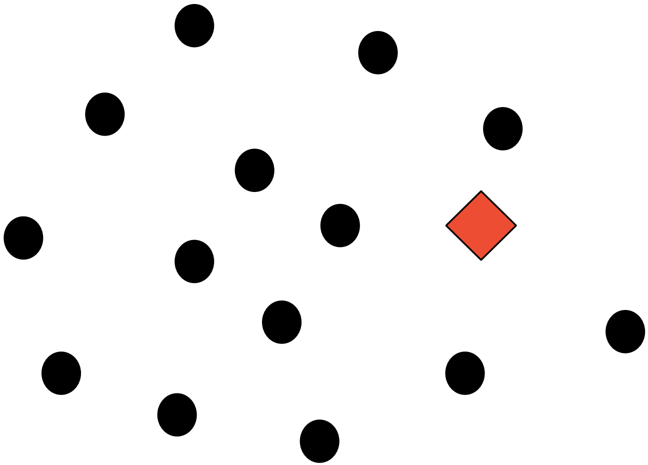
- Assume Y_{it} follows a factor model: $Y_{it} = \sum_{j=1}^{J} \phi_{ij}\mu_{jt} + \varepsilon_{it}$
- Assume $\hat{\gamma}^{\text{scm}}$ achieves exact balance even as T_0 grows
- Then SCM bias decreases as T_0 increases
- Intuition: For large T_0, can only balance observed X by balancing latent ϕ.

In practice

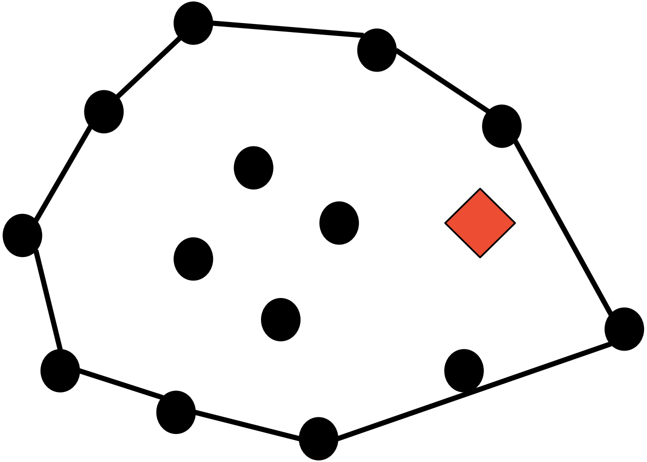
- T_0 is typically larger than or on the same order of N
- Exact balance is elusive
- Abadie et al. [2015] recommend against using SCM when
  - "the pre-treatment fit is poor or the number of pre-treatment periods is small"

SCM is biased

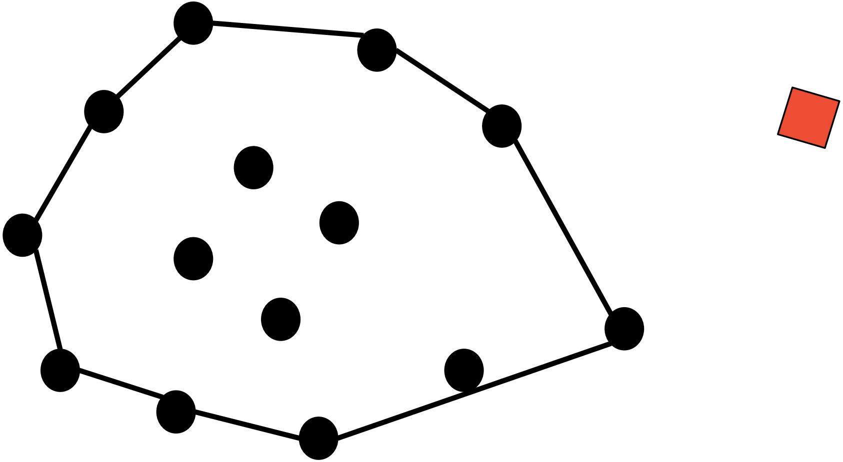
Inside the Convex Hull



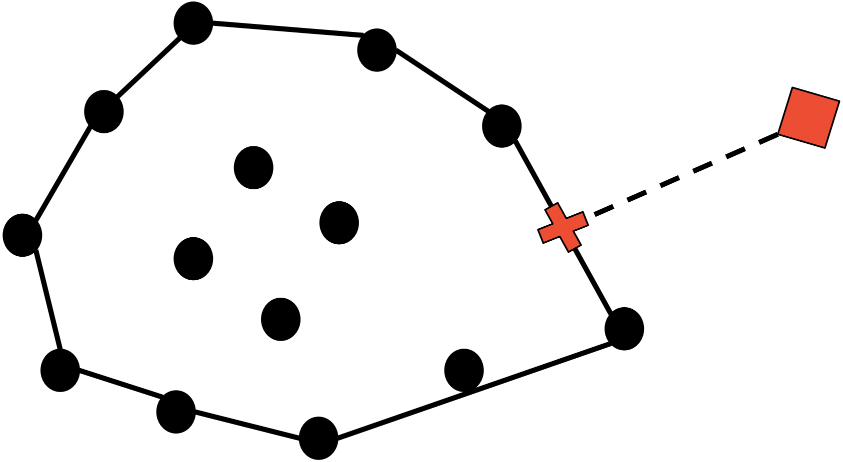
Inside the Convex Hull



Curse of Dimensionality ⇒ Unlikely to be in Convex Hull



Curse of Dimensionality ⇒ Unlikely to be in Convex Hull



Estimating the Bias

Fit a working model for the prognostic score

$$\hat{m}(X_i): \quad Y_i(0) \sim X_i$$

Estimating the Bias

Fit a working model for the prognostic score

$$\hat{m}(X_i): \quad Y_i(0) \sim X_i$$

Estimate bias

$$\hat{m}(X_1) - \sum_{W_i=0} \hat{\gamma}_i \hat{m}(X_i)$$

Estimating the Bias

Fit a working model for the prognostic score

$$\hat{m}(X_i): \quad Y_i(0) \sim X_i$$

Estimate bias

$$\hat{m}(X_1) - \sum_{W_i=0} \hat{\gamma}_i \hat{m}(X_i)$$

Examples:

- Regularized linear model, `gsynth` [Xu, 2017] and matrix completion [Athey et al., 2017]
- Black box prediction algorithms

Estimating the Bias

Fit a working model for the prognostic score

$$\hat{m}(X_i): \quad Y_i(0) \sim X_i$$

Estimate bias

$$\hat{m}(X_1) - \sum_{W_i=0} \hat{\gamma}_i \hat{m}(X_i)$$

Examples:

- Regularized linear model, `gsynth` [Xu, 2017] and matrix completion [Athey et al., 2017]
- Black box prediction algorithms

Related to within-time placebo balance check [Abadie et al., 2010]

Augmented Synthetic Control Method

Augmented Synthetic Control Method (ASCM)

Adjust SCM for estimated bias:

$$\hat{Y}_1^{\text{aug}}(0) = \underbrace{\sum_{W_i=0} \hat{\gamma}_i Y_i}_{\text{SCM estimate}} + \underbrace{\hat{m}(X_1) - \sum_{W_i=0} \hat{\gamma}_i \hat{m}(X_i)}_{\text{Estimate of bias}} \qquad \text{(Bias Correction)}$$

Augmented Synthetic Control Method (ASCM)

Adjust SCM for estimated bias:

$$\hat{Y}_1^{\text{aug}}(0) = \underbrace{\sum_{W_i=0} \hat{\gamma}_i Y_i}_{\text{SCM estimate}} + \underbrace{\hat{m}(X_1) - \sum_{W_i=0} \hat{\gamma}_i \hat{m}(X_i)}_{\text{Estimate of bias}} \quad \text{(Bias Correction)}$$

$$= \underbrace{\hat{m}(X_1)}_{\text{Outcome model}} + \underbrace{\sum_{W_i=0} \hat{\gamma}_i (Y_i - \hat{m}(X_i))}_{\text{Re-weight residuals}} \quad \text{(Augmented IPW)}$$

Augmented Synthetic Control Method (ASCM)

Adjust SCM for estimated bias:

$$\hat{Y}_1^{\text{aug}}(0) = \underbrace{\sum_{W_i=0} \hat{\gamma}_i Y_i}_{\text{SCM estimate}} + \underbrace{\hat{m}(X_1) - \sum_{W_i=0} \hat{\gamma}_i \hat{m}(X_i)}_{\text{Estimate of bias}} \qquad \text{(Bias Correction)}$$

$$= \underbrace{\hat{m}(X_1)}_{\text{Outcome model}} + \underbrace{\sum_{W_i=0} \hat{\gamma}_i (Y_i - \hat{m}(X_i))}_{\text{Re-weight residuals}} \qquad \text{(Augmented IPW)}$$

Deep connections to existing methods:

- Model assisted survey sampling [Cassel et al., 1976; Breidt and Opsomer, 2017]
- Approximate residual balancing [Athey et al., 2018; Tan, 2018]

Ridge ASCM is a Weighting Estimator

$$\hat{Y}_1^{\text{aug}}(0) = \underbrace{\sum_{W_i=0} {\color{red}\hat{\gamma}_i^{\text{scm}}} Y_i}_{\text{SCM estimate}} + \underbrace{\left(X_{1\cdot} - \sum_{W_i=0} {\color{red}\hat{\gamma}_i^{\text{scm}}} X_{i\cdot}\right) \cdot \hat{\eta}}_{\text{Estimate of bias}}$$

Ridge ASCM is a Weighting Estimator

$$\hat{Y}_1^{\text{aug}}(0) = \underbrace{\sum_{W_i=0} \hat{\gamma}_i^{\text{scm}} Y_i}_{\text{SCM estimate}} + \underbrace{\left(X_{1\cdot} - \sum_{W_i=0} \hat{\gamma}_i^{\text{scm}} X_{i\cdot} \right) \cdot \hat{\eta}}_{\text{Estimate of bias}} = \sum_{W_i=0} \hat{\gamma}_i^{\text{aug}} Y_i$$

Ridge ASCM is a Weighting Estimator

$$\hat{Y}_1^{\text{aug}}(0) = \underbrace{\sum_{W_i=0} \hat{\gamma}_i^{\text{scm}} Y_i}_{\text{SCM estimate}} + \underbrace{\left(X_{1\cdot} - \sum_{W_i=0} \hat{\gamma}_i^{\text{scm}} X_{i\cdot}\right) \cdot \hat{\eta}}_{\text{Estimate of bias}} = \sum_{W_i=0} \hat{\gamma}_i^{\text{aug}} Y_i$$

Adjust the SCM weights

$$\hat{\gamma}_i^{\text{aug}} = \hat{\gamma}_i^{\text{scm}} + \underbrace{(X_1 - X_{0\cdot}'\hat{\gamma}^{\text{scm}})'(X_{0\cdot}'X_{0\cdot} + \lambda I_{T_0})^{-1} X_{i\cdot}}_{\text{Adjust SCM weights for better balance}}$$

Ridge ASCM is a Weighting Estimator

$$\hat{Y}_1^{\text{aug}}(0) = \underbrace{\sum_{W_i=0} \hat{\gamma}_i^{\text{scm}} Y_i}_{\text{SCM estimate}} + \underbrace{\left(X_{1\cdot} - \sum_{W_i=0} \hat{\gamma}_i^{\text{scm}} X_{i\cdot} \right) \cdot \hat{\eta}}_{\text{Estimate of bias}} = \sum_{W_i=0} \hat{\gamma}_i^{\text{aug}} Y_i$$

Adjust the SCM weights

$$\hat{\gamma}_i^{\text{aug}} = \hat{\gamma}_i^{\text{scm}} + \underbrace{(X_1 - X_{0\cdot}' \hat{\gamma}^{\text{scm}})'(X_{0\cdot}' X_{0\cdot} + \lambda I_{T_0})^{-1} X_{i\cdot}}_{\text{Adjust SCM weights for better balance}}$$

Achieve better balance: $\|X_1 - X_{0\cdot}' \hat{\gamma}^{\text{aug}}\|_2 < \|X_1 - X_{0\cdot}' \hat{\gamma}^{\text{scm}}\|_2$

Ridge ASCM is a Weighting Estimator

$$\hat{Y}_1^{\text{aug}}(0) = \underbrace{\sum_{W_i=0} \hat{\gamma}_i^{\text{scm}} Y_i}_{\text{SCM estimate}} + \underbrace{\left(X_{1\cdot} - \sum_{W_i=0} \hat{\gamma}_i^{\text{scm}} X_{i\cdot}\right) \cdot \hat{\eta}}_{\text{Estimate of bias}} = \sum_{W_i=0} \hat{\gamma}_i^{\text{aug}} Y_i$$

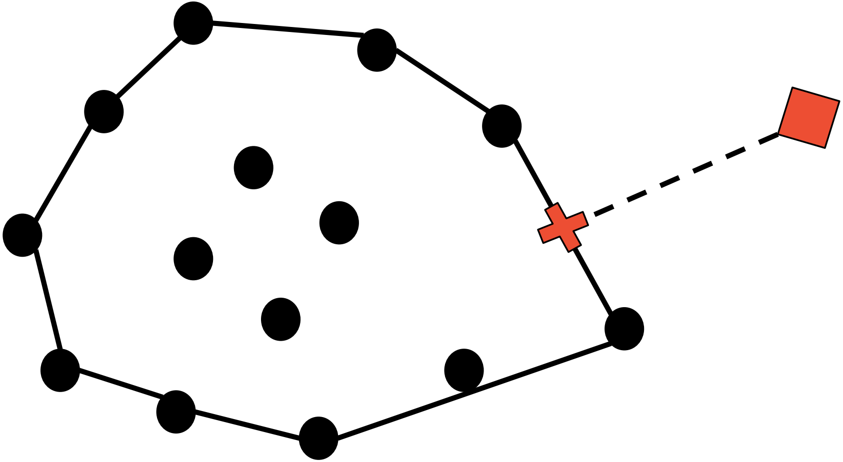
Adjust the SCM weights

$$\hat{\gamma}_i^{\text{aug}} = \hat{\gamma}_i^{\text{scm}} + \underbrace{(X_1 - X_{0\cdot}'\hat{\gamma}^{\text{scm}})'(X_{0\cdot}'X_{0\cdot} + \lambda I_{T_0})^{-1} X_{i\cdot}}_{\text{Adjust SCM weights for better balance}}$$

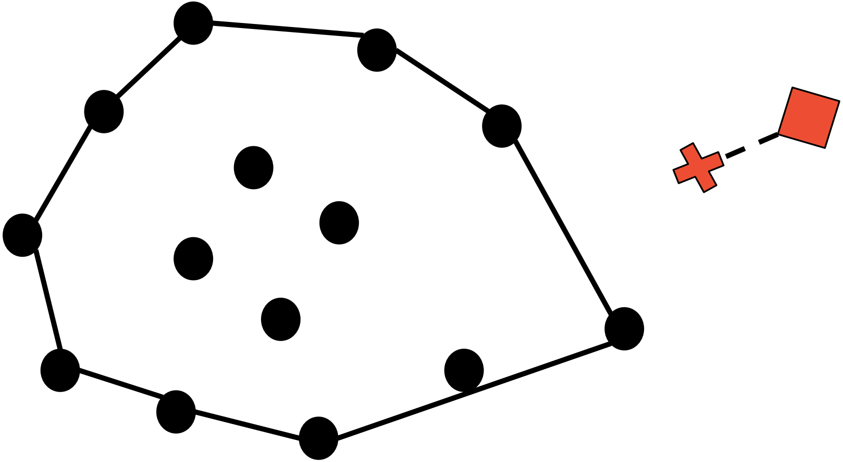
Achieve better balance: $\|X_1 - X_{0\cdot}'\hat{\gamma}^{\text{aug}}\|_2 < \|X_1 - X_{0\cdot}'\hat{\gamma}^{\text{scm}}\|_2$

...but higher variance and possibly negative weights

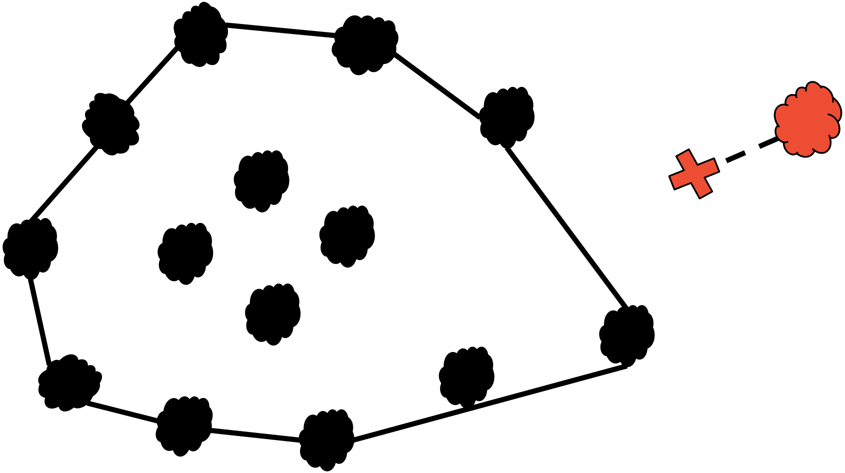
Ridge ASCM Extrapolates



Ridge ASCM Extrapolates



Noisy Proxies



Bias under a linear factor model

Under a linear factor model, $Y_{it} = \sum \phi_{ij}\mu_{jt} + \varepsilon_{it}$

SCM bias $\propto$ imbalance in ϕ

Bias under a linear factor model

Under a linear factor model, $Y_{it} = \sum \phi_{ij}\mu_{jt} + \varepsilon_{it}$

SCM bias $\propto$ imbalance in ϕ

SCM bias $\leq$ imbalance in X (RMSE) $+$ approximation error

Bias under a linear factor model

Under a linear factor model, $Y_{it} = \sum \phi_{ij}\mu_{jt} + \varepsilon_{it}$

SCM bias $\propto$ imbalance in ϕ

SCM bias $\leq$ imbalance in X (RMSE) $+$ approximation error

- Recovers Abadie et al. [2010] result for perfect balance
- Ridge ASCM has **lower bias** than SCM or ridge alone
- Related to Ferman and Pinto [2018]: SCM doesn't balance ϕ as $T_0 \to \infty$

SCM = IPW

Penalize SCM to Ensure a Unique Solution

Add a **dispersion penalty** [Abadie et al., 2015; Abadie and L'Hour, 2018]

$$\min_{\gamma} \quad \frac{1}{2\zeta}\|X_{1\cdot} - X_{0\cdot}'\gamma\|_2^2$$

$$\text{subject to} \quad \sum_{W_i=0} \gamma_i = 1$$

$$\gamma_i \geq 0 \quad i = 2, \ldots, N$$

Penalize SCM to Ensure a Unique Solution

Add a **dispersion penalty** [Abadie et al., 2015; Abadie and L'Hour, 2018]

$$\min_{\gamma} \quad \frac{1}{2\zeta}\|X_{1\cdot} - X_{0\cdot}'\gamma\|_2^2 + \sum_{W_i=0} \gamma_i \log \gamma_i$$

$$\text{subject to} \quad \sum_{W_i=0} \gamma_i = 1$$

$$\gamma_i \geq 0 \quad i = 2, \ldots, N$$

Penalize SCM to Ensure a Unique Solution

Add a **dispersion penalty** [Abadie et al., 2015; Abadie and L'Hour, 2018]

$$\min_{\gamma} \quad \frac{1}{2\zeta}\|X_{1\cdot} - X_{0\cdot}'\gamma\|_2^2 + \sum_{W_i=0} \gamma_i \log \gamma_i$$

$$\text{subject to} \quad \sum_{W_i=0} \gamma_i = 1$$

$$\gamma_i \geq 0 \quad i = 2, \ldots, N$$

- Many possible penalties [Doudchenko and Imbens, 2017; Robbins et al., 2017].
- When unpenalized SCM weights are unique, equivalent for sufficiently small ζ.

Dual: Inverse Propensity Score Weighting

Implicit estimate of propensity score model with **ridge regularization**

$$\min_{\alpha,\beta} \underbrace{\sum_{W_i=0} \exp(\alpha + \beta' X_{i\cdot}) - (\alpha + \beta' X_{1\cdot})}_{\text{Calibration loss}}$$

Dual: Inverse Propensity Score Weighting

Implicit estimate of propensity score model with **ridge regularization**

$$\min_{\alpha,\beta} \underbrace{\sum_{W_i=0} \exp(\alpha + \beta' X_{i\cdot}) - (\alpha + \beta' X_{1\cdot})}_{\text{Calibration loss}} + \underbrace{\frac{\zeta}{2}\|\beta\|_2^2}_{\text{Regularization}}$$

Dual: Inverse Propensity Score Weighting

Implicit estimate of propensity score model with **ridge regularization**

$$\min_{\alpha,\beta} \underbrace{\sum_{W_i=0} \exp(\alpha + \beta' X_{i\cdot}) - (\alpha + \beta' X_{1\cdot})}_{\text{Calibration loss}} + \underbrace{\frac{\zeta}{2}\|\beta\|_2^2}_{\text{Regularization}}$$

Weights are odds of treatment (ATT weights)

$$\hat{\gamma}_i = \exp(\hat{\alpha} + \hat{\beta}' X_{i\cdot}) = \frac{\text{logit}^{-1}(\hat{\alpha} + \hat{\beta}' X_{i\cdot})}{1 - \text{logit}^{-1}(\hat{\alpha} + \hat{\beta}' X_{i\cdot})} = \frac{\hat{\pi}(X_{i\cdot})}{1 - \hat{\pi}(X_{i\cdot})}$$

Dual: Inverse Propensity Score Weighting

Implicit estimate of propensity score model with **ridge regularization**

$$\min_{\alpha,\beta} \underbrace{\sum_{W_i=0} \exp(\alpha + \beta' X_{i\cdot}) - (\alpha + \beta' X_{1\cdot})}_{\text{Calibration loss}} + \underbrace{\frac{\zeta}{2}\|\beta\|_2^2}_{\text{Regularization}}$$

Weights are odds of treatment (ATT weights)

$$\hat{\gamma}_i = \exp(\hat{\alpha} + \hat{\beta}' X_{i\cdot}) = \frac{\text{logit}^{-1}(\hat{\alpha} + \hat{\beta}' X_{i\cdot})}{1 - \text{logit}^{-1}(\hat{\alpha} + \hat{\beta}' X_{i\cdot})} = \frac{\hat{\pi}(X_{i\cdot})}{1 - \hat{\pi}(X_{i\cdot})}$$

Calibrated propensity score estimation [Graham et al., 2012; Zhao and Percival, 2017; Tan, 2017]

Implications for inference

Typical inference for SCM: **uniform permutation** of placebo estimates

- Estimate *placebo gap* for each unit, $Y_i - \tilde{Y}_i$
- Compare observed gap to unweighted dist. of gaps
- Many interpretations in the literature [Ando and Sävje, 2013; Hahn and Shi, 2017]

Implications for inference

Typical inference for SCM: **uniform permutation** of placebo estimates

- Estimate *placebo gap* for each unit, $Y_i - \tilde{Y}_i$
- Compare observed gap to unweighted dist. of gaps
- Many interpretations in the literature [Ando and Sävje, 2013; Hahn and Shi, 2017]

Implicit propensity score model $\rightarrow$ randomization test

- From this perspective, uniform permutation is invalid

Implications for inference

Typical inference for SCM: **uniform permutation** of placebo estimates

- Estimate *placebo gap* for each unit, $Y_i - \tilde{Y}_i$
- Compare observed gap to unweighted dist. of gaps
- Many interpretations in the literature [Ando and Sävje, 2013; Hahn and Shi, 2017]

Implicit propensity score model $\rightarrow$ randomization test

- From this perspective, uniform permutation is invalid
- *In theory*: weighted permutation test

Implications for inference

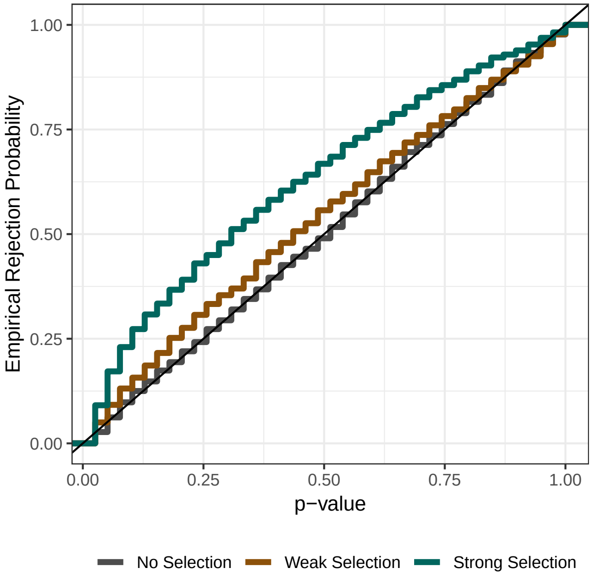
Typical inference for SCM: **uniform permutation** of placebo estimates

- Estimate *placebo gap* for each unit, $Y_i - \tilde{Y}_i$
- Compare observed gap to unweighted dist. of gaps
- Many interpretations in the literature [Ando and Sävje, 2013; Hahn and Shi, 2017]

Implicit propensity score model $\rightarrow$ randomization test

- From this perspective, uniform permutation is invalid
- *In theory:* weighted permutation test
- *In practice:* approach is infeasible
  - Lousy estimates of $\hat{\pi}$ in SCM settings
  - Typically only a few units have positive $\hat{\pi}$, so $p < 0.05$ is impossible

Uniform Permutation is Invalid Under Selection



Our Approach: Model Based Inference

A generic outcome mode with independent, additive, homoscedastic noise ε_{it}:

$$Y_{it}(0) = m_{it} + \varepsilon_{it}$$

Our Approach: Model Based Inference

A generic outcome mode with independent, additive, homoscedastic noise ε_{it}:

$$Y_{it}(0) = m_{it} + \varepsilon_{it}$$

Variance of ATT estimate is:

$$\mathsf{Var}_{\varepsilon_T}(\hat{\tau}) = \mathsf{Var}_{\varepsilon_T}\left(Y_{1T} - \sum_{W_i=0} \hat{\gamma}_i Y_{iT}\right) = \left(\frac{1}{N_1} + \|\hat{\gamma}\|_2^2\right)\mathsf{Var}(\varepsilon_{iT})$$

Our Approach: Model Based Inference

A generic outcome mode with independent, additive, homoscedastic noise ε_{it}:

$$Y_{it}(0) = m_{it} + \varepsilon_{it}$$

Variance of ATT estimate is:

$$\text{Var}_{\varepsilon_T}(\hat{\tau}) = \text{Var}_{\varepsilon_T}\left(Y_{1T} - \sum_{W_i=0} \hat{\gamma}_i Y_{iT}\right) = \left(\frac{1}{N_1} + \|\hat{\gamma}\|_2^2\right)\text{Var}(\varepsilon_{iT})$$

Get a conservative estimate variance with placebo distribution

$$\widehat{\text{Var}}(\varepsilon_{iT}) = \frac{1}{N_0}\sum_{W_i=0}(Y_i - \tilde{Y}_i)^2$$

Our Approach: Model Based Inference

A generic outcome mode with independent, additive, homoscedastic noise ε_{it}:

$$Y_{it}(0) = m_{it} + \varepsilon_{it}$$

Variance of ATT estimate is:

$$\mathsf{Var}_{\varepsilon_T}(\hat{\tau}) = \mathsf{Var}_{\varepsilon_T}\left(Y_{1T} - \sum_{W_i=0} \hat{\gamma}_i Y_{iT}\right) = \left(\frac{1}{N_1} + \|\hat{\gamma}\|_2^2\right)\mathsf{Var}(\varepsilon_{iT})$$

Get a conservative estimate variance with placebo distribution

$$\widehat{\mathsf{Var}}(\varepsilon_{iT}) = \frac{1}{N_0}\sum_{W_i=0}(Y_i - \tilde{Y}_i)^2$$

Fundamentally hard problem, difficult to estimate σ_1 without homoscedasticity

Simulations

Evaluation with Calibrated Simulations

Linear Factor model with unit and time fixed effects:

$$Y_{it}(0) = \alpha_i + \nu_t + \sum_{j=1}^{J} \phi_{ij}\mu_{jt} + \varepsilon_{it}.$$

Fit parameters with `gsynth` [Xu, 2017] and sample from empirical distributions

Evaluation with Calibrated Simulations

Linear Factor model with unit and time fixed effects:

$$Y_{it}(0) = \alpha_i + \nu_t + \sum_{j=1}^{J} \phi_{ij}\mu_{jt} + \varepsilon_{it}.$$

Fit parameters with `gsynth` [Xu, 2017] and sample from empirical distributions

Model selection into treatment based on factors

$$\text{logit}^{-1} P(W_i = 1 \mid \alpha_i, \phi_i) = \theta\left(\alpha_i + \sum_{j=1}^{J} \phi_i\right)$$

Evaluation with Calibrated Simulations

Linear Factor model with unit and time fixed effects:

$$Y_{it}(0) = \alpha_i + \nu_t + \sum_{j=1}^{J} \phi_{ij}\mu_{jt} + \varepsilon_{it}.$$

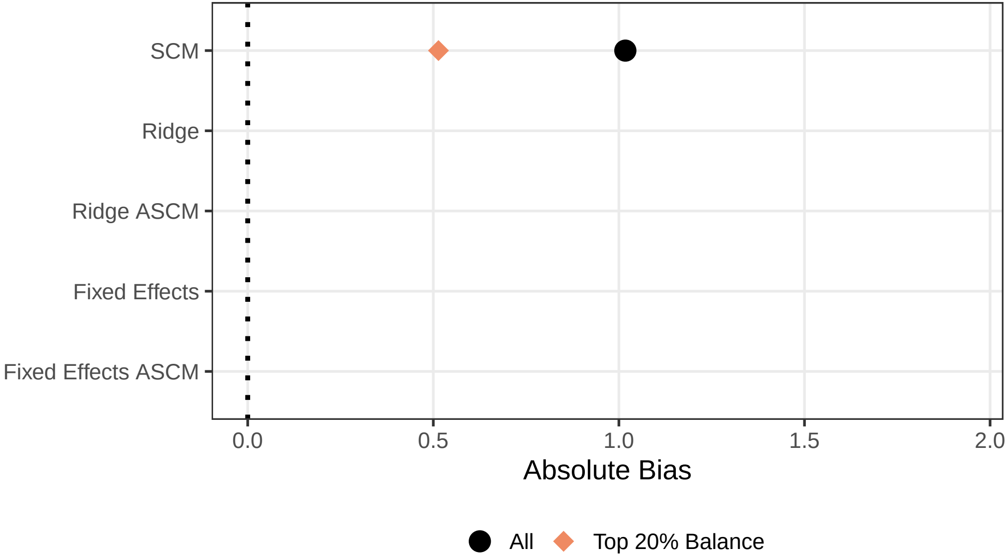
Fit parameters with `gsynth` [Xu, 2017] and sample from empirical distributions

Model selection into treatment based on factors

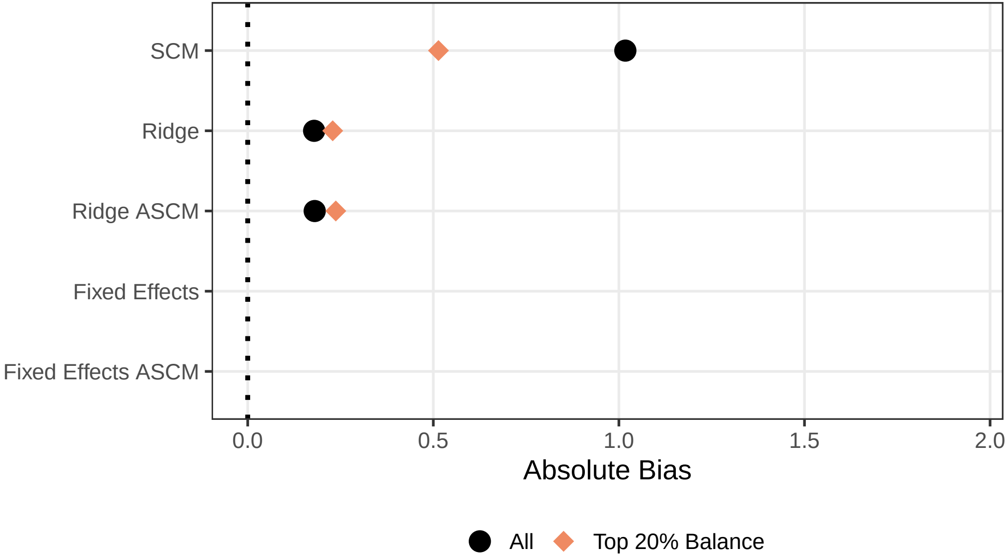
$$\text{logit}^{-1} P(W_i = 1 \mid \alpha_i, \phi_i) = \theta \left(\alpha_i + \sum_{j=1}^{J} \phi_i \right)$$

No treatment effect whatsoever

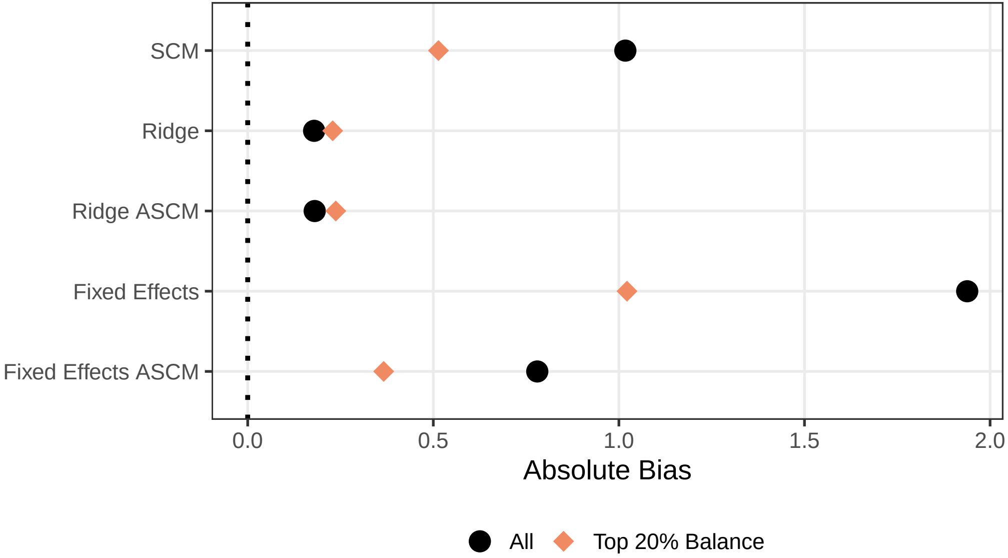
SCM is Biased



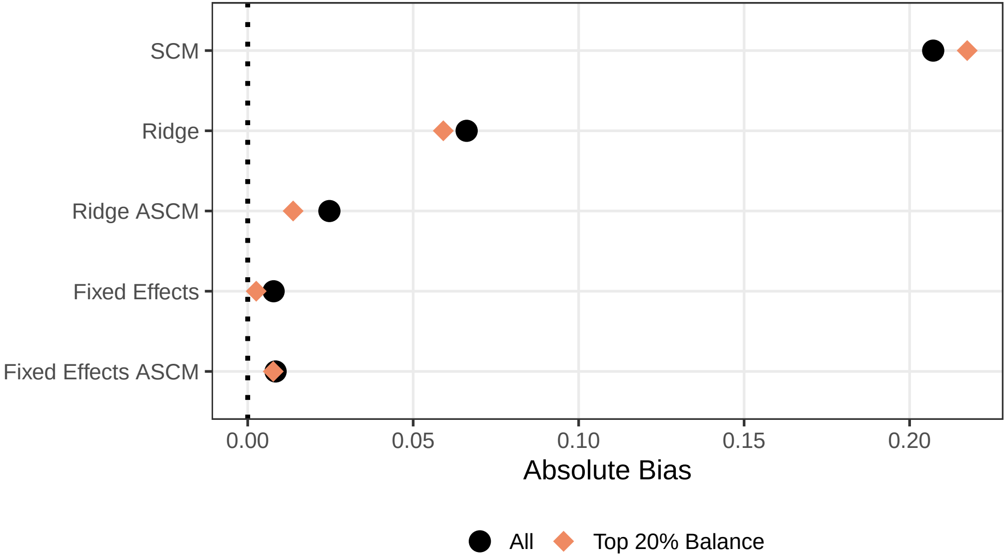
Ridge/Ridge ASCM is Less Biased



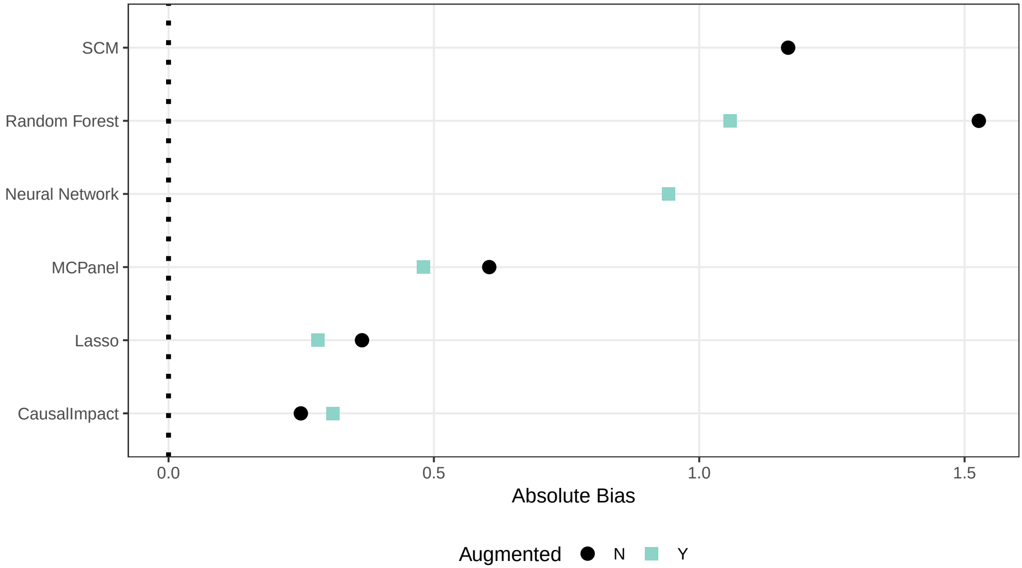
Augmentation Helps When Outcome Model is Only OK



Bias Under Fixed Effects



Flexible Outcome Models



Conclusion

Understanding SCM

- Pre-treatment imbalance is linked to bias
- In applied settings imbalance can be large so SCM is biased
- IPW perspective connects to wider balancing weights literature and informs testing

Augmenting SCM

- Account for imbalance and adjust
- Reduces bias at cost of extrapolation and slight increase in variance
- Can incorporate flexible ML and panel data methods, penalized regression works well
- Can also incorporate auxiliary covariates
- `R` implementation: `augsynth`

Thank you!

arxiv.org/abs/1811.04170

ebenmichael.github.io

References

References I

Abadie, A., Diamond, A., and Hainmueller, J. (2010). Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statistical Association*, 105(490):493–505.

Abadie, A., Diamond, A., and Hainmueller, J. (2015). Comparative Politics and the Synthetic Control Method. *American Journal of Political Science*, 59(2):495–510.

Abadie, A. and Gardeazabal, J. (2003). The Economic Costs of Conflict: A Case Study of the Basque Country. *The American Economic Review*, 93(1):113–132.

Abadie, A. and L'Hour, J. (2018). A penalized synthetic control estimator for disaggregated data.

Ando, M. and Sävje, F. (2013). Hypothesis testing with the synthetic control method.

Athey, S., Bayati, M., Doudchenko, N., Imbens, G., and Khosravi, K. (2017). Matrix Completion Methods for Causal Panel Data Models. *arxiv 1710.10251*.

Athey, S. and Imbens, G. W. (2017). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives*, 31(2):3–32.

References II

Athey, S., Imbens, G. W., and Wager, S. (2018). Approximate Residual Balancing: De-Biased Inference of Average Treatment Effects in High Dimensions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*.

Breidt, F. J. and Opsomer, J. D. (2017). Model-Assisted Survey Estimation with Modern Prediction Techniques. *Statistical Science*, 32(2):190–205.

Cassel, C. M., Sarndal, C.-E., and Wretman, J. H. (1976). Some results on generalized difference estimation and generalized regression estimation for finite populations. *Biometrika*, 63(3):615–620.

Doudchenko, N. and Imbens, G. W. (2017). Difference-In-Differences and Synthetic Control Methods: A Synthesis. *arxiv 1610.07748*.

Ferman, B. and Pinto, C. (2018). Synthetic controls with imperfect pre-treatment fit.

Graham, B. S., de Xavier Pinto, C. C., and Egel, D. (2012). Inverse probability tilting for moment condition models with missing data. *The Review of Economic Studies*, 79(3):1053–1079.

Hahn, J. and Shi, R. (2017). Synthetic control and inference. *Econometrics*, 5(4):52.

References III

Hainmueller, J. (2011). Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. *Political Analysis*, 20:25–46.

Kline, P. (2011). Oaxaca-Blinder as a reweighting estimator. In *American Economic Review*, volume 101, pages 532–537.

Robbins, M., Saunders, J., and Kilmer, B. (2017). A Framework for Synthetic Control Methods With High-Dimensional, Micro-Level Data: Evaluating a Neighborhood-Specific Crime Intervention. *Journal of the American Statistical Association*, 112(517):109–126.

Tan, Z. (2017). Regularized calibrated estimation of propensity scores with model misspecification and high-dimensional data.

Tan, Z. (2018). Model-assisted inference for treatment effects using regularized calibrated estimation with high-dimensional data. *arXiv preprint arXiv:1801.09817*.

Wang, Y. and Zubizarreta, J. R. (2018). Minimal Approximately Balancing Weights: Asymptotic Properties and Practical Considerations.

Xu, Y. (2017). Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models. *Political Analysis*, 25:57–76.

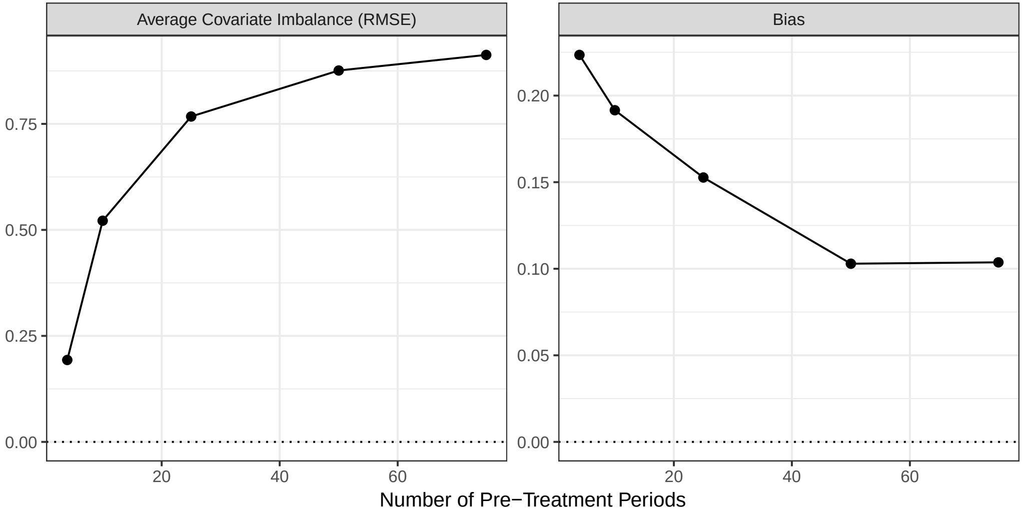
References IV

Zhao, Q. and Percival, D. (2017). Entropy balancing is doubly robust. *Journal of Causal Inference*, 5(1).

Zubizarreta, J. R. (2015). Stable Weights that Balance Covariates for Estimation With Incomplete Outcome Data. *Journal of the American Statistical Association*, 110(511):910–922.

Appendix

A Long Long Time Period Doesn't Fix the Bias



Auxiliary Covariates

Original formulation and practical applications have auxiliary covariates Z_i

- Default procedure: use Z in p-score, tune to balance pre-treatment outcomes X_i
- IPW perspective: include Z with X in p-score and outcome models
- Alternatively, balance X with SCM and fit outcome model with Z

Auxiliary Covariates

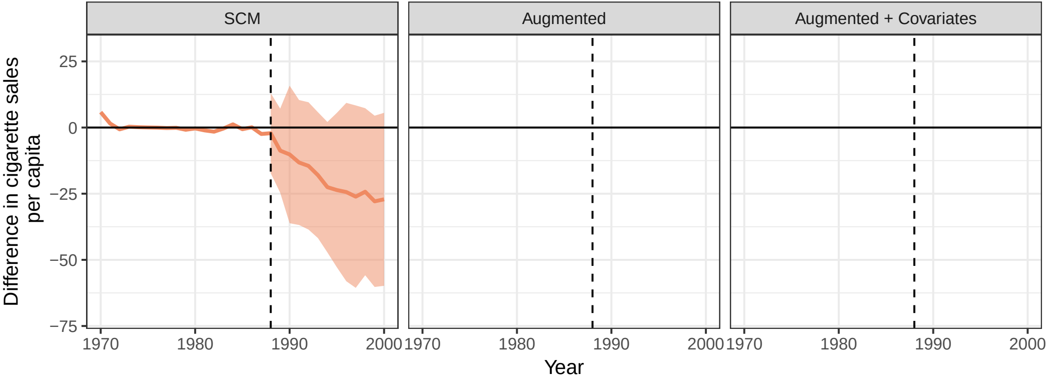
Original formulation and practical applications have auxiliary covariates Z_i

- Default procedure: use Z in p-score, tune to balance pre-treatment outcomes X_i
- IPW perspective: include Z with X in p-score and outcome models
- Alternatively, balance X with SCM and fit outcome model with Z

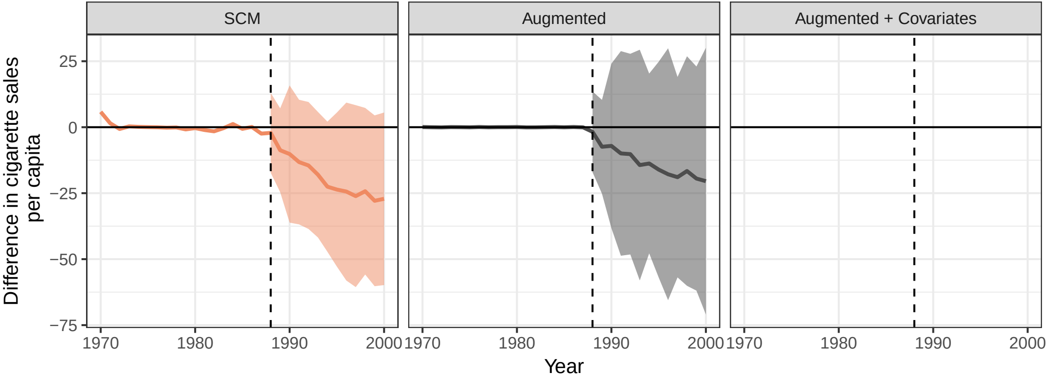
Partitioned regression approach:

- Regress Y and X on Z, get residuals $\check{Y} = Y - \hat{Y}$, $\check{X} = X - \hat{X}$
- Fit (A)SCM with residuals and get estimate $\check{Y}_1 - \check{Y}_0'\hat{\gamma}$
- This is ASCM with an OLS outcome model on $Z \Rightarrow$ also a weighting estimator

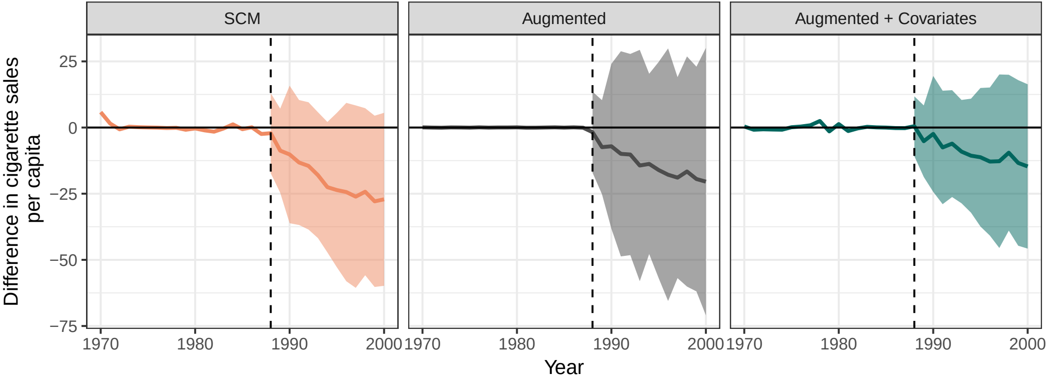
California Prop. 99



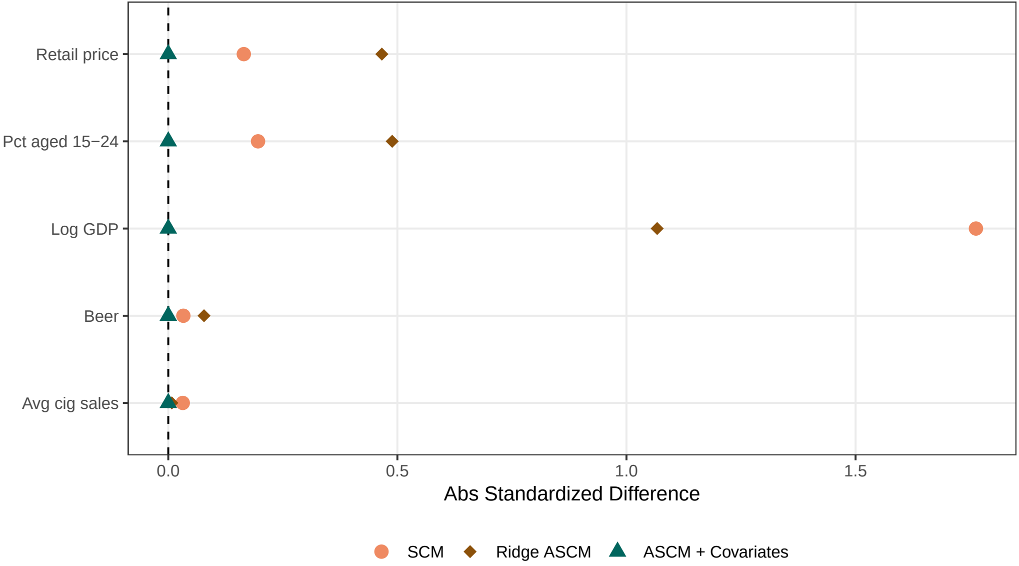
California Prop. 99



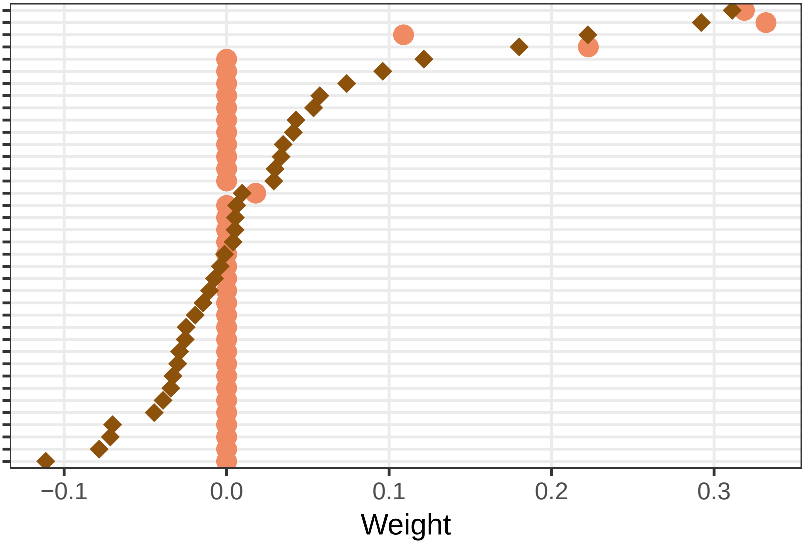
California Prop. 99



Prop 99: Auxiliary Covariate Balance



Prop 99: Weights



Bias for a Weighting Estimator (Linear Model)

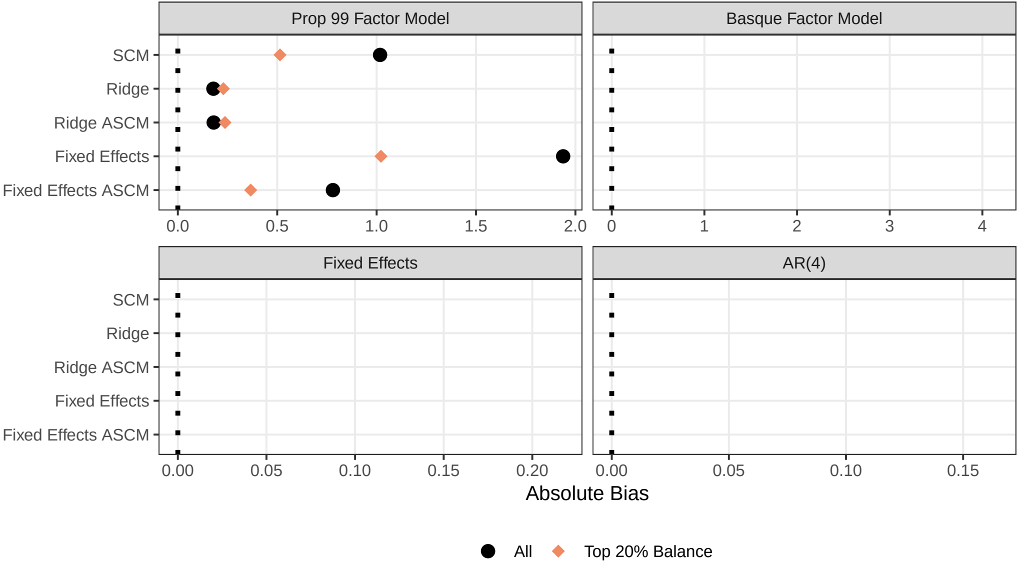
Under linearity in lagged outcomes (e.g. AR(k))

$$Y_{it} = \sum_{j=t-1-k}^{t-1} \beta_j Y_{ij} + \varepsilon_{it}$$

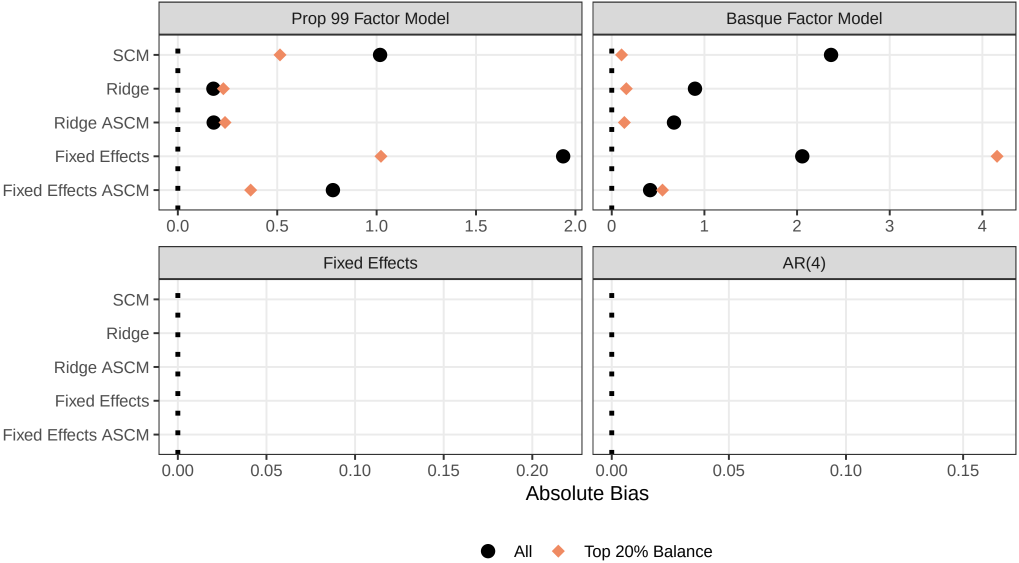
Then bias scales with imbalance in lagged outcomes:

$$\mathbb{E}_{\varepsilon_T}\left[Y_1 - Y_0'\gamma\right] = \sum_{j=t-1-k}^{t-1} \beta_j \left(X_{1j} - X_{0j}'\gamma\right) \leq \|\beta\|_2 \|X_{1\cdot} - X_{0\cdot}\|_2$$

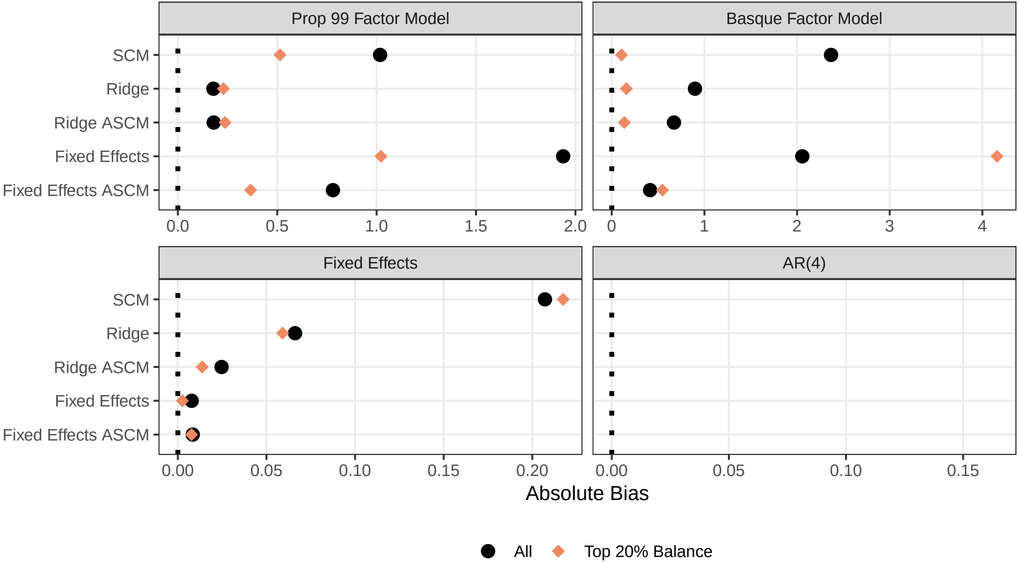
More Simulations



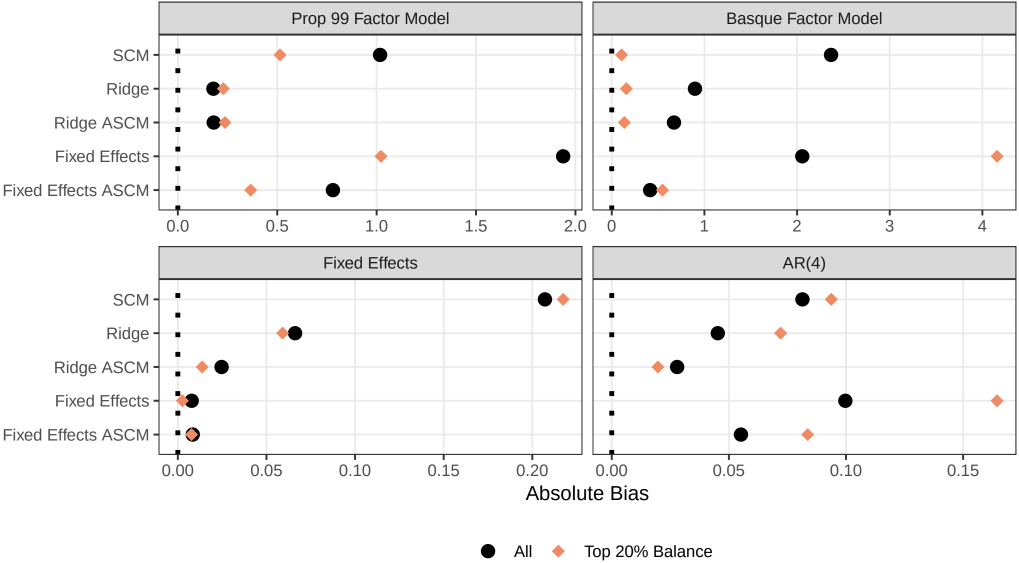
More Simulations



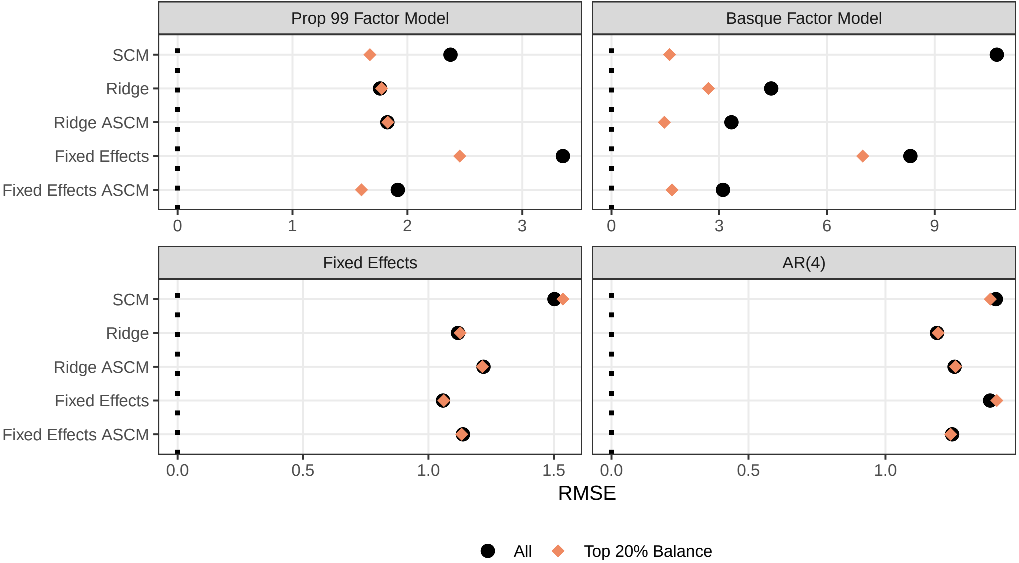
More Simulations



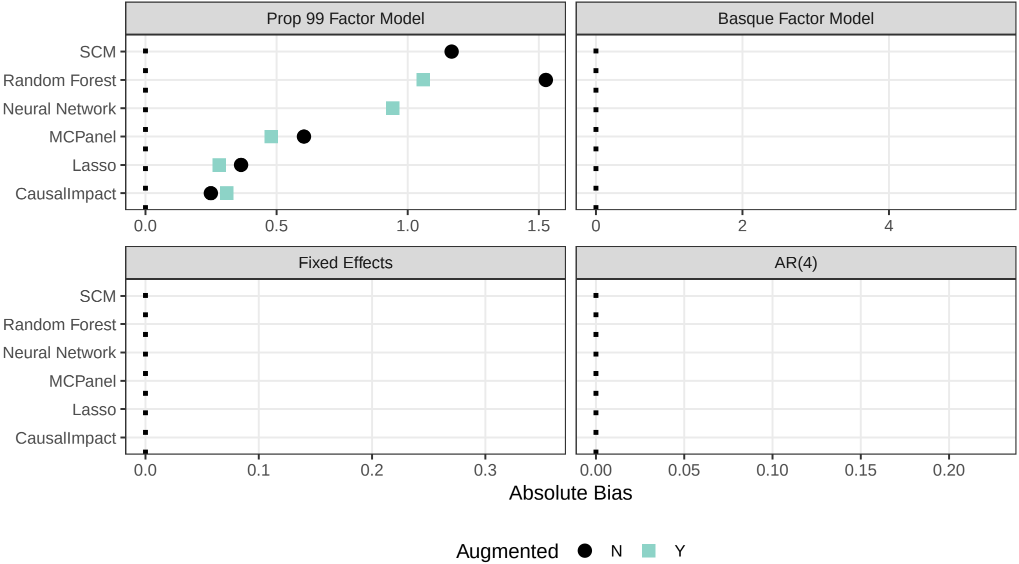
More Simulations



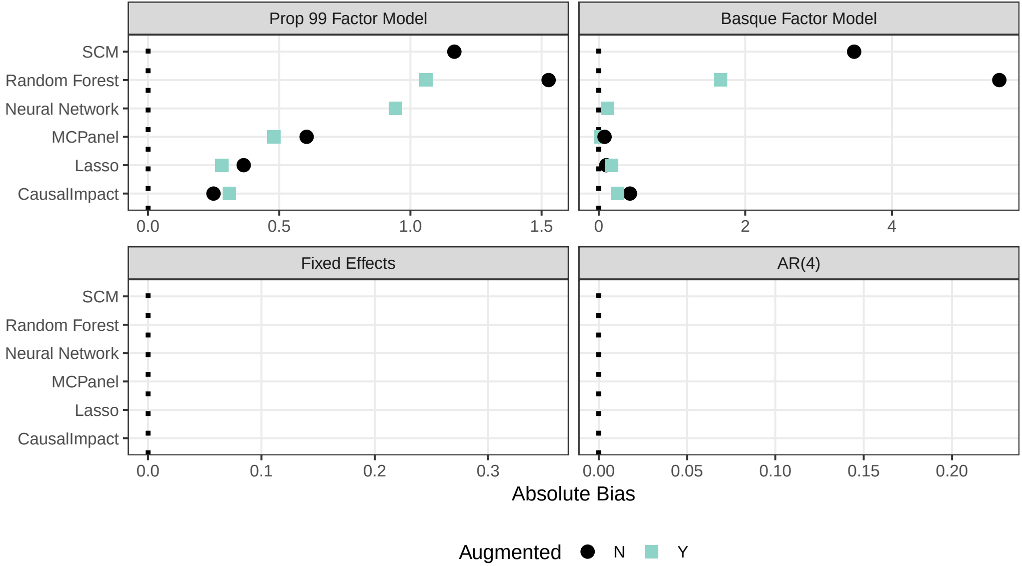
RMSE


Prop 99 Factor Model
Basque Factor Model
Fixed Effects
AR(4)
SCM
Ridge
Ridge ASCM
Fixed Effects
Fixed Effects ASCM
0
1
2
3
0
3
6
9
0.0
0.5
1.0
1.5
0.0
0.5
1.0
RMSE
All
Top 20% Balance

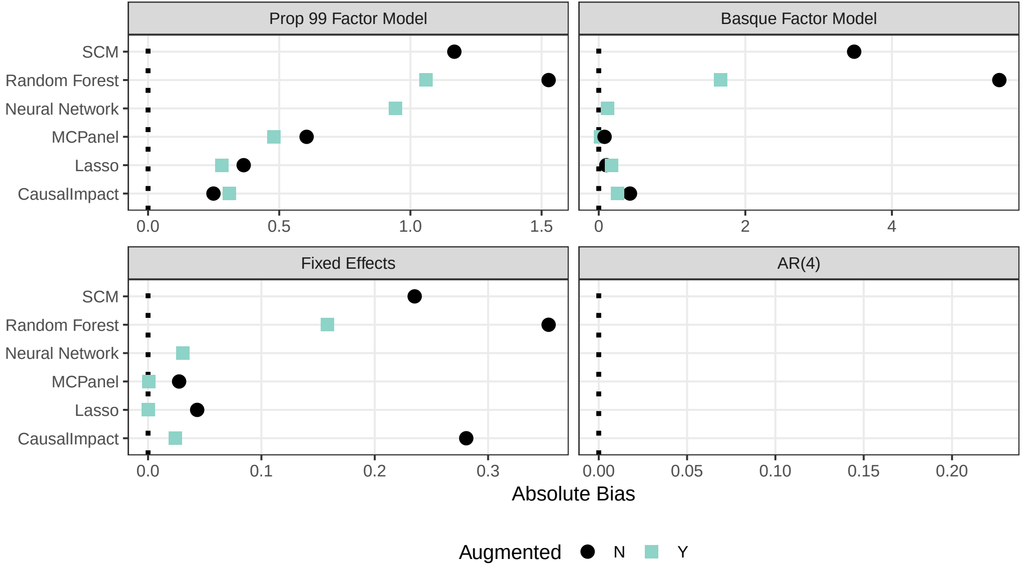
Flexible ML and Modern Panel Methods



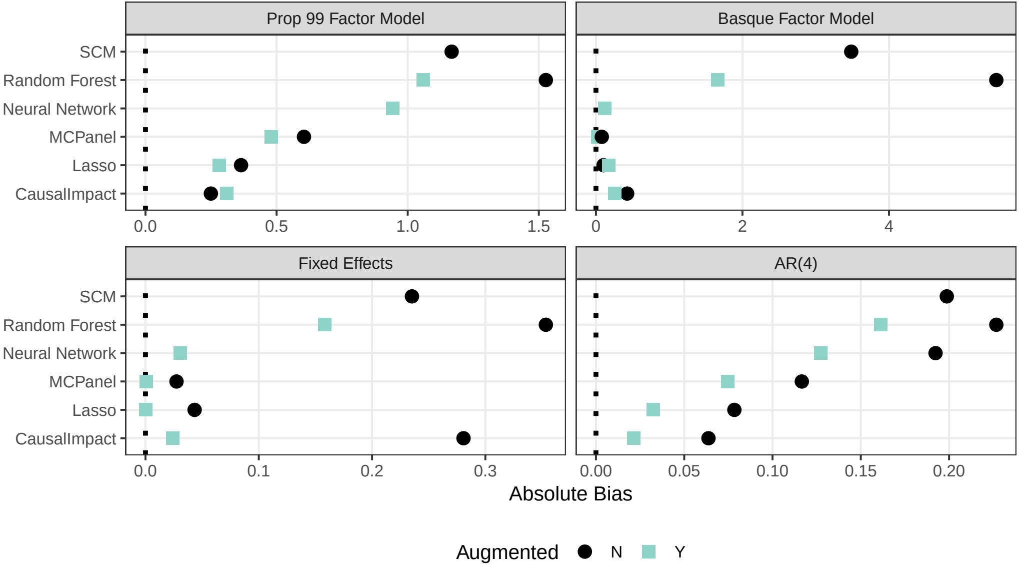
Flexible ML and Modern Panel Methods



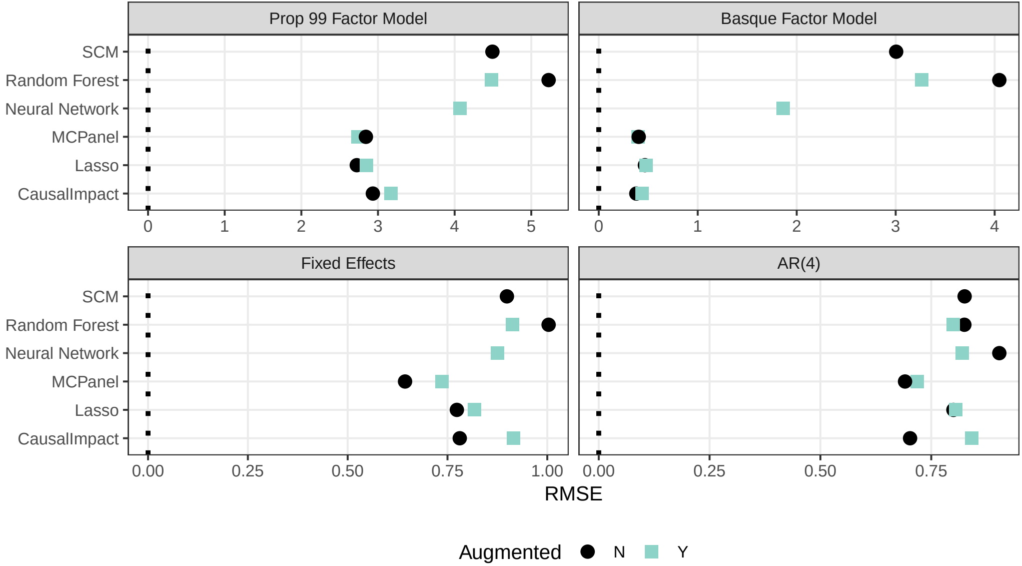
Flexible ML and Modern Panel Methods



Flexible ML and Modern Panel Methods



Flexible ML and Modern Panel Methods



General Duality (1)

General balancing weights problem has:

- Dispersion measure $f : \mathbb{R} \to \mathbb{R}$
  - Entropy penalty: $f(\gamma_i) = \gamma_i \log \gamma_i$ [Hainmueller, 2011; Robbins et al., 2017]
  - 2-Norm: $f(\gamma_i) = \frac{1}{2}\gamma_i^2$ [Zubizarreta, 2015]
  - Elastic net: $f(\gamma_i) = \frac{1-\alpha}{2}\gamma_i^2 + \alpha|\gamma_i|$ [Doudchenko and Imbens, 2017]
- Balance criterion $h : \mathbb{R}^{T_0} \to \bar{\mathbb{R}}$
  - L^∞ constraint: $h(x) = \begin{cases} 0 & \|x\|_\infty \leq \lambda \\ \infty & \|x\|_\infty > \lambda \end{cases}$ [Wang and Zubizarreta, 2018; Athey et al., 2018]
  - L^2 penalty: $h(x) = \frac{1}{2}\|x\|_2^2$

$$\min_{\gamma} \; h\left(X_{1\cdot} - X_{0\cdot}\right) + \sum_{W_i=0} f(\gamma_i)$$

General Duality (2)

Dual view as p-score estimator

- Dispersion measure controls odds function $f^{*\prime}(\cdot)$
  - Entropy penalty $\Rightarrow$ exponential odds $f^{*\prime}(X'\beta) = \exp(X'\beta)$ [Zhao and Percival, 2017]
  - 2-Norm $\Rightarrow$ linear odds $f^{*\prime}(X'\beta) = X'\beta$ [Kline, 2011]
- Balance criterion controls regularization $h^*(\cdot)$
  - L^∞ constraint $\Rightarrow$ Lasso penalty: $h^*(\beta) = \lambda\|\beta\|_1$
  - L^2 penalty $\Rightarrow$ ridge penalty: $h^*(\beta) = \frac{\lambda}{2}\|\beta\|_2^2$

$$\min_{\alpha,\beta} \sum_{W_i=0} f^*(\alpha + \beta' X_i) - (\alpha + \beta' X_{1\cdot}) + h^*(\beta)$$