Using Multiple Outcomes to Improve the Synthetic Control Method

Eli Ben-Michael

CMU

Joint work with Liyang Sun (UCL) and Avi Feller (UC Berkeley)

Analyzing multiple outcomes with synthetic controls

Synthetic Control Method (SCM)

- Re-weight control units ("synthetic control") to closely match treated unit's pre-treatment outcomes

Analyzing multiple outcomes with synthetic controls

Synthetic Control Method (SCM)

- Re-weight control units ("synthetic control") to closely match treated unit's pre-treatment outcomes

Often interested in effects on multiple outcomes

- Common SCM practice $\longrightarrow$ run separate analyses
- Incompatible SCs and potential over-fitting

Analyzing multiple outcomes with synthetic controls

Synthetic Control Method (SCM)

- Re-weight control units ("synthetic control") to closely match treated unit's pre-treatment outcomes

Often interested in effects on multiple outcomes

- Common SCM practice $\longrightarrow$ run separate analyses
- Incompatible SCs and potential over-fitting

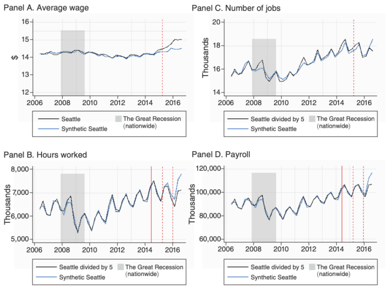


FIGURE 3. LEVELS OF EMPLOYMENT, WAGES, AND PAYROLL IN SEATTLE COMPARED TO SYNTHETIC SEATTLE IN JOBS PAYING LESS THAN $19 PER HOUR

[Jardim et al., 2022]

Analyzing multiple outcomes with synthetic controls

Synthetic Control Method (SCM)

- Re-weight control units ("synthetic control") to closely match treated unit's pre-treatment outcomes

Often interested in effects on multiple outcomes

- Common SCM practice ⟶ run separate analyses
- Incompatible SCs and potential over-fitting

We propose to find a single SC by:

- Fitting on all outcomes simultaneously
- Fitting on an index/avg of outcomes

Combines info across outcomes to reduce the bias

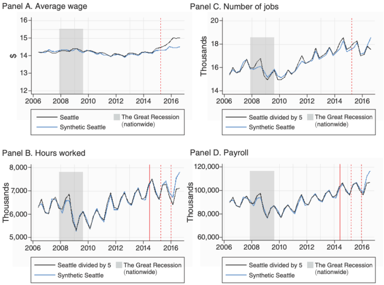


Figure 3. Levels of Employment, Wages, and Payroll in Seattle Compared to Synthetic Seattle in Jobs Paying Less than $19 per Hour

[Jardim et al., 2022]

Analyzing multiple outcomes with synthetic controls

Synthetic Control Method (SCM)

- Re-weight control units ("synthetic control") to closely match treated unit's pre-treatment outcomes

Often interested in effects on multiple outcomes

- Common SCM practice ⟶ run separate analyses
- Incompatible SCs and potential over-fitting

We propose to find a single SC by:

- Fitting on all outcomes simultaneously
- Fitting on an index/avg of outcomes

Combines info across outcomes to reduce the bias

Case study: Trejo et al. [2024] study on the 2014 Flint water crisis

- Math, reading, attendance, special needs

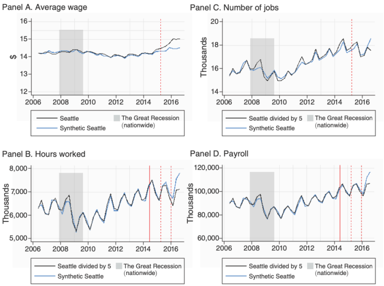


FIGURE 3. LEVELS OF EMPLOYMENT, WAGES, AND PAYROLL IN SEATTLE COMPARED TO SYNTHETIC SEATTLE IN JOBS PAYING LESS THAN $19 PER HOUR

[Jardim et al., 2022]

Notation and estimands

Units: $i = 1, \ldots, N$

Time: $t = 1, \ldots, T$

Outcomes: $k = 1, \ldots, K$

k^{th} outcome for unit i at time t: Y_{itk}

First unit is treated at time T_0

Potential outcomes $Y_{itk}(0), Y_{itk}(1)$

$$\text{treat} = \begin{pmatrix} & \checkmark & \checkmark & \checkmark \\ & & & \\ & & & \end{pmatrix}$$

Notation and estimands

Units: $i = 1, \ldots, N$

Time: $t = 1, \ldots, T$

Outcomes: $k = 1, \ldots, K$

k^{th} outcome for unit i at time t: Y_{itk}

First unit is treated at time T_0

Potential outcomes $Y_{itk}(0), Y_{itk}(1)$

Goal: Estimate effect on k^{th} outcome for treated unit at time $t \geq T_0$:

$$\tau_{tk} = Y_{1tk}(1) \quad - \quad Y_{1tk}(0)$$

$$\text{treat} = \begin{pmatrix} & \checkmark & \checkmark & \checkmark \\ & & & \\ & & & \end{pmatrix}$$

Impute the counterfactual via weighting

Synthetic control: weighted average of comparison units' outcomes

[Abadie et al., 2010, 2015]

$$\widehat{Y}_{1tk}(0) = \sum_{i \in \text{ctrls}} \hat{\gamma}_i Y_{itk}$$

Impute the counterfactual via weighting

Synthetic control: weighted average of comparison units' outcomes

[Abadie et al., 2010, 2015]

$$\widehat{Y}_{1tk}(0) = \sum_{i \in \text{ctrls}} \hat{\gamma}_i Y_{itk}$$

Weights optimize pre-treatment fit

$$\min_{\gamma \in \Delta} \sum_{t=1}^{T_0 - 1} \left(Y_{1tk} - \sum_{\text{controls}} \gamma_i Y_{itk} \right)^2$$

Impute the counterfactual via weighting

Synthetic control: weighted average of comparison units' outcomes

[Abadie et al., 2010, 2015]

$$\widehat{Y}_{1tk}(0) = \sum_{i \in \text{ctrls}} \hat{\gamma}_i Y_{itk}$$

Weights optimize pre-treatment fit

$$\min_{\gamma \in \Delta} \quad \|\text{imbalance}_k\|_2^2$$

Impute the counterfactual via weighting

Synthetic control: weighted average of comparison units' outcomes

[Abadie et al., 2010, 2015]

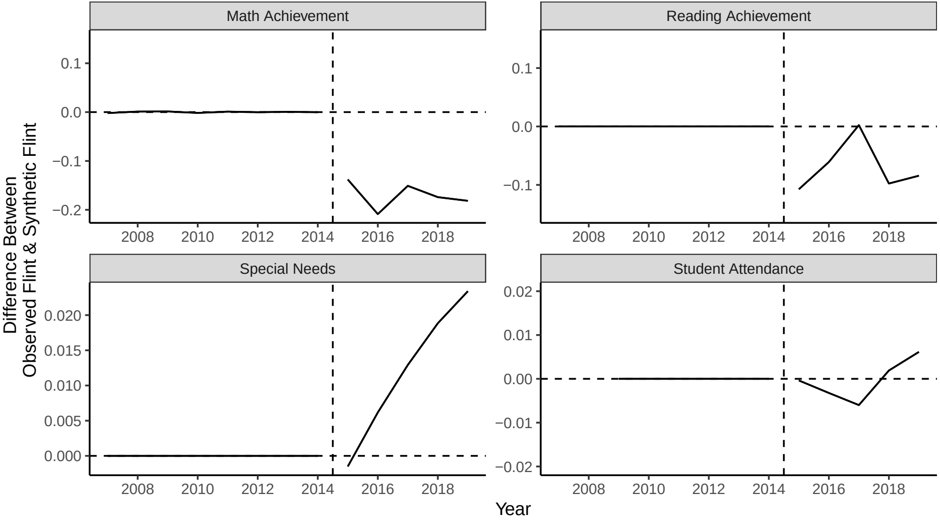
$$\widehat{Y}_{1tk}(0) = \sum_{i \in \text{ctrls}} \hat{\gamma}_i Y_{itk}$$

Weights optimize pre-treatment fit

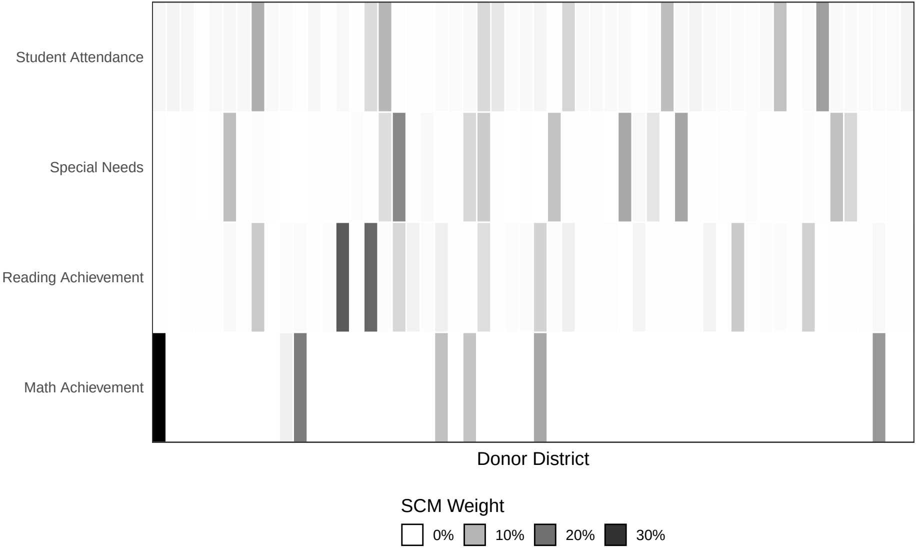
$$\min_{\gamma \in \Delta} \quad \|\text{imbalance}_k\|_2^2$$

Abadie et al. [2010]: low bias if **excellent pre-treatment fit** and a **long pre-period**

Perfect fit on all outcomes ⟶ over-fitting?



Very different weights across outcomes ⟶ inconsistent analyses?



To adjudicate this, let's take a deeper dive into the bias

Typically assume a linear factor model: $Y_{it}(0) = \sum_{r=1}^{R} \phi_{ir}\mu_{tr} + \varepsilon_{it}$

- μ_t are J latent factors vary over time, fixed over units
- ϕ_i are J latent factor loadings vary over units, fixed over time

can't observe these

To adjudicate this, let's take a deeper dive into the bias

Typically assume a linear factor model: $Y_{it}(0) = \sum_{r=1}^{R} \phi_{ir}\mu_{tr} + \varepsilon_{it}$

- μ_t are J latent factors vary over time, fixed over units
- ϕ_i are J latent factor loadings vary over units, fixed over time

can't observe these

Challenge: Want to balance **latent** loadings, can only balance **observed** outcomes

- Only a noisy proxy

To adjudicate this, let's take a deeper dive into the bias

Typically assume a linear factor model: $Y_{it}(0) = \sum_{r=1}^{R} \phi_{ir}\mu_{tr} + \varepsilon_{it}$

- μ_t are J latent factors vary over time, fixed over units
- ϕ_i are J latent factor loadings vary over units, fixed over time

can't observe these

Challenge: Want to balance **latent** loadings, can only balance **observed** outcomes

- Only a noisy proxy

Bias for a single outcome

$$\mathbb{E}[\hat{\tau} - \tau] \lesssim \frac{1}{\text{signal}} \times \text{pre-treatment fit} + \text{noise} \times \sqrt{\frac{\log N}{T}}$$

To adjudicate this, let's take a deeper dive into the bias

Typically assume a linear factor model: $Y_{it}(0) = \sum_{r=1}^{R} \phi_{ir}\mu_{tr} + \varepsilon_{it}$

- μ_t are J latent factors vary over time, fixed over units
- ϕ_i are J latent factor loadings vary over units, fixed over time

can't observe these

Challenge: Want to balance **latent** loadings, can only balance **observed** outcomes

- Only a noisy proxy

Bias for a single outcome

$$\mathbb{E}[\hat{\tau} - \tau] \lesssim \frac{1}{\text{signal}} \times \text{pre-treatment fit} + \text{noise} \times \sqrt{\frac{\log N}{T}}$$

Tradeoff between good pre-treatment fit and low approximation error

- Large # of time periods $\implies$ low approximation error
- Large # of time periods $\implies$ poor pre-treatment fit

To adjudicate this, let's take a deeper dive into the bias

Typically assume a linear factor model: $Y_{it}(0) = \sum_{r=1}^{R} \phi_{ir}\mu_{tr} + \varepsilon_{it}$

- μ_t are J latent factors vary over time, fixed over units
- ϕ_i are J latent factor loadings vary over units, fixed over time

can't observe these

Challenge: Want to balance **latent** loadings, can only balance **observed** outcomes

- Only a noisy proxy

Bias for a single outcome

$$\mathbb{E}[\hat{\tau} - \tau] \lesssim \frac{1}{\text{signal}} \times \text{pre-treatment fit} + \text{noise} \times \sqrt{\frac{\log N}{T}}$$

Tradeoff between good pre-treatment fit and low approximation error

- Large # of time periods $\Longrightarrow$ low approximation error
- Large # of time periods $\Longrightarrow$ poor pre-treatment fit

Back to Flint: low # of time periods, might be overfitting + large approx error

One set of weights for all outcomes

Tackle both problems by using a common set of weights for outcomes $k = 1, \ldots, K$

- Share information across outcomes $\longrightarrow$ more info on latent factor loadings

One set of weights for all outcomes

Tackle both problems by using a common set of weights for outcomes $k = 1, \ldots, K$

- Share information across outcomes $\longrightarrow$ more info on latent factor loadings

Option 1: **concatenate** the outcomes together

[contemporaneously proposed by Tian et al. [2023]]

$$\min_{\gamma \in \Delta} \frac{1}{K} \sum_{k=1}^{K} \|\text{imbalance}_k\|_2^2$$

One set of weights for all outcomes

Tackle both problems by using a common set of weights for outcomes $k = 1, \ldots, K$

- Share information across outcomes $\longrightarrow$ more info on latent factor loadings

Option 1: **concatenate** the outcomes together

[contemporaneously proposed by Tian et al. [2023]]

$$\min_{\gamma \in \Delta} \frac{1}{K} \sum_{k=1}^{K} \|\text{imbalance}_k\|_2^2$$

Option 2: **average** the outcomes together

$$\min_{\gamma \in \Delta} \left\| \frac{1}{K} \sum_{k=1}^{K} \text{imbalance}_k \right\|_2^2$$

One set of weights for all outcomes

Tackle both problems by using a common set of weights for outcomes $k = 1, \ldots, K$

- Share information across outcomes $\longrightarrow$ more info on latent factor loadings

Option 1: **concatenate** the outcomes together

[contemporaneously proposed by Tian et al. [2023]]

$$\min_{\gamma \in \Delta} \frac{1}{K} \sum_{k=1}^{K} \|\text{imbalance}_k\|_2^2$$

Option 2: **average** the outcomes together

$$\min_{\gamma \in \Delta} \left\| \frac{1}{K} \sum_{k=1}^{K} \text{imbalance}_k \right\|_2^2$$

Also include unit fixed effects (intercept-shifted SCM)

[Ferman and Pinto, 2021]

A shared latent structure across outcomes

Link outcomes together via a common set of latent factor loadings

[in the paper: generalize this in terms of rank conditions]

$$Y_{itk}(0) = \alpha_{ik} + \beta_{tk} + \sum_{r=1}^{R} \phi_{ir}\mu_{tkr} + \epsilon_{itk}$$

- Fixed effects α_{ik} & β_{tk} + factors μ_{kr} differ by outcomes
- Factor loadings ϕ_i common across outcomes
- Flexibility through number of factors R

A shared latent structure across outcomes

Link outcomes together via a common set of latent factor loadings

[in the paper: generalize this in terms of rank conditions]

$$Y_{itk}(0) = \alpha_{ik} + \beta_{tk} + \sum_{r=1}^{R} \phi_{ir}\mu_{tkr} + \epsilon_{itk}$$

- Fixed effects α_{ik} & β_{tk} + factors μ_{kr} differ by outcomes
- Factor loadings ϕ_i common across outcomes
- Flexibility through number of factors R

If R_0 common factors and ΔR idiosyncratic factors per outcome, sufficient condition:

$$R_0 + K \times \Delta R < N - 1$$

- Test scores [Duflo et al., 2011]
- Finer temporal resolution [Sun, EBM & Feller (2024)]

A shared latent structure across outcomes

Link outcomes together via a common set of latent factor loadings

[in the paper: generalize this in terms of rank conditions]

$$Y_{itk}(0) = \alpha_{ik} + \beta_{tk} + \sum_{r=1}^{R} \phi_{ir}\mu_{tkr} + \epsilon_{itk}$$

- Fixed effects α_{ik} & β_{tk} + factors μ_{kr} differ by outcomes
- Factor loadings ϕ_i common across outcomes
- Flexibility through number of factors R

If R_0 common factors and ΔR idiosyncratic factors per outcome, sufficient condition:

$$R_0 + K \times \Delta R < N - 1$$

- Test scores [Duflo et al., 2011]
- Finer temporal resolution [Sun, EBM & Feller (2024)]

Gives a common set of **oracle** weights that balance the common latent factor loadings

$$\sum_{\text{controls}} \phi_i \gamma_i^* = \phi_{\text{trt}}$$

+ add'l regularity condition that $\|\gamma^*\|_1$ is bounded

How well can each method expect to do?

How well can each method expect to do?

Quantify by comparing to **oracle** weights that balance latent factors

- SCM weights fit the outcome the best, so they must fit at least as well as the oracle weights

How well can each method expect to do?

Quantify by comparing to **oracle** weights that balance latent factors

- SCM weights fit the outcome the best, so they must fit at least as well as the oracle weights

For separate SCM weights:

$$\text{bias} \lesssim \frac{\text{noise}}{\text{signal}} + \text{noise} \times \sqrt{\frac{\log N}{T}}$$

add'l bias from fixed effects like Nickell [1981] bias

- Approx error $\downarrow$ as $T \uparrow$
- Pre-treatment fit stays the same

[see also Ferman and Pinto, 2021]

How well can each method expect to do?

Quantify by comparing to **oracle** weights that balance latent factors

- SCM weights fit the outcome the best, so they must fit at least as well as the oracle weights

For separate SCM weights:

$$\text{bias} \lesssim \frac{\text{noise}}{\text{signal}} + \text{noise} \times \sqrt{\frac{\log N}{T}}$$

add'l bias from fixed effects like Nickell [1981] bias

- Approx error ↓ as T ↑
- Pre-treatment fit stays the same

[see also Ferman and Pinto, 2021]

For concatenated SCM weights:

$$\text{bias} \lesssim \frac{\text{noise}}{\text{signal}} + \frac{\text{noise}}{\sqrt{K}} \times \sqrt{\frac{\log N}{T}}$$

- Reduces approx error by a factor of $\frac{1}{\sqrt{K}}$

How well can each method expect to do?

Quantify by comparing to **oracle** weights that balance latent factors

- SCM weights fit the outcome the best, so they must fit at least as well as the oracle weights

For separate SCM weights:

$$\text{bias} \lesssim \frac{\text{noise}}{\text{signal}} + \text{noise} \times \sqrt{\frac{\log N}{T}}$$

add'l bias from fixed effects like Nickell [1981] bias

- Approx error $\downarrow$ as $T \uparrow$
- Pre-treatment fit stays the same

[see also Ferman and Pinto, 2021]

For concatenated SCM weights:

$$\text{bias} \lesssim \frac{\text{noise}}{\text{signal}} + \frac{\text{noise}}{\sqrt{K}} \times \sqrt{\frac{\log N}{T}}$$

- Reduces approx error by a factor of $\frac{1}{\sqrt{K}}$

For averaged SCM weights:

$$\text{bias} \lesssim \frac{1}{\sqrt{K}} \times \left(\frac{\text{noise}}{\text{signal}} + \text{noise} \times \sqrt{\frac{\log N}{T}}\right)$$

- Reduces approx error by a factor of $\frac{1}{\sqrt{K}}$
- Improves pre-treatment fit by a factor of $\frac{1}{\sqrt{K}}$

A robust combined approach

Averaging might remove the signal $\rightarrow$ average SCM has large bias

A robust combined approach

Averaging might remove the signal $\rightarrow$ average SCM has large bias

Idea: achieve good fit on **both** concatenated and average objectives

- can achieve better of the two bounds

A robust combined approach

Averaging might remove the signal $\rightarrow$ average SCM has large bias

Idea: achieve good fit on **both** concatenated and average objectives

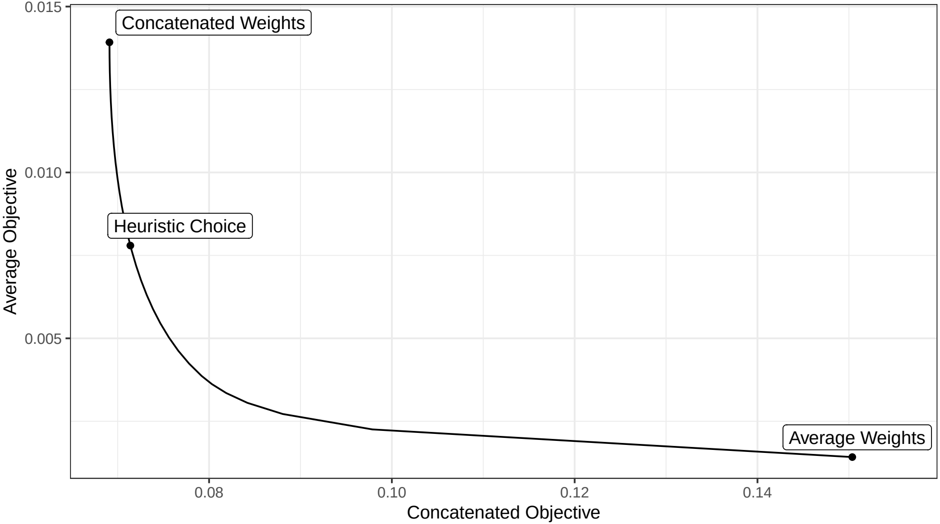
- can achieve better of the two bounds

Option 3: combined approach

$$\min_{\gamma \in \Delta} \nu \left\| \frac{1}{K} \sum_{k=1}^{K} \text{imbalance}_k \right\|_2^2 + \frac{1-\nu}{K} \sum_{k=1}^{K} \|\text{imbalance}_k\|_2^2$$

- In principle, a correct ν^* exists, but depends on the model
- Heuristic $\hat{\nu}$: ratio of avg and concatenated fit for concatenated SCM
- Vary ν as a sensitivity parameter

The balance frontier



Adapt the conformal inference approach from Chernozhukov et al. [2021]

Operates as a randomization test of a sharp null $H_0 : (\tau_1, \ldots, \tau_K) = (\tau_{10}, \ldots, \tau_{K0})$

Adapt the conformal inference approach from Chernozhukov et al. [2021]

Operates as a randomization test of a sharp null $H_0 : (\tau_1, \ldots, \tau_K) = (\tau_{10}, \ldots, \tau_{K0})$

1. Enforce the sharp null by adjusting post-treatment outcomes $Y_{itk} - \tau_{0k}$

Adapt the conformal inference approach from Chernozhukov et al. [2021]

Operates as a randomization test of a sharp null $H_0 : (\tau_1, \ldots, \tau_K) = (\tau_{10}, \ldots, \tau_{K0})$

1. Enforce the sharp null by adjusting post-treatment outcomes $Y_{itk} - \tau_{0k}$
2. Fit weights on all outcomes, incl. adjusted post-treatment outcomes

Inference on treatment effects

Adapt the conformal inference approach from Chernozhukov et al. [2021]

Operates as a randomization test of a sharp null $H_0 : (\tau_1, \dots, \tau_K) = (\tau_{10}, \dots, \tau_{K0})$

1. Enforce the sharp null by adjusting post-treatment outcomes $Y_{itk} - \tau_{0k}$
2. Fit weights on all outcomes, incl. adjusted post-treatment outcomes
3. Compute a test statistic on the residuals

Inference on treatment effects

Adapt the conformal inference approach from Chernozhukov et al. [2021]

Operates as a randomization test of a sharp null $H_0 : (\tau_1, \ldots, \tau_K) = (\tau_{10}, \ldots, \tau_{K0})$

1. Enforce the sharp null by adjusting post-treatment outcomes $Y_{itk} - \tau_{0k}$
2. Fit weights on all outcomes, incl. adjusted post-treatment outcomes
3. Compute a test statistic on the residuals
4. Randomly scramble pre-post treatment time indicator and compute p-value by comparing observed test stat to the distribution

Adapt the conformal inference approach from Chernozhukov et al. [2021]

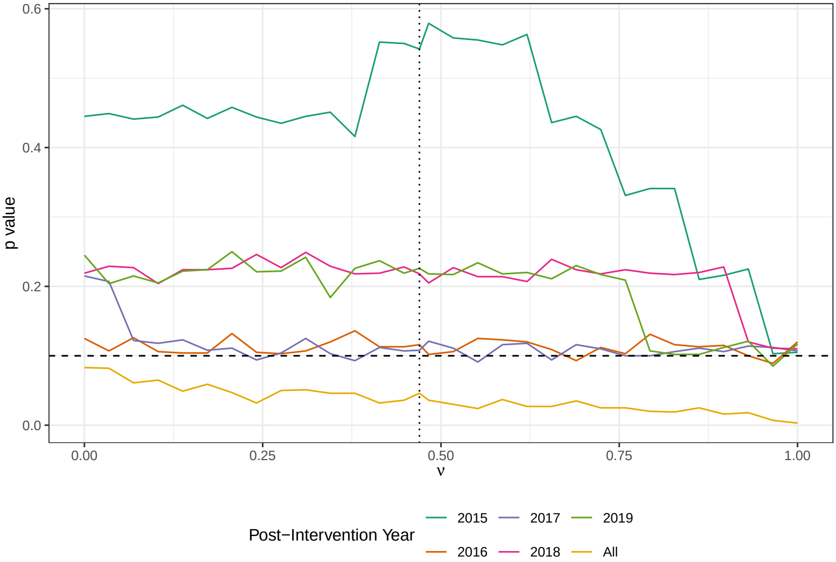
Operates as a randomization test of a sharp null $H_0 : (\tau_1, \dots, \tau_K) = (\tau_{10}, \dots, \tau_{K0})$

1. Enforce the sharp null by adjusting post-treatment outcomes $Y_{itk} - \tau_{0k}$
2. Fit weights on all outcomes, incl. adjusted post-treatment outcomes
3. Compute a test statistic on the residuals
4. Randomly scramble pre-post treatment time indicator and compute p-value by comparing observed test stat to the distribution

- Asymptotically correct size as $T \to \infty$
- But requires us to specify joint null on all outcomes together

Effects measured via different approaches

Sensitivity to ν



Recap

Synthetic controls with multiple outcomes

- Common practice: run a separate SCM analysis for each outcome
- Practical and theoretical pitfalls: potential for overfitting, inconsistent analyses

Recap

Synthetic controls with multiple outcomes

- Common practice: run a separate SCM analysis for each outcome
- Practical and theoretical pitfalls: potential for overfitting, inconsistent analyses

We propose to find a single set of weights

- Concatenate outcomes: ↓ overfitting
- Average outcomes: ↓ overfitting, ↑ pre-treatment fit
- Combined: more robust, advantages of both

Recap

Synthetic controls with multiple outcomes

- Common practice: run a separate SCM analysis for each outcome
- Practical and theoretical pitfalls: potential for overfitting, inconsistent analyses

We propose to find a single set of weights

- Concatenate outcomes: ↓ overfitting
- Average outcomes: ↓ overfitting, ↑ pre-treatment fit
- Combined: more robust, advantages of both

Many open questions and next steps

- Tests for shared factor structure & diagnostics for averaging?
- Weaken shared factor structure, e.g. hierarchical models?
- Less demanding form of inference?

Recap

Synthetic controls with multiple outcomes

- Common practice: run a separate SCM analysis for each outcome
- Practical and theoretical pitfalls: potential for overfitting, inconsistent analyses

We propose to find a single set of weights

- Concatenate outcomes: $\downarrow$ overfitting
- Average outcomes: $\downarrow$ overfitting, $\uparrow$ pre-treatment fit
- Combined: more robust, advantages of both

Many open questions and next steps

- Tests for shared factor structure & diagnostics for averaging?
- Weaken shared factor structure, e.g. hierarchical models?
- Less demanding form of inference?

Thank you!

ebenmichael.github.io

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.

Chernozhukov, V., Wuthrich, K., and Zhu, Y. (2021). An exact and robust conformal inference method for counterfactual and synthetic controls. *Journal of the American Statistical Association*, 116(536):1849–1864.

Duflo, E., Dupas, P., and Kremer, M. (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in kenya. *American Economic Review*, 101(5):1739–74.

Ferman, B. and Pinto, C. (2021). Synthetic controls with imperfect pre-treatment fit. *Quantitative Economics*.

Jardim, E., Long, M. C., Plotnick, R., van Inwegen, E., Vigdor, J., and Wething, H. (2022). Minimum wage increases, wages, and low-wage employment: Evidence from seattle. *American Economic Journal: Economic Policy*, 14(2):263–314.

References II

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica: Journal of the Econometric Society*, pages 1417–1426.

Tian, W., Lee, S., and Panchenko, V. (2023). Synthetic controls with multiple outcomes: Estimating the effects of non-pharmaceutical interventions in the covid-19 pandemic.

Trejo, S., Yeomans-Maldonado, G., and Jacob, B. (2024). The effects of the flint water crisis on the educational outcomes of school-age children. *Science Advances*, 10(11):eadk4737.