Uncertainty Quantification for Synthetic Controls

Matias D. Cattaneo, Princeton University

Yingjie Feng, Tsinghua University

Filippo Palomba, Princeton University

Rocío Titiunik, Princeton University

Jun 3, 2022

DataX Synthetic Control Methods Workshop

Introduction

This presentation: a mixture of two papers

- "Prediction Intervals for Synthetic Control Methods," with Matias Cattaneo and Rocío Titiunik. *Journal of the American Statistical Association*, 116(536): 1865-1880, Dec 2021.
- "Uncertainty Quantification in Synthetic Controls with Staggered Treatment Adoption," with Matias Cattaneo, Filippo Palomba, Rocío Titiunik.

I will mostly focus on the canonical case, but the idea can be generalized

Introduction

Synthetic control (SC) method

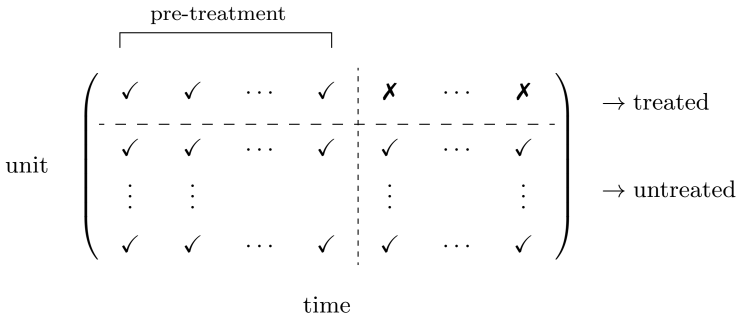
- Introduced by Abadie and Gardeazabal (2003)
- Causal effect of a treatment on a single or a few units
- Popular in comparative case studies
  - German reunification, California Tobacco Control, Compulsory Voting, etc
- Key idea
  - Predict $Y(0)$ of the treated by a linear combination of control units

We focus on uncertainty quantification of SC

- Prediction Intervals for treatment effect on the treated
- Two sources of uncertainty
- Non-asymptotic coverage guarantees
- Practical guidance
- Software package `scpi`

Data Structure: Canonical Case

Treatment: $D_{it} = 0, 1$ Outcome: $Y_{it} = Y_{it}(1)D_{it} + Y_{it}(0)(1 - D_{it})$

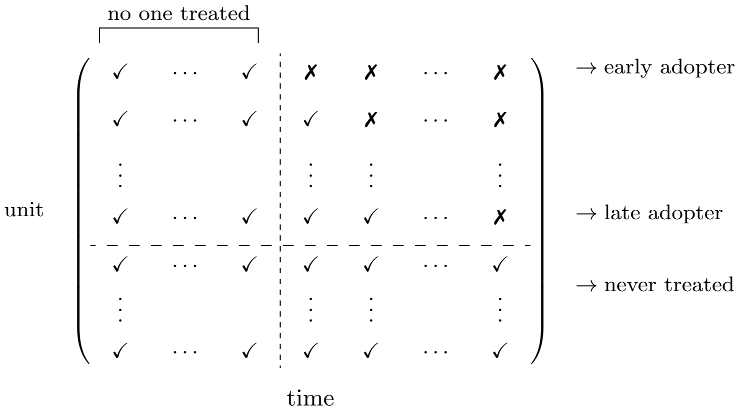


not treated : ✓; treated : ✗

Quantity of interest: (individual) treatment effect $Y_{1T}(1) - Y_{1T}(0)$ (random!)

Data Structure: Staggered Adoption

Adoption time: T_i Outcome: $Y_{it} = Y_{it}(T_i)\mathbb{1}(t \geq T_i) + Y_{it}(\infty)\mathbb{1}(t < T_i)$



not treated : ✓; treated : ✗

Quantity of interest: much more flexible

Data Structure: Staggered Adoption

Adoption time: T_i Outcome: $Y_{it} = Y_{it}(T_i)\mathbb{1}(t \geq T_i) + Y_{it}(\infty)\mathbb{1}(t < T_i)$

Many quantities of interest. For example,

- Individual treatment effect after k periods:

$$Y_{i(T_i+k)}(T_i) - Y_{i(T_i+k)}(\infty)$$

- Average treatment effect on units treated at time s after k periods:

$$\frac{1}{\#\{j : T_j = s\}} \sum_{i:T_i=s} \left(Y_{i(s+k)}(s) - Y_{i(s+k)}(\infty)\right)$$

  - Apply SC to each treated unit or an "aggregate" unit

- Average treatment effect on the treated (size N_1) after k periods:

$$\frac{1}{N_1} \sum_{i=1}^{N_1} \left(Y_{i(T_i+k)}(T_i) - Y_{i(T_i+k)}(\infty)\right)$$

Synthetic Control: Basics

In the canonical case,

$$Y_{it} = \begin{cases} Y_{it}(0) & \text{if } i = 2, \ldots N+1 \\ Y_{it}(0) & \text{if } i = 1 \text{ and } t \in \{1, 2, \ldots, T_0\} \\ Y_{it}(1) & \text{if } i = 1 \text{ and } t \in \{T_0 + 1, \ldots, T_0 + T_1\}. \end{cases}$$

Treatment effect on the treated (random!)

$$\tau_T = Y_{1T}(1) - Y_{1T}(0), \quad \text{for} \quad T > T_0$$

Find $\{w_i\}$

$$\sum_{i=2}^{N+1} w_i Y_{it}(0) \approx Y_{1t}(0), \quad \text{for} \quad t = 1, \cdots, T_0,$$

Hopefully,

$$\sum_{i=2}^{N+1} w_i Y_{iT}(0) \approx Y_{1T}(0), \quad \text{for} \quad T > T_0.$$

- Intuition: stable cross-sectional relation over time

Synthetic Control: Basics

In the canonical case,

$$Y_{it} = \begin{cases} Y_{it}(0) & \text{if } i = 2, \ldots N+1 \\ Y_{it}(0) & \text{if } i = 1 \text{ and } t \in \{1, 2, \ldots, T_0\} \\ Y_{it}(1) & \text{if } i = 1 \text{ and } t \in \{T_0+1, \ldots, T_0+T_1\}. \end{cases}$$

Treatment effect on the treated (random!)

$$\tau_T = Y_{1T}(1) - Y_{1T}(0), \quad \text{for} \quad T > T_0$$

Find $\{w_i\}$

$$\sum_{i=2}^{N+1} w_i Y_{it}(0) \approx Y_{1t}(0), \quad \text{for} \quad t = 1, \cdots, T_0,$$

Canonical SC

$$\widehat{\boldsymbol{w}} = \arg\min_{\boldsymbol{w} \in \mathcal{W}} \sum_{t=1}^{T_0} \left(Y_{1t}(0) - \boldsymbol{x}_t' \boldsymbol{w} \right)^2,$$

- $\mathcal{W} = \{\boldsymbol{w} \in \mathbb{R}^N : \sum_{i=2}^{N+1} w_i = 1, w_i \geq 0, \ \forall i \geq 2\}$ $\qquad \boldsymbol{x}_t = (Y_{2t}(0), \cdots, Y_{N+1,t}(0))'$

Prediction Intervals

Treatment effect on the treated

$$\tau_T = Y_{1T}(1) - Y_{1T}(0), \qquad \widehat{\tau}_T = Y_{1T}(1) - \boldsymbol{x}_T'\widehat{\boldsymbol{w}}, \qquad T > T_0$$

Prediction interval (PI) for *random* τ_T

$$\mathbb{P}\Big\{\mathbb{P}\Big[\tau_T \in \mathcal{I} \,\Big|\, \mathscr{H}\Big] \geq 1-\alpha\Big\} \geq 1-\pi$$

- $\mathcal{I}$: prediction interval
- $\mathscr{H}$: conditioning σ-algebra
  - unconditional if $\mathscr{H}$ is trivial
  - This paper: $\{\boldsymbol{x}_t : 1 \leq t \leq T\}$
- $(1-\alpha)$: conditional coverage prob., e.g., 95%
- π: failure over $\mathscr{H}$

Two Sources of Uncertainty

Treatment effect on the treated

$$\tau_T = Y_{1T}(1) - Y_{1T}(0), \qquad \widehat{\tau}_T = Y_{1T}(1) - \boldsymbol{x}_T' \widehat{\boldsymbol{w}}$$

For a pseudo-true value $\boldsymbol{w}_0$ ("estimand" of $\widehat{\boldsymbol{w}}$) ▸ Def.

$$Y_{1T}(0) = \boldsymbol{x}_T' \boldsymbol{w}_0 + u_T$$

A simple decomposition

$$\begin{aligned} \widehat{\tau}_T - \tau_T &= \quad Y_{1T}(0) \quad - \boldsymbol{x}_T' \widehat{\boldsymbol{w}} \\ &= (\boldsymbol{x}_T' \boldsymbol{w}_0 + u_T) - \boldsymbol{x}_T' \widehat{\boldsymbol{w}} \\ &= \quad u_T \quad - \boldsymbol{x}_T' (\widehat{\boldsymbol{w}} - \boldsymbol{w}_0) \end{aligned}$$

- In-sample error: $\boldsymbol{x}_T' (\widehat{\boldsymbol{w}} - \boldsymbol{w}_0)$
- Out-of-sample error: u_T
- Non-asymptotically, both are important

Prediction Intervals: Basic Construction

$$\widehat{\tau}_T - \tau_T = u_T - \boldsymbol{x}_T'(\widehat{\boldsymbol{w}} - \boldsymbol{w}_0)$$

- **In-Sample Error**: with prob. $\geq 1 - \pi_1$ (over $\mathscr{H}$)

$$\mathbb{P}\Big[M_{1,\mathtt{L}} \leq \boldsymbol{x}_T'(\boldsymbol{w}_0 - \widehat{\boldsymbol{w}}) \leq M_{1,\mathtt{U}} \;\Big|\; \mathscr{H}\Big] \geq 1 - \alpha_1$$

- **Out-of-Sample Error**: with prob. $\geq 1 - \pi_2$ (over $\mathscr{H}$)

$$\mathbb{P}\Big[M_{2,\mathtt{L}} \leq u_T \leq M_{2,\mathtt{U}} \;\Big|\; \mathscr{H}\Big] \geq 1 - \alpha_2$$

Prediction Interval for τ_T: with prob. $\geq 1 - \pi_1 - \pi_2$ (over $\mathscr{H}$)

$$\mathbb{P}\Big[M_{1,\mathtt{L}} + M_{2,\mathtt{L}} \leq \widehat{\tau}_T - \tau_T \leq M_{1,\mathtt{U}} + M_{2,\mathtt{U}} \Big| \mathscr{H}\Big] \geq 1 - \alpha_1 - \alpha_2$$

- Conservative, but offer non-asymptotic probability guarantee

In-Sample Error: Stationary Case

$$\sqrt{T_0}(\widehat{\boldsymbol{w}} - \boldsymbol{w}_0) = \underset{\boldsymbol{\delta} \in \sqrt{T_0}(\mathcal{W} - \boldsymbol{w}_0)}{\arg\min} \; \ell(\boldsymbol{\delta}), \quad \ell(\boldsymbol{\delta}) = \boldsymbol{\delta}' \underbrace{\left(\frac{1}{T_0} \sum_{t=1}^{T_0} \boldsymbol{x}_t \boldsymbol{x}_t' \right)}_{\widehat{\boldsymbol{Q}}} \boldsymbol{\delta} - 2 \underbrace{\left(\frac{1}{\sqrt{T_0}} \sum_{t=1}^{T_0} \boldsymbol{x}_t' u_t \right)}_{\widehat{\boldsymbol{\gamma}}} \boldsymbol{\delta}$$

- Assume $\mathcal{W}$ is convex
- Possibly **misspecified**: $\boldsymbol{\gamma} := \mathbb{E}[\widehat{\boldsymbol{\gamma}} | \mathscr{H}] \neq 0$
- By optimality of $\widehat{\boldsymbol{w}}$ and $\boldsymbol{w}_0$,

$$\sqrt{T_0}(\widehat{\boldsymbol{w}} - \boldsymbol{w}_0) \in \left\{ \boldsymbol{\delta} \in \Delta : \boldsymbol{\delta}' \widehat{\boldsymbol{Q}} \boldsymbol{\delta} - 2(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})' \boldsymbol{\delta} \leq 0 \right\}, \quad \Delta = \sqrt{T_0}(\mathcal{W} - \boldsymbol{w}_0)$$

  - $\widehat{\boldsymbol{Q}}$ is fixed conditional on $\mathscr{H}$
  - Approximate $\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}$

In-Sample Error: Distributional Approximation

Goal: bound $\boldsymbol{x}_T'(\widehat{\boldsymbol{w}} - \boldsymbol{w}_0)$, e.g.

$$\sqrt{T_0}\boldsymbol{x}_T'(\widehat{\boldsymbol{w}} - \boldsymbol{w}_0) \leq \sup_{\boldsymbol{\delta} \in \mathcal{M}_{\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}}} \boldsymbol{x}_T'\boldsymbol{\delta}$$

$$\mathcal{M}_{\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}} = \left\{\boldsymbol{\delta} \in \Delta : \boldsymbol{\delta}'\widehat{\boldsymbol{Q}}\boldsymbol{\delta} - 2(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})'\boldsymbol{\delta} \leq 0\right\}$$

Reduce to distributional approximation of $\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}$: for any κ,

$$\sup_{\boldsymbol{\delta} \in \mathcal{M}_{\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}}} \boldsymbol{x}_T'\boldsymbol{\delta} \leq \kappa \quad \Leftrightarrow \quad \widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma} \in \mathcal{A}_\kappa$$

- Normal approximation: ▸ detail

$$\begin{aligned} \mathbb{P}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma} \in \mathcal{A}_\kappa) &\approx \mathbb{P}(\boldsymbol{G} \in \mathcal{A}_\kappa), \quad \boldsymbol{G} \sim \mathsf{N}(0, \mathbb{V}[\widehat{\boldsymbol{\gamma}}|\mathscr{H}]) \\ &\approx \mathbb{P}(\widehat{\boldsymbol{G}} \in \mathcal{A}_\kappa), \quad \widehat{\boldsymbol{G}} \sim \mathsf{N}(0, \widehat{\mathbb{V}}[\widehat{\boldsymbol{\gamma}}|\mathscr{H}]) \end{aligned}$$

- Bounds can be obtained by simulation
  - A constraint set $\Delta^\star$ in simulation: locally equivalent to $\Delta = \sqrt{T_0}(\mathcal{W} - \boldsymbol{w}_0)$

Out-of-Sample Error

Three approaches:

- Concentration inequalities, e.g. subgaussian

$$\mathbb{P}\Big(|u_T - \mathbb{E}[u_T|\mathscr{H}]| \geq \varpi_u \Big| \mathscr{H}\Big) \leq 2\exp\left(-\frac{\varpi_u^2}{2\sigma_{\mathscr{H}}^2}\right)$$

- Location-scale model

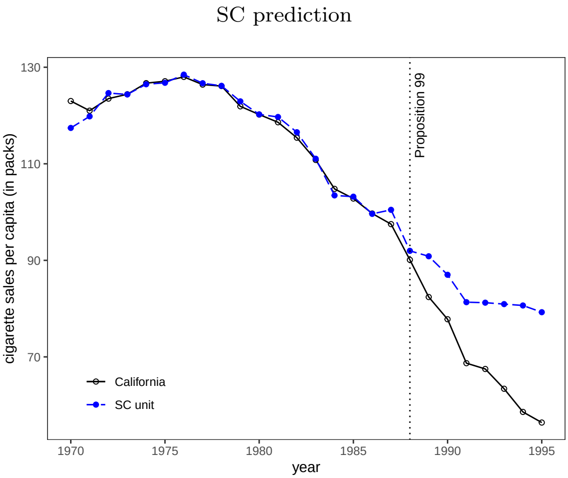
$$u_t = \mathbb{E}[u_t|\mathscr{H}] + (\mathbb{V}[u_t|\mathscr{H}])^{1/2} e_t, \quad \{e_t\} \perp\!\!\!\perp \mathscr{H}$$

- Quantile regression: model conditional quantiles of u_t given $\mathscr{H}$

Sensitivity analysis

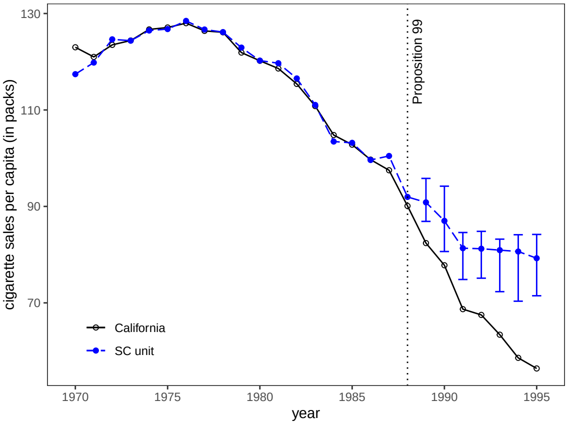
- To cancel out the effect, how large u_T needs to be?

Example: Proposition 99 (California)



Example: Proposition 99 (California)

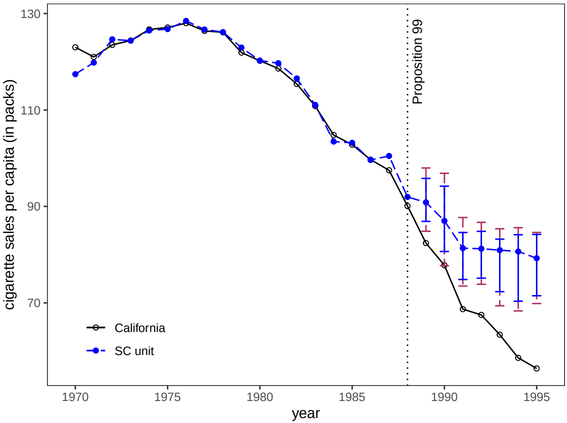
SC prediction with PI, in-sample error



Note: PI has $\geq 95\%$ coverage.

Example: Proposition 99 (California)

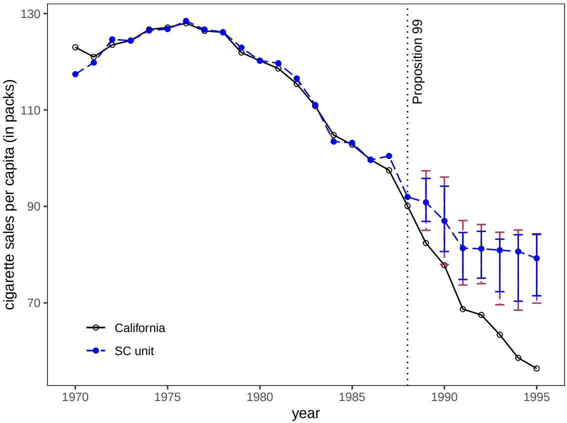
SC prediction with PI for $Y_{1T}(0)$, concentration-based



Note: PI has $\geq 90\%$ coverage.

Example: Proposition 99 (California)

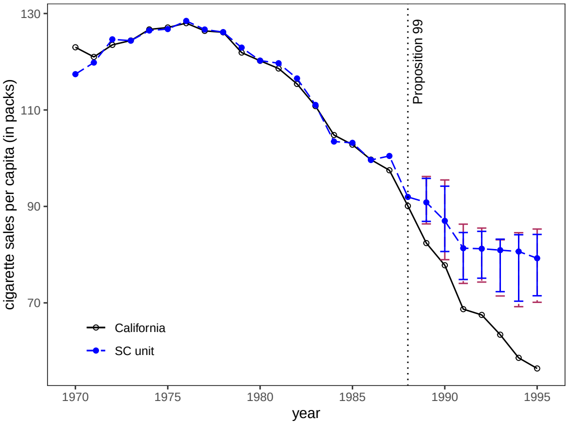
SC prediction with PI for $Y_{1T}(0)$, local-scale model



Note: PI has $\geq 90\%$ coverage.

Example: Proposition 99 (California)

SC prediction with PI for $Y_{1T}(0)$, quantile-based



Note: PI has $\geq 90\%$ coverage.

Extension: Multiple Treated Units, Staggered Adoption

- $\mathfrak{J}$ treated units
- Staggered adoption: $T_1 \leq \cdots \leq T_{N_1}$, $T_0 := T_1 - 1$
- M features (not just pre-treatment outcomes)
- General convex constraint set $\mathcal{W}$ (possibly nonlinear)
- Weighting matrix $\boldsymbol{V}$

$$\underbrace{\boldsymbol{A}_j}_{\substack{\text{treated} \\ T_0 \cdot M \times 1}} \approx \underbrace{\boldsymbol{B}_j}_{\substack{\text{untreated} \\ T_0 \cdot M \times N}} \boldsymbol{w}_j, \qquad \text{for} \quad j = 1, \cdots, \mathfrak{J},$$

$$\underbrace{\boldsymbol{A}}_{T_0 \cdot M \cdot \mathfrak{J} \times 1} = \begin{bmatrix} \boldsymbol{A}_1 \\ \vdots \\ \boldsymbol{A}_{\mathfrak{J}} \end{bmatrix}, \qquad \underbrace{\boldsymbol{B}}_{T_0 \cdot M \cdot \mathfrak{J} \times N \cdot \mathfrak{J}} = \begin{bmatrix} \boldsymbol{B}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{B}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{B}_{\mathfrak{J}} \end{bmatrix}$$

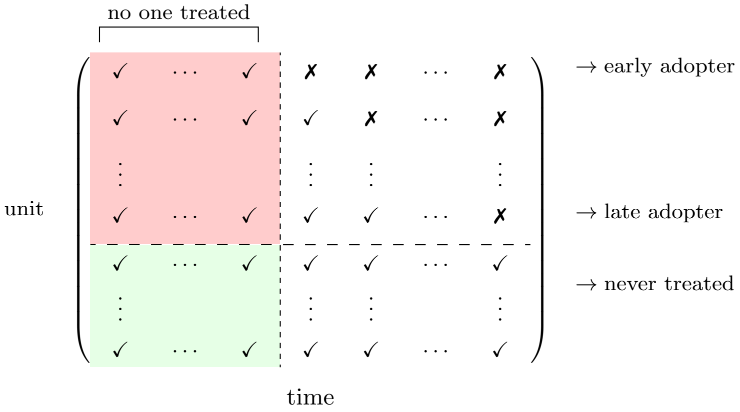
$$\widehat{\boldsymbol{w}} = (\widehat{\boldsymbol{w}}_1', \cdots, \widehat{\boldsymbol{w}}_{\mathfrak{J}}')' = \underset{\boldsymbol{w} \in \mathcal{W}}{\arg\min}\, (\boldsymbol{A} - \boldsymbol{B}\boldsymbol{w})' \boldsymbol{V} (\boldsymbol{A} - \boldsymbol{B}\boldsymbol{w})$$

Similar optimization bounds can be constructed

Extension: Multiple Treated Units, Staggered Adoption

For example,

- Fix the sample used to construct weights
- $\boldsymbol{A}$ matrix: use the "red" portion
- $\boldsymbol{B}$ matrix: use the "green" portion



Extension: Simultaneous Prediction Intervals

- The previous PIs have coverage for *each* post-treatment period
- Can be generalized to have **simultaneous** coverage for multiple periods
- For example, in the canonical case, treatment effect on "unit 1" at time t is

$$\tau_t = Y_{1t}(1) - Y_{1t}(0)$$
$$\hat{\tau}_t - \tau_t = u_t - \boldsymbol{x}_t'(\widehat{\boldsymbol{w}} - \boldsymbol{w}_0)$$

Simultaneous PI:

$$\mathbb{P}\Big\{\mathbb{P}\big[\tau_t \in \mathcal{I}, \text{ for all } T_0 + 1 \leq t \leq T_0 + L \,\big|\, \mathscr{H}\big] \geq 1 - \alpha\Big\} \geq 1 - \pi.$$

- In-sample error: easy to accommodate in the optimization
  - Take sup/inf over more evaluation vectors $\{\boldsymbol{x}_t\}$
- Out-of-sample error:
  - Maximal inequality for $\{u_t\}$

Conclusion

- Prediction Intervals for general SC methods
- Conditional validity
- Non-asymptotic probability guarantee
- Two sources of uncertainty
- Staggered adoption allowed

Thanks!

In-Sample Error: Plug-in Approximation

$$\mathbb{P}(\sqrt{T_0}\boldsymbol{x}_T'(\widehat{\boldsymbol{w}}-\boldsymbol{w}_0)\leq\kappa)\geq\mathbb{P}(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\in\mathcal{A}_\kappa)\approx\mathbb{P}(\boldsymbol{G}\in\mathcal{A}_\kappa),\quad \boldsymbol{G}\sim\mathsf{N}(0,\mathbb{V}[\widehat{\boldsymbol{\gamma}}|\mathscr{H}])$$
$$\approx\mathbb{P}(\widehat{\boldsymbol{G}}\in\mathcal{A}_\kappa),\quad \widehat{\boldsymbol{G}}\sim\mathsf{N}(0,\widehat{\mathbb{V}}[\widehat{\boldsymbol{\gamma}}|\mathscr{H}])$$

Basic approximation strategy:

$$\sup\left\{\boldsymbol{x}_T'\boldsymbol{\delta}:\boldsymbol{\delta}\in\Delta,\ \ \boldsymbol{\delta}'\widehat{\boldsymbol{Q}}\boldsymbol{\delta}-2(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})'\boldsymbol{\delta}\leq 0\right\}$$
$$\Downarrow$$
$$\sup\left\{\boldsymbol{x}_T'\boldsymbol{\delta}:\boldsymbol{\delta}\in\Delta,\ \ \boldsymbol{\delta}'\widehat{\boldsymbol{Q}}\boldsymbol{\delta}-2\boldsymbol{G}'\boldsymbol{\delta}\leq 0\right\}$$
$$\Downarrow$$
$$\sup\left\{\boldsymbol{x}_T'\boldsymbol{\delta}:\boldsymbol{\delta}\in\Delta^\star,\ \boldsymbol{\delta}'\widehat{\boldsymbol{Q}}\boldsymbol{\delta}-2\widehat{\boldsymbol{G}}'\boldsymbol{\delta}\leq 0\right\}$$

- $\Delta^\star$: locally equivalent to $\Delta=\sqrt{T_0}(\mathcal{W}-\boldsymbol{w}_0)$
- Thresholding

What is $\boldsymbol{w}_0$?

Definition of $\boldsymbol{w}_0$ relies on

- Constraint $\mathcal{W}$
- Assumptions on $\{Y_{1t}(0), \boldsymbol{x}_t\}$
- Conditioning set $\mathscr{H}$ ($\boldsymbol{w}_0$ may be random)

Examples

- $(Y_{1t}(0), \boldsymbol{x}_t)$ stationary

$$\boldsymbol{w}_0 = \arg\min_{\boldsymbol{w}\in\mathcal{W}} \mathbb{E}\left[\frac{1}{T_0}\sum_{t=1}^{T_0}\left(Y_{1t}(0) - \boldsymbol{x}_t'\boldsymbol{w}\right)^2 \Big| \mathscr{H}\right]$$

  - (Constrained) best linear prediction

- $Y_{1t}(0), \boldsymbol{x}_t \sim I(1)$ (integrated process)
  - Cointegration: $Y_{1t}(0) - \boldsymbol{x}_t'\boldsymbol{w}_0 \sim I(0)$
  - $(1, -\boldsymbol{w}_0)$: cointegrating vector (unique given $\boldsymbol{w}_0 \in \mathcal{W}$)

back

In-Sample Error: Non-Stationary Case

$$T_0(\widehat{\boldsymbol{w}} - \boldsymbol{w}_0) = \underset{\boldsymbol{\delta} \in T_0(\mathcal{W} - \boldsymbol{w}_0)}{\arg\min} \ \ell(\boldsymbol{\delta}), \quad \ell(\boldsymbol{\delta}) = \boldsymbol{\delta}' \underbrace{\Big(\frac{1}{T_0^2} \sum_{t=1}^{T_0} \boldsymbol{x}_t \boldsymbol{x}_t' \Big)}_{\widehat{\boldsymbol{Q}}} \boldsymbol{\delta} - 2 \underbrace{\Big(\frac{1}{T_0} \sum_{t=1}^{T_0} \boldsymbol{x}_t' u_t \Big)}_{\widehat{\boldsymbol{\gamma}}} \boldsymbol{\delta}$$

Cointegrated system, e.g.,

$$Y_{1t}(0) = \boldsymbol{x}_t' \boldsymbol{w}_0 + u_t, \quad \boldsymbol{x}_t = \boldsymbol{x}_{t-1} + \boldsymbol{\epsilon}_t, \quad (u_t, \boldsymbol{\epsilon}_t) \sim \text{i.i.d } I(0)$$

Again, bound $T_0(\widehat{\boldsymbol{w}} - \boldsymbol{w}_0)$ by simulating

$$\sup \Big\{ \boldsymbol{x}_T' \boldsymbol{\delta} : \boldsymbol{\delta} \in \Delta^\star, \ \boldsymbol{\delta}' \widehat{\boldsymbol{Q}} \boldsymbol{\delta} - 2 \widehat{\boldsymbol{G}}' \boldsymbol{\delta} \leq 0 \Big\}, \quad \widehat{\boldsymbol{G}} \sim \mathsf{N}(0, \widehat{\mathbb{V}}[\widehat{\boldsymbol{\gamma}} | \mathscr{H}])$$

- $\widehat{\boldsymbol{Q}}$ fixed conditional on $\mathscr{H}$
- Probably $\mathbb{E}[u_t | \mathscr{H}] \neq 0$
- Non-stationarity affects the analysis of $\widehat{\boldsymbol{Q}}$ and $\mathbb{V}[\widehat{\boldsymbol{\gamma}} | \mathscr{H}]$