Inference on Union Bounds

Xinyue Bei
<xinyue.bei@austin.utexas.edu>
UT Austin

Cornell Econometrics Workshop
September 19, 2024

Motivation

In many empirical applications, the target object is in a **union bound**

$$\theta \in \left[\min_{b \in \mathcal{B}} \lambda_{\ell,b},\ \max_{b \in \mathcal{B}} \lambda_{u,b}\right]$$

- θ is the target object
- $(\lambda_\ell, \lambda_u) \in \mathbb{R}^{2|\mathcal{B}|}$ is unknown but estimable
- $\mathcal{B}$ is the set of indices: known and finite

The goal of this paper: Construct a confidence interval for θ

Examples

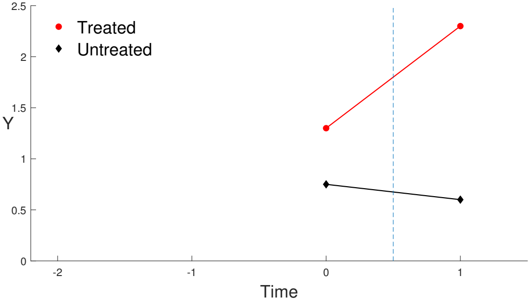
DiD: Manski and Pepper (2018, REStat), Rambachan and Roth (2023, REStud)

What if the parallel trends assumption does not hold?

Examples

DiD: Manski and Pepper (2018, REStat), Rambachan and Roth (2023, REStud)

What if the parallel trends assumption does not hold?

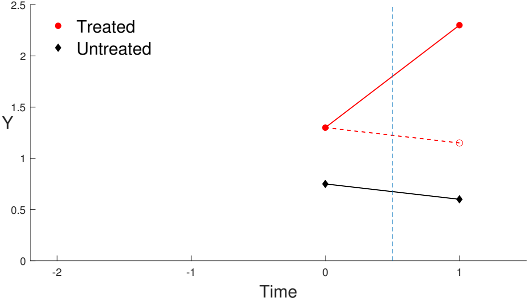


Units with $D = 1$ receive a treatment at $t = 1$

Examples

DiD: Manski and Pepper (2018, REStat), Rambachan and Roth (2023, REStud)

What if the parallel trends assumption does not hold?



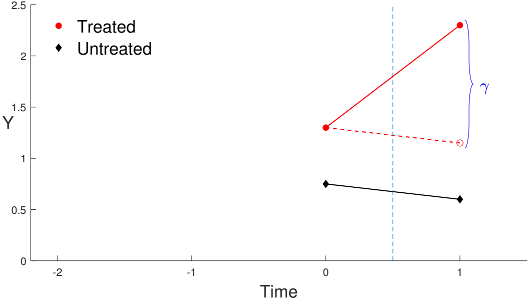
$Y_t(d)$: the potential outcome at t with treatment d
With parallel trends

$$0 = \mathbb{E}\left[Y_t(0) - Y_{t-1}(0) \mid D = 1\right] - \mathbb{E}\left[Y_t(0) - Y_{t-1}(0) \mid D = 0\right]$$

Examples

DiD: Manski and Pepper (2018, REStat), Rambachan and Roth (2023, REStud)

What if the parallel trends assumption does not hold?



$Y_t(d)$: the potential outcome at t with treatment d
With parallel trends

$$0 = \mathbb{E}\left[Y_t(0) - Y_{t-1}(0) \mid D = 1\right] - \mathbb{E}\left[Y_t(0) - Y_{t-1}(0) \mid D = 0\right]$$

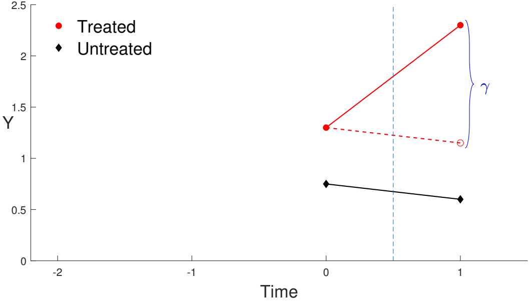
The DiD estimand γ is

$$\gamma = \underbrace{ATT}_{\theta}$$

Examples

DiD: Manski and Pepper (2018, REStat), Rambachan and Roth (2023, REStud)

What if the parallel trends assumption does not hold?



$Y_t(d)$: the potential outcome at t with treatment d

Without parallel trends

$$\Delta_t = \mathbb{E}\left[Y_t(0) - Y_{t-1}(0) \mid D = 1\right] - \mathbb{E}\left[Y_t(0) - Y_{t-1}(0) \mid D = 0\right]$$

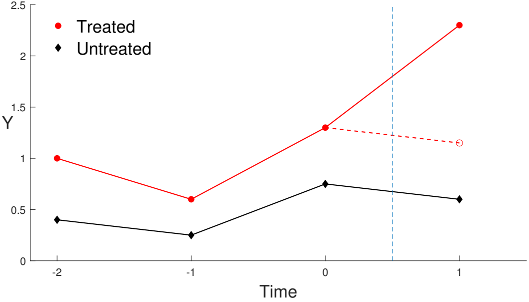
The DiD estimand γ is

$$\gamma = \underbrace{ATT}_{\theta} + \Delta_1$$

Examples

DiD: Manski and Pepper (2018, REStat), Rambachan and Roth (2023, REStud)

What if the parallel trends assumption does not hold?



$Y_t(d)$: the potential outcome at t with treatment d
Without parallel trends

$$\Delta_t = \mathbb{E}\left[Y_t(0) - Y_{t-1}(0) \mid D = 1\right] - \mathbb{E}\left[Y_t(0) - Y_{t-1}(0) \mid D = 0\right]$$

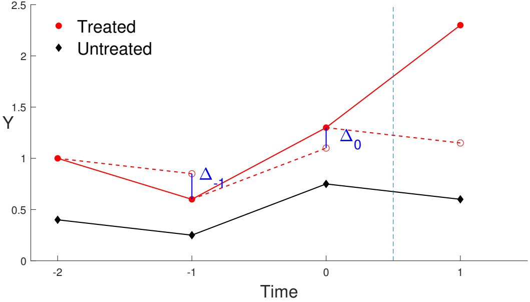
The DiD estimand γ is

$$\gamma = \underbrace{ATT}_{\theta} + \Delta_1$$

Examples

DiD: Manski and Pepper (2018, REStat), Rambachan and Roth (2023, REStud)

What if the parallel trends assumption does not hold?



$Y_t(d)$: the potential outcome at t with treatment d
Without parallel trends

$$\Delta_t = \mathbb{E}\left[Y_t(0) - Y_{t-1}(0) \mid D = 1\right] - \mathbb{E}\left[Y_t(0) - Y_{t-1}(0) \mid D = 0\right]$$

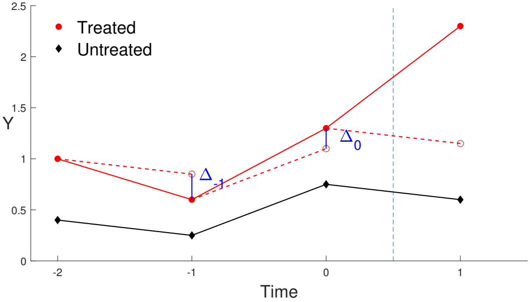
The DiD estimand γ is

$$\gamma = \underbrace{ATT}_{\theta} + \Delta_1$$

Examples

DiD: Manski and Pepper (2018, REStat), Rambachan and Roth (2023, REStud)

What if the parallel trends assumption does not hold?



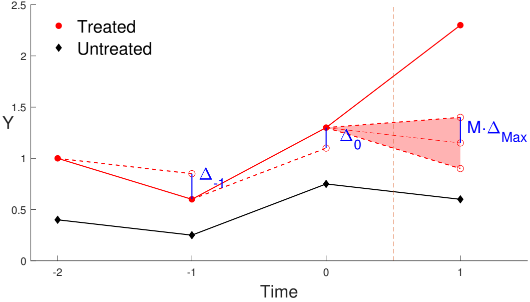
Relax the parallel trends assumption by

$$|\Delta_1| \leqslant M \cdot \max_{t=-T+1,\dots,0} |\Delta_t|$$

Examples

DiD: Manski and Pepper (2018, REStat), Rambachan and Roth (2023, REStud)

What if the parallel trends assumption does not hold?



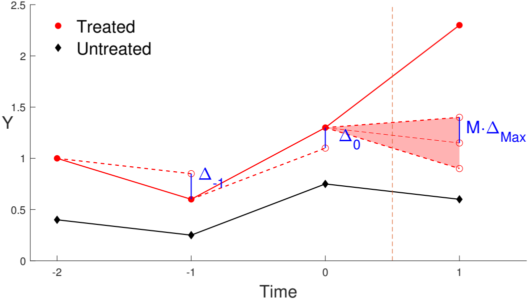
Relax the parallel trends assumption by

$$|\Delta_1| \leqslant M \cdot \max_{t=-T+1,\ldots,0} |\Delta_t|$$

Examples

DiD: Manski and Pepper (2018, REStat), Rambachan and Roth (2023, REStud)

What if the parallel trends assumption does not hold?



$$\theta = ATT \in \left[\min_{b \in \mathcal{B}} \lambda_{\ell,b},\ \max_{b \in \mathcal{B}} \lambda_{u,b}\right]$$

where $\mathcal{B} = \{-(T-1), ..., T-1, T\}$,

$$\lambda_{\ell,b} = \lambda_{u,b} = \begin{cases} \gamma - M\Delta_b & \text{if } b = -(T-1), ..., 0, \\ \gamma + M\Delta_{b-T} & \text{if } b = 1, ..., T \end{cases}$$

Examples

DiD: Manski and Pepper (2018, REStat), Rambachan and Roth (2023, REStud)

Examples

DiD: Manski and Pepper (2018, REStat), Rambachan and Roth (2023, REStud)

Misspecification Analysis: Stoye (2022, WP)

The identified set for θ is $[\theta_L, \theta_u]$, and this set is empty if $\theta_L > \theta_U$

Stoye (2022, WP) suggests reporting the misspecification robust identified set

$$[\theta_L, \theta_U] \cup \left\{ \frac{\sigma_U \theta_L + \sigma_L \theta_U}{\sigma_L + \sigma_U} \right\}$$

where σ_L and σ_U are the standard deviations for $\hat{\theta}_L$ and $\hat{\theta}_U$

$$\theta \in \left[\min \left\{ \theta_L, \frac{\sigma_U \theta_L + \sigma_L \theta_U}{\sigma_L + \sigma_U} \right\}, \ \max \left\{ \theta_U, \frac{\sigma_U \theta_L + \sigma_L \theta_U}{\sigma_L + \sigma_U} \right\} \right]$$

Examples

DiD: Manski and Pepper (2018, REStat), Rambachan and Roth (2023, REStud)

Misspecification Analysis: Stoye (2022, WP)

The identified set for θ is $[\theta_L, \theta_u]$, and this set is empty if $\theta_L > \theta_U$

Stoye (2022, WP) suggests reporting the misspecification robust identified set

$$[\theta_L, \theta_U] \cup \left\{ \frac{\sigma_U \theta_L + \sigma_L \theta_U}{\sigma_L + \sigma_U} \right\}$$

where σ_L and σ_U are the standard deviations for $\hat{\theta}_L$ and $\hat{\theta}_U$

$$\theta \in \left[\min \left\{ \theta_L, \frac{\sigma_U \theta_L + \sigma_L \theta_U}{\sigma_L + \sigma_U} \right\},\ \max \left\{ \theta_U, \frac{\sigma_U \theta_L + \sigma_L \theta_U}{\sigma_L + \sigma_U} \right\} \right]$$

Sign congruence: Molinari, Miller, Stoye (2024, WP)

Whether an average treatment effect has the same sign accross two groups

$$H_0 : \theta_1 \theta_2 \geq 0 \Leftrightarrow H_0 : 0 \in [\min\{\theta_1, -\theta_1\}, \max\{\theta_2, -\theta_2\}]$$

Examples

- Difference-in-Differences: Hasegawa, Small, Webster (2019, Epidemiology), Ye, Keele, Hasegawa, and Small (2023, JASA), Ban and Kédagni (2023, WP)
- Regression Discontinuity Design: Kolesár and Rothe (2018, AER)
- Bunching and Income Elasticity: Blomquist, Newey, Kumar, and Liang (2021, JPE)
- Sign congruence: Brinch, Mogstad, Wiswall (2017, JPE), Kowalski (2022, RES), Kim (2024, WP)
- Misspecification Analysis: Masten and Poirier (2021, ECTA), Apfel and Windmeijer (2022, WP)
- Instrumental Variables: Machado, Shaikh, and Vytlacil (2019, JoE)

Main Contributions

I propose a novel CI based on modified conditional inference

- **Valid**: CI covers θ with prob. $\geq 1 - \alpha$ under mild regularity conditions
- **Short**: higher power than existing methods under a large set of DGPs

Main Contributions

I propose a novel CI based on modified conditional inference

- **Valid**: CI covers θ with prob. $\geq 1 - \alpha$ under mild regularity conditions
- **Short**: higher power than existing methods under a large set of DGPs

Main Contributions

I propose a novel CI based on modified conditional inference

- **Valid**: CI covers θ with prob. $\geq 1 - \alpha$ under mild regularity conditions
- **Short**: higher power than existing methods under a large set of DGPs

Compare with Existing Procedures

My method improves upon existing valid procedures

	My	Simple	Hybrid	Adj. Boot.
Adjust for union	✓	×	×	✓
Apply to general cases	✓	✓	×	✓
$\sqrt{n}$ conv. rate to id set	✓	✓	✓	×

- Simple CI: Kolesár and Rothe (2018, AER), among others
- Hybrid CI: Rambachan and Roth (2023, RES)
- Adjusted Bootstrap: Ye, Keele, Hasegawa and Small (2023, JASA)

Contributions to Other Related Literature

Intersection Bounds & Moment Inequalities

Chernozhukov, Hong, and Tamer (2007), Romano and Shaikh (2008), Rosen (2008), D. Andrews and Guggenberger (2009), D. Andrews and Soares (2010), Chernozhukov, Lee, and Rosen (2013), D. Andrews and Shi (2013), Bugni, Canay and Shi (2015), among others

- This paper complements the intersection bounds Intersection

Contributions to Other Related Literature

Intersection Bounds & Moment Inequalities
Chernozhukov, Hong, and Tamer (2007), Romano and Shaikh (2008), Rosen (2008), D. Andrews and Guggenberger (2009), D. Andrews and Soares (2010), Chernozhukov, Lee, and Rosen (2013), D. Andrews and Shi (2013), Bugni, Canay and Shi (2015), among others

- This paper complements the intersection bounds Intersection

Directionally Differentiable Functions
Hirano and Porter (2012), Fang and Santos (2019), Fang (2018), Ponomarev (2022), among others

- This paper sheds light on inference of direct. diff. func. w/o convex null

Contributions to Other Related Literature

Intersection Bounds & Moment Inequalities

Chernozhukov, Hong, and Tamer (2007), Romano and Shaikh (2008), Rosen (2008), D. Andrews and Guggenberger (2009), D. Andrews and Soares (2010), Chernozhukov, Lee, and Rosen (2013), D. Andrews and Shi (2013), Bugni, Canay and Shi (2015), among others

- This paper complements the intersection bounds Intersection

Directionally Differentiable Functions

Hirano and Porter (2012), Fang and Santos (2019), Fang (2018), Ponomarev (2022), among others

- This paper sheds light on inference of direct. diff. func. w/o convex null

Conditional Inference

I. Andrews and Mikusheva (2016), I. Andrews, Roth and Pakes (2016), I. Andrews, Kitagawa, McCloskey (2021, 2023), Rambachan and Roth (2023), among others

- This paper widens the use of the conditional inference technique

Outline

Outline

1. Inference Procedure
2. Simulation
3. Empirical Illustration
4. Conclusion

Setting

The target object

$$\theta \in \left[\min_{b \in \mathcal{B}} \lambda_{\ell,b},\ \max_{b \in \mathcal{B}} \lambda_{u,b}\right]$$

where λ_ℓ and λ_u are $|\mathcal{B}|$-dimensional vectors

Assume that $\widehat{\lambda}_\ell$, $\widehat{\lambda}_u$ are asymptotically normal

$$\sqrt{n}\begin{pmatrix} \widehat{\lambda}_\ell - \lambda_\ell \\ \widehat{\lambda}_u - \lambda_u \end{pmatrix} \xrightarrow{d} N\left(0, \Sigma\right)$$

Goal: construct a *uniformly valid* and *short* CI for θ

$$\liminf_n \inf_{P \in \mathcal{P}} \inf_{\theta \in [\lambda_{\ell,\min}, \lambda_{u,\max}]} P\left(\theta \in CI\right) \geq 1 - \alpha$$

A Simple Example

Consider the simplest possible case where

$$\theta \in [\min\{\lambda_1, \lambda_2\},\ \max\{\lambda_1, \lambda_2\}]$$

with

$$\begin{pmatrix} \widehat{\lambda}_1 \\ \widehat{\lambda}_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right)$$

A Simple Example

Construct the CI by inverting tests of the hypothesis

$$H_0 : \min\{\lambda_1, \lambda_2\} \leq \theta \leq \max\{\lambda_1, \lambda_2\}$$

A Simple Example

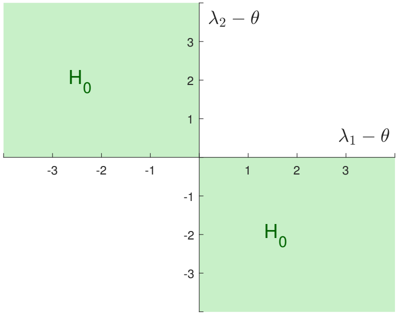
Construct the CI by inverting tests of the hypothesis

$$H_0 : \min\{\lambda_1 - \theta, \lambda_2 - \theta\} \leq 0 \leq \max\{\lambda_1 - \theta, \lambda_2 - \theta\}$$

A Simple Example

Construct the CI by inverting tests of the hypothesis

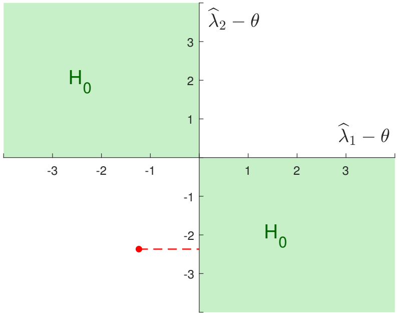
$$H_0 : \min\{\lambda_1 - \theta, \lambda_2 - \theta\} \leq 0 \leq \max\{\lambda_1 - \theta, \lambda_2 - \theta\}$$



A Simple Example

Construct the CI by inverting tests of the hypothesis

$$H_0 : \min\{\lambda_1 - \theta, \lambda_2 - \theta\} \leq 0 \leq \max\{\lambda_1 - \theta, \lambda_2 - \theta\}$$



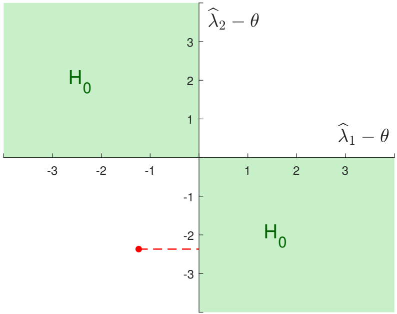
Test stat is

$$\widehat{T}(\theta) = \max\Big\{\min\{\widehat{\lambda}_1 - \theta, \widehat{\lambda}_2 - \theta\}, \min\{\theta - \widehat{\lambda}_1, \theta - \widehat{\lambda}_2\}\Big\}$$

A Simple Example

Construct the CI by inverting tests of the hypothesis

$$H_0 : \min\{\lambda_1 - \theta, \lambda_2 - \theta\} \leq 0 \leq \max\{\lambda_1 - \theta, \lambda_2 - \theta\}$$



Test stat is

$$\widehat{T}(\theta) = \max\Big\{\min\{\widehat{\lambda}_1 - \theta, \widehat{\lambda}_2 - \theta\}, \min\{\theta - \widehat{\lambda}_1, \theta - \widehat{\lambda}_2\}\Big\}$$

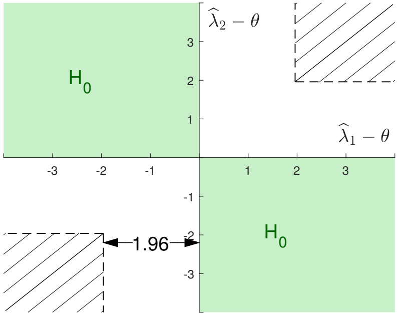
The CI is

$$\left[\min_{\widehat{T}(\theta) \leq c(\theta)} \theta, \max_{\widehat{T}(\theta) \leq c(\theta)} \theta\right]$$

A Simple Example

Construct the CI by inverting tests of the hypothesis

$$H_0 : \min\{\lambda_1 - \theta, \lambda_2 - \theta\} \leq 0 \leq \max\{\lambda_1 - \theta, \lambda_2 - \theta\}$$



Test stat is

$$\widehat{T}(\theta) = \max\Big\{\min\{\widehat{\lambda}_1 - \theta, \widehat{\lambda}_2 - \theta\}, \min\{\theta - \widehat{\lambda}_1, \theta - \widehat{\lambda}_2\}\Big\}$$

The CI is

$$\left[\min_{\widehat{T}(\theta) \leq c(\theta)} \theta, \max_{\widehat{T}(\theta) \leq c(\theta)} \theta\right]$$

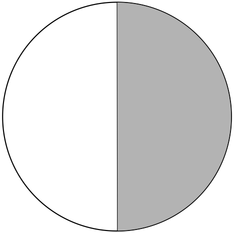
Current practice: $c^{\text{sim}} = \Phi^{-1}(1 - \frac{\alpha}{2})$, 1.96 for $\alpha = 0.05$

A Simple Example - Modified Conditional CV

To improve upon c^{sim}, I propose a modified conditional cv

A Simple Example - Modified Conditional CV

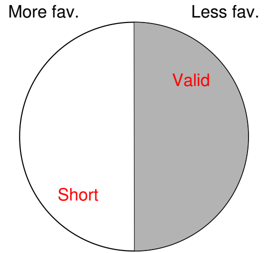
To improve upon c^{sim}, I propose a modified conditional cv



1. Define less and more favorable DGPs

A Simple Example - Modified Conditional CV

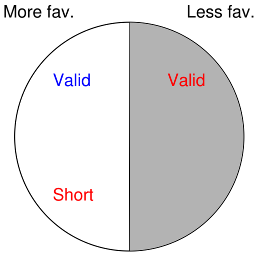
To improve upon c^{sim}, I propose a modified conditional cv



1. Define less and more favorable DGPs
2. Construct a conditional cv

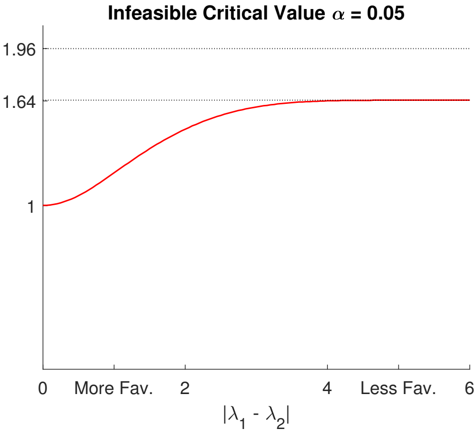
A Simple Example - Modified Conditional CV

To improve upon c^{sim}, I propose a modified conditional cv

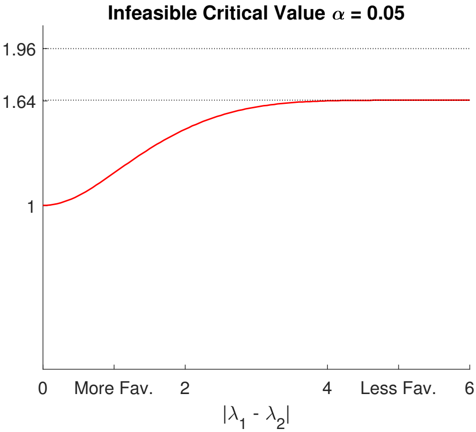


1. Define less and more favorable DGPs
2. Construct a conditional cv
3. Modify the conditional cv

A Simple Example - Step 1 More & Less fav. DGPs

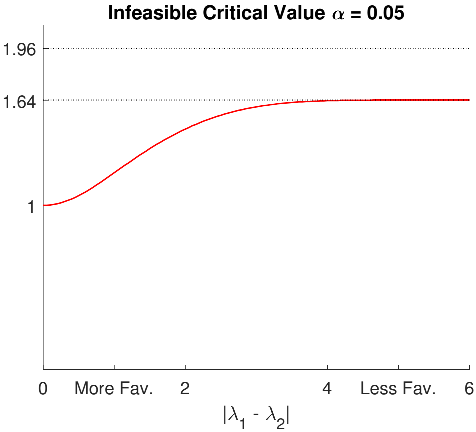


A Simple Example - Step 1 More & Less fav. DGPs



If $\lambda_1 = \lambda_2$, the infeasible CI is $\left[\min\{\widehat{\lambda}_1, \widehat{\lambda}_2\} - 1,\ \max\{\widehat{\lambda}_1, \widehat{\lambda}_2\} + 1\right]$

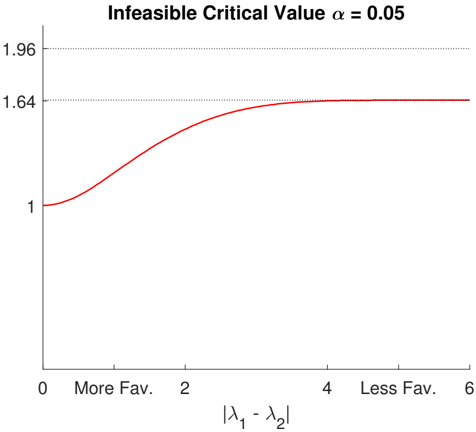
A Simple Example - Step 1 More & Less fav. DGPs



If $\lambda_1 = \lambda_2$, the infeasible CI is $\left[\min\{\widehat{\lambda}_1, \widehat{\lambda}_2\} - 1,\ \max\{\widehat{\lambda}_1, \widehat{\lambda}_2\} + 1\right]$

- we can use $\left[\widehat{\lambda}_1 - 1.96,\ \widehat{\lambda}_1 + 1.96\right]$ *or* $\left[\widehat{\lambda}_2 - 1.96,\ \widehat{\lambda}_2 + 1.96\right]$

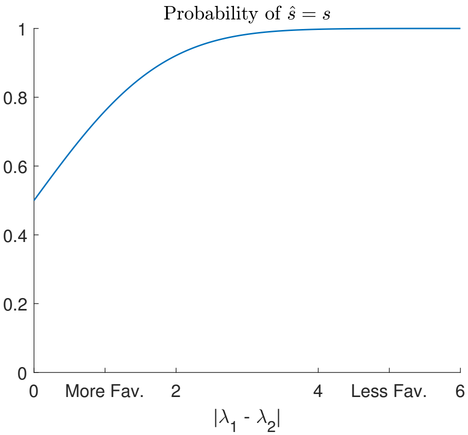
A Simple Example - Step 1 More & Less fav. DGPs



If $\lambda_1 = \lambda_2$, the infeasible CI is $\left[\min\{\widehat{\lambda}_1, \widehat{\lambda}_2\} - 1,\ \max\{\widehat{\lambda}_1, \widehat{\lambda}_2\} + 1\right]$

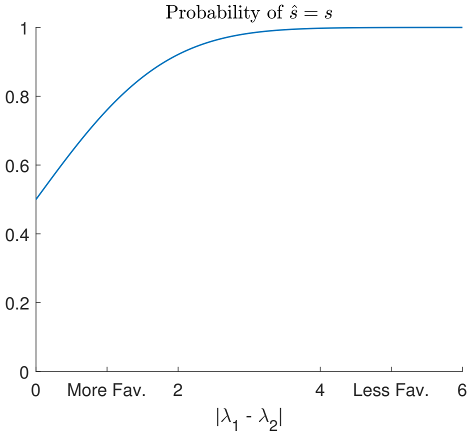
If $\lambda_1 \ll \lambda_2$, the infeasible CI is $\left[\min\{\widehat{\lambda}_1, \widehat{\lambda}_2\} - 1.64,\ \max\{\widehat{\lambda}_1, \widehat{\lambda}_2\} + 1.64\right]$

A Simple Example - Step 2 Conditional CV



where $s = \arg\min_{b=1,2} \lambda_b, \quad \widehat{s} = \arg\min_{b=1,2} \widehat{\lambda}_b$

A Simple Example - Step 2 Conditional CV



where $s = \arg\min_{b=1,2} \lambda_b$, $\quad \widehat{s} = \arg\min_{b=1,2} \widehat{\lambda}_b$

Construct conditional cv based on the distribution of $\widehat{T}(\theta)\Big|\,\widehat{s} = s$

A Simple Example - Step 2 Conditional CV

Construct conditional cv based on the distribution of $\widehat{T}(\theta)\Big|\,\widehat{s} = s$

A Simple Example - Step 2 Conditional CV

Construct conditional cv based on the distribution of $\widehat{T}(\theta)\Big|\,\widehat{s}=s$

Recall

$$\widehat{T}(\theta)=\max\left\{\min\left\{\theta-\widehat{\lambda}_1,\theta-\widehat{\lambda}_2\right\},\min\left\{\widehat{\lambda}_1-\theta,\widehat{\lambda}_2-\theta\right\}\right\}$$

A Simple Example - Step 2 Conditional CV

Construct conditional cv based on the distribution of $\widehat{T}(\theta)\Big|\, \widehat{s} = s$

Recall

$$\widehat{T}(\theta) = \max\left\{\min\left\{\theta - \widehat{\lambda}_1, \theta - \widehat{\lambda}_2\right\}, \min\left\{\widehat{\lambda}_1 - \theta, \widehat{\lambda}_2 - \theta\right\}\right\}$$

Specifically, I consider

$$\widehat{T}(\theta)\,\Big|\,\widehat{T}(\theta) = \theta - \widehat{\lambda}_1, \widehat{s} = s$$

$$\widehat{T}(\theta)\,\Big|\,\widehat{T}(\theta) = \theta - \widehat{\lambda}_2, \widehat{s} = s$$

$$\widehat{T}(\theta)\,\Big|\,\widehat{T}(\theta) = \widehat{\lambda}_1 - \theta, \widehat{s} = s$$

$$\widehat{T}(\theta)\,\Big|\,\widehat{T}(\theta) = \widehat{\lambda}_2 - \theta, \widehat{s} = s$$

A Simple Example - Step 2 Conditional CV

Construct conditional cv based on the distribution of $\widehat{T}(\theta)\Big|\widehat{s}=s$

Recall

$$\widehat{T}(\theta)=\max\left\{\min\left\{\theta-\widehat{\lambda}_1,\theta-\widehat{\lambda}_2\right\},\min\left\{\widehat{\lambda}_1-\theta,\widehat{\lambda}_2-\theta\right\}\right\}$$

Specifically, I consider

$$\widehat{T}(\theta)\,\Big|\,\widehat{T}(\theta)=\theta-\widehat{\lambda}_1,\widehat{s}=s$$

$$\widehat{T}(\theta)\,\Big|\,\widehat{T}(\theta)=\theta-\widehat{\lambda}_2,\widehat{s}=s$$

$$\widehat{T}(\theta)\,\Big|\,\widehat{T}(\theta)=\widehat{\lambda}_1-\theta,\widehat{s}=s$$

$$\widehat{T}(\theta)\,\Big|\,\widehat{T}(\theta)=\widehat{\lambda}_2-\theta,\widehat{s}=s$$

A Simple Example - Step 2 Conditional CV

I illustrate with ▸ Con. CV

$$\widehat{T}(\theta) \,\Big|\, \widehat{T}(\theta) = \widehat{\lambda}_1 - \theta, \widehat{s} = s$$

A Simple Example - Step 2 Conditional CV

I illustrate with ▸ Con. CV

$$\widehat{T}(\theta) \,\Big|\, \widehat{T}(\theta) = \widehat{\lambda}_1 - \theta, \widehat{s} = s \quad \overset{\text{FOSD}}{\preceq} \quad \mathcal{TN}\left(0, \left[\theta - \widehat{\lambda}_2, \widehat{\lambda}_2 - \theta\right]\right)$$

A Simple Example - Step 2 Conditional CV

I illustrate with Con. CV

$$\widehat{T}(\theta) \,\Big|\, \widehat{T}(\theta) = \widehat{\lambda}_1 - \theta, \widehat{s} = s \quad \overset{\text{FOSD}}{\preceq} \quad \mathcal{TN}\left(0, \left[\theta - \widehat{\lambda}_2, \widehat{\lambda}_2 - \theta\right]\right)$$

Let $\alpha^{\text{con}} \in (\frac{\alpha}{2}, \alpha)$ be the conditional rejection rate, recommend $\alpha^{\text{con}} = \frac{4}{5}\alpha$

Define $\widehat{c}^{\text{con}}$ as the $1 - \alpha^{\text{con}}$ quantile of $\mathcal{TN}\left(0, \left[\theta - \widehat{\lambda}_2, \widehat{\lambda}_2 - \theta\right]\right)$

A Simple Example - Step 2 Conditional CV

I illustrate with ▸ Con. CV

$$\widehat{T}(\theta) \,\Big|\, \widehat{T}(\theta) = \widehat{\lambda}_1 - \theta, \widehat{s} = s \qquad \overset{\text{FOSD}}{\preceq} \qquad \mathcal{TN}\left(0, \left[\theta - \widehat{\lambda}_2, \widehat{\lambda}_2 - \theta\right]\right)$$

Let $\alpha^{\text{con}} \in (\frac{\alpha}{2}, \alpha)$ be the conditional rejection rate, recommend $\alpha^{\text{con}} = \frac{4}{5}\alpha$

Define $\widehat{c}^{\text{con}}$ as the $1 - \alpha^{\text{con}}$ quantile of $\mathcal{TN}\left(0, \left[\theta - \widehat{\lambda}_2, \widehat{\lambda}_2 - \theta\right]\right)$

- ✓ Valid size under less favorable DGPs:

$$P\left(\widehat{T}(\theta) > \widehat{c}^{con}\right) = P\left(\widehat{T}(\theta) > \widehat{c}^{con}, \widehat{s} = s\right) + P\left(\widehat{T}(\theta) > \widehat{c}^{con}, \widehat{s} \neq s\right)$$

A Simple Example - Step 2 Conditional CV

I illustrate with ▸ Con. CV

$$\widehat{T}(\theta) \,\Big|\, \widehat{T}(\theta) = \widehat{\lambda}_1 - \theta, \widehat{s} = s \quad \overset{\text{FOSD}}{\preceq} \quad \mathcal{TN}\left(0, \left[\theta - \widehat{\lambda}_2, \widehat{\lambda}_2 - \theta\right]\right)$$

Let $\alpha^{\text{con}} \in (\frac{\alpha}{2}, \alpha)$ be the conditional rejection rate, recommend $\alpha^{\text{con}} = \frac{4}{5}\alpha$

Define $\widehat{c}^{\text{con}}$ as the $1 - \alpha^{\text{con}}$ quantile of $\mathcal{TN}\left(0, \left[\theta - \widehat{\lambda}_2, \widehat{\lambda}_2 - \theta\right]\right)$

✓ Valid size under less favorable DGPs:

$$P\left(\widehat{T}(\theta) > \widehat{c}^{con}\right) \leq P\left(\widehat{T}(\theta) > \widehat{c}^{con}, \widehat{s} = s\right) + P\left(\widehat{s} \neq s\right)$$

A Simple Example - Step 2 Conditional CV

I illustrate with ▸ Con. CV

$$\widehat{T}(\theta) \,\Big|\, \widehat{T}(\theta) = \widehat{\lambda}_1 - \theta, \widehat{s} = s \quad \overset{\text{FOSD}}{\preceq} \quad \mathcal{TN}\left(0, \left[\theta - \widehat{\lambda}_2, \widehat{\lambda}_2 - \theta\right]\right)$$

Let $\alpha^{\text{con}} \in \left(\frac{\alpha}{2}, \alpha\right)$ be the conditional rejection rate, recommend $\alpha^{\text{con}} = \frac{4}{5}\alpha$

Define $\widehat{c}^{\text{con}}$ as the $1 - \alpha^{\text{con}}$ quantile of $\mathcal{TN}\left(0, \left[\theta - \widehat{\lambda}_2, \widehat{\lambda}_2 - \theta\right]\right)$

✓ Valid size under less favorable DGPs:

$$P\left(\widehat{T}(\theta) > \widehat{c}^{con}\right) \leq P\left(\widehat{T}(\theta) > \widehat{c}^{con} \,|\widehat{s} = s\right) P\left(\widehat{s} = s\right) + P\left(\widehat{s} \neq s\right)$$

A Simple Example - Step 2 Conditional CV

I illustrate with ‣ Con. CV

$$\widehat{T}(\theta) \,\Big|\, \widehat{T}(\theta) = \widehat{\lambda}_1 - \theta, \widehat{s} = s \quad \overset{\text{FOSD}}{\preceq} \quad \mathcal{TN}\left(0, \left[\theta - \widehat{\lambda}_2, \widehat{\lambda}_2 - \theta\right]\right)$$

Let $\alpha^{\text{con}} \in (\frac{\alpha}{2}, \alpha)$ be the conditional rejection rate, recommend $\alpha^{\text{con}} = \frac{4}{5}\alpha$

Define $\widehat{c}^{\text{con}}$ as the $1 - \alpha^{\text{con}}$ quantile of $\mathcal{TN}\left(0, \left[\theta - \widehat{\lambda}_2, \widehat{\lambda}_2 - \theta\right]\right)$

✓ Valid size under less favorable DGPs:

$$P\left(\widehat{T}(\theta) > \widehat{c}^{con}\right) \le \underbrace{P\left(\widehat{T}(\theta) > \widehat{c}^{con} \,|\widehat{s} = s\right)}_{\le \alpha^{\mathbf{con}} < \alpha} \underbrace{P\left(\widehat{s} = s\right)}_{\approx 1} + \underbrace{P\left(\widehat{s} \neq s\right)}_{\approx 0}$$

A Simple Example - Step 2 Conditional CV

I illustrate with ▸ Con. CV

$$\widehat{T}(\theta) \Big| \widehat{T}(\theta) = \widehat{\lambda}_1 - \theta, \widehat{s} = s \quad \overset{\text{FOSD}}{\preceq} \quad \mathcal{TN}\left(0, \left[\theta - \widehat{\lambda}_2, \widehat{\lambda}_2 - \theta\right]\right)$$

Let $\alpha^{\text{con}} \in (\frac{\alpha}{2}, \alpha)$ be the conditional rejection rate, recommend $\alpha^{\text{con}} = \frac{4}{5}\alpha$

Define $\widehat{c}^{\text{con}}$ as the $1 - \alpha^{\text{con}}$ quantile of $\mathcal{TN}\left(0, \left[\theta - \widehat{\lambda}_2, \widehat{\lambda}_2 - \theta\right]\right)$

✓ Valid size under less favorable DGPs:

$$P\left(\widehat{T}(\theta) > \widehat{c}^{con}\right) \leq \underbrace{P\left(\widehat{T}(\theta) > \widehat{c}^{con} \,|\widehat{s} = s\right)}_{\leq \alpha^{\textbf{con}} < \alpha} \underbrace{P\left(\widehat{s} = s\right)}_{\approx 1} + \underbrace{P\left(\widehat{s} \neq s\right)}_{\approx 0} \leq \alpha$$

A Simple Example - Step 2 Conditional CV

I illustrate with ▸ Con. CV

$$\widehat{T}(\theta)\,\Big|\,\widehat{T}(\theta)=\widehat{\lambda}_1-\theta,\widehat{s}=s \qquad \overset{\text{FOSD}}{\preceq} \qquad \mathcal{TN}\left(0,\left[\theta-\widehat{\lambda}_2,\widehat{\lambda}_2-\theta\right]\right)$$

Let $\alpha^{\text{con}}\in\left(\frac{\alpha}{2},\alpha\right)$ be the conditional rejection rate, recommend $\alpha^{\text{con}}=\frac{4}{5}\alpha$

Define $\widehat{c}^{\text{con}}$ as the $1-\alpha^{\text{con}}$ quantile of $\mathcal{TN}\left(0,\left[\theta-\widehat{\lambda}_2,\widehat{\lambda}_2-\theta\right]\right)$

✓ Valid size under less favorable DGPs:

$$P\left(\widehat{T}(\theta)>\widehat{c}^{con}\right)\leq P\left(\widehat{T}(\theta)>\widehat{c}^{con}\,|\widehat{s}=s\right)P\left(\widehat{s}=s\right)+P\left(\widehat{s}\neq s\right)\leq\alpha$$

✓ Smaller than c^{sim}

$$\widehat{c}^{con}=\Phi^{-1}\left(\alpha^{con}+(1-2\alpha^{con})\Phi\left(\widehat{\lambda}_2-\theta\right)\right)$$

A Simple Example - Step 2 Conditional CV

I illustrate with ▸ Con. CV

$$\widehat{T}(\theta) \,\Big|\, \widehat{T}(\theta) = \widehat{\lambda}_1 - \theta, \widehat{s} = s \quad \overset{\text{FOSD}}{\preceq} \quad \mathcal{TN}\left(0, \left[\theta - \widehat{\lambda}_2, \widehat{\lambda}_2 - \theta\right]\right)$$

Let $\alpha^{\text{con}} \in (\frac{\alpha}{2}, \alpha)$ be the conditional rejection rate, recommend $\alpha^{\text{con}} = \frac{4}{5}\alpha$

Define $\widehat{c}^{\text{con}}$ as the $1 - \alpha^{\text{con}}$ quantile of $\mathcal{TN}\left(0, \left[\theta - \widehat{\lambda}_2, \widehat{\lambda}_2 - \theta\right]\right)$

✓ Valid size under less favorable DGPs:

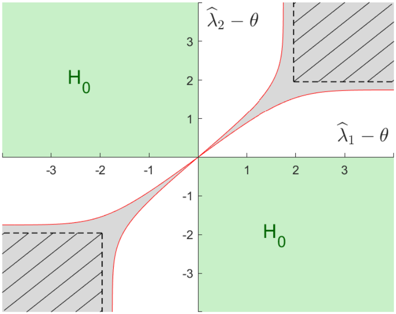
$$P\left(\widehat{T}(\theta) > \widehat{c}^{con}\right) \leq P\left(\widehat{T}(\theta) > \widehat{c}^{con} \,|\widehat{s} = s\right) P\left(\widehat{s} = s\right) + P\left(\widehat{s} \neq s\right) \leq \alpha$$

✓ Smaller than c^{sim}

$$\begin{aligned} \widehat{c}^{con} &= \Phi^{-1}\left(\alpha^{con} + (1 - 2\alpha^{con})\Phi\left(\widehat{\lambda}_2 - \theta\right)\right) \\ &\leq \Phi^{-1}\left(\alpha^{con} + (1 - 2\alpha^{con})\right) \\ &= \Phi^{-1}\left(1 - \alpha^{con}\right) < \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = c^{sim} \end{aligned}$$

A Simple Example - Step 2 Conditional CV

The rejection region of $\hat{c}^{\text{con}}$ for $H_0 : \min\{\lambda_1, \lambda_2\} \leq \theta \leq \max\{\lambda_1, \lambda_2\}$



A Simple Example - Step 3 Modification

I introduce a novel modification

$$c^m(\theta, c^t) = \begin{cases} \widehat{c}^{con}(\theta) & \text{if } \widehat{c}^{con}(\theta) \geq c^t \\ c^t & \text{if } \widehat{c}^{con}(\theta) < c^t \end{cases}$$

A Simple Example - Step 3 Modification

I introduce a novel modification

$$c^m(\theta, c^t) = \begin{cases} \widehat{c}^{con}(\theta) & \text{if } \widehat{c}^{con}(\theta) \geq c^t \\ c^t & \text{if } \widehat{c}^{con}(\theta) < c^t \end{cases}$$

To determine the value of c^{t}, fix $c^{\text{t}} = c$

$$P(\theta \notin CI; \lambda)$$

with $\lambda = (\lambda_1, \lambda_2)$

A Simple Example - Step 3 Modification

I introduce a novel modification

$$c^m(\theta, c^t) = \begin{cases} \widehat{c}^{con}(\theta) & \textit{if } \widehat{c}^{con}(\theta) \geq c^t \\ c^t & \textit{if } \widehat{c}^{con}(\theta) < c^t \end{cases}$$

To determine the value of c^{t}, fix $c^{\text{t}} = c$

$$P(\theta \notin CI; \lambda) \leq \max\left\{P\left([\theta_\ell, \theta_{\text{mid}}] \not\subseteq CI; \lambda\right), P\left([\theta_{\text{mid}}, \theta_u] \not\subseteq CI; \lambda\right)\right\}$$

with $\lambda = (\lambda_1, \lambda_2)$, $\theta_\ell = \min\{\lambda_1, \lambda_2\}$, $\theta_u = \max\{\lambda_1, \lambda_2\}$, $\theta_{\text{mid}} = (\theta_\ell + \theta_u)/2$

A Simple Example - Step 3 Modification

I introduce a novel modification

$$c^m(\theta, c^t) = \begin{cases} \hat{c}^{con}(\theta) & \text{if } \hat{c}^{con}(\theta) \geq c^t \\ c^t & \text{if } \hat{c}^{con}(\theta) < c^t \end{cases}$$

To determine the value of c^{t}, fix $c^{\text{t}} = c$

$$\begin{aligned} P(\theta \notin CI; \lambda) \leq & \max \{ P([\theta_\ell, \theta_{\text{mid}}] \not\subseteq CI; \lambda), P([\theta_{\text{mid}}, \theta_u] \not\subseteq CI; \lambda) \} \\ \leq & \max \{ P(T(\theta_\ell) > c^{\text{m}}(\theta_\ell, c) \text{ or } T(\theta_{\text{mid}}) > c^{\text{m}}(\theta_{\text{mid}}, c); \lambda) \\ & P(T(\theta_u) > c^{\text{m}}(\theta_u, c) \text{ or } T(\theta_{\text{mid}}) > c^{\text{m}}(\theta_{\text{mid}}, c); \lambda) \} \end{aligned}$$

A Simple Example - Step 3 Modification

I introduce a novel modification

$$c^m(\theta, c^t) = \begin{cases} \widehat{c}^{con}(\theta) & \text{if } \widehat{c}^{con}(\theta) \geq c^t \\ c^t & \text{if } \widehat{c}^{con}(\theta) < c^t \end{cases}$$

To determine the value of c^{t}, fix $c^{\mathrm{t}} = c$

$$\begin{aligned} P(\theta \notin CI; \lambda) \leq & \max\left\{P\left([\theta_\ell, \theta_{\text{mid}}] \not\subseteq CI; \lambda\right), P\left([\theta_{\text{mid}}, \theta_u] \not\subseteq CI; \lambda\right)\right\} \\ \leq & \max\left\{P\left(T(\theta_\ell) > c^{\mathrm{m}}(\theta_\ell, c) \text{ or } T(\theta_{\text{mid}}) > c^{\mathrm{m}}(\theta_{\text{mid}}, c); \lambda\right)\right. \\ & \left. P\left(T(\theta_u) > c^{\mathrm{m}}(\theta_u, c) \text{ or } T(\theta_{\text{mid}}) > c^{\mathrm{m}}(\theta_{\text{mid}}, c); \lambda\right)\right\} \\ =: & \bar{p}(c, \lambda) \end{aligned}$$

A Simple Example - Step 3 Modification

I introduce a novel modification

$$c^m(\theta, c^t) = \begin{cases} \widehat{c}^{con}(\theta) & \textit{if } \widehat{c}^{con}(\theta) \geq c^t \\ c^t & \textit{if } \widehat{c}^{con}(\theta) < c^t \end{cases}$$

To determine the value of c^{t}, fix $c^{\mathrm{t}} = c$

$$\begin{aligned} P(\theta \notin CI; \lambda) \leq & \max\left\{P\left([\theta_\ell, \theta_{\text{mid}}] \not\subseteq CI; \lambda\right), P\left([\theta_{\text{mid}}, \theta_u] \not\subseteq CI; \lambda\right)\right\} \\ \leq & \max\left\{P\left(T(\theta_\ell) > c^{\mathrm{m}}(\theta_\ell, c) \text{ or } T(\theta_{\text{mid}}) > c^{\mathrm{m}}(\theta_{\text{mid}}, c); \lambda\right)\right. \\ & \left. P\left(T(\theta_u) > c^{\mathrm{m}}(\theta_u, c) \text{ or } T(\theta_{\text{mid}}) > c^{\mathrm{m}}(\theta_{\text{mid}}, c); \lambda\right)\right\} \\ =: & \bar{p}(c, \lambda) \end{aligned}$$

- $\bar{p}(c, \lambda)$ is easier to calculate

A Simple Example - Step 3 Modification

I introduce a novel modification

$$c^m(\theta, c^t) = \begin{cases} \widehat{c}^{con}(\theta) & \text{if } \widehat{c}^{con}(\theta) \geq c^t \\ c^t & \text{if } \widehat{c}^{con}(\theta) < c^t \end{cases}$$

To determine the value of c^{t}, fix $c^{\mathrm{t}} = c$

$$\begin{aligned} P(\theta \notin CI; \lambda) \leq & \max \left\{ P\left([\theta_\ell, \theta_{\mathrm{mid}}] \not\subseteq CI; \lambda\right), P\left([\theta_{\mathrm{mid}}, \theta_u] \not\subseteq CI; \lambda\right)\right\} \\ \leq & \max \left\{ P\left(T(\theta_\ell) > c^{\mathrm{m}}(\theta_\ell, c) \text{ or } T(\theta_{\mathrm{mid}}) > c^{\mathrm{m}}(\theta_{\mathrm{mid}}, c); \lambda\right) \right. \\ & \left. P\left(T(\theta_u) > c^{\mathrm{m}}(\theta_u, c) \text{ or } T(\theta_{\mathrm{mid}}) > c^{\mathrm{m}}(\theta_{\mathrm{mid}}, c); \lambda\right)\right\} \\ =: & \bar{p}(c, \lambda) \end{aligned}$$

- $\bar{p}(c, \lambda)$ is easier to calculate
- $\bar{p}(c, \lambda)$ is not overly conservative

A Simple Example - Step 3 Modification

I introduce a novel modification

$$c^m(\theta, c^t) = \begin{cases} \widehat{c}^{con}(\theta) & \textit{if } \widehat{c}^{con}(\theta) \geq c^t \\ c^t & \textit{if } \widehat{c}^{con}(\theta) < c^t \end{cases}$$

To determine the value of c^{t}, fix $c^{\mathrm{t}} = c$

$$\begin{aligned} P(\theta \notin CI; \lambda) \leq & \max \left\{ P\left([\theta_\ell, \theta_{\text{mid}}] \not\subseteq CI; \lambda \right), P\left([\theta_{\text{mid}}, \theta_u] \not\subseteq CI; \lambda \right) \right\} \\ \leq & \max \left\{ P\left(T(\theta_\ell) > c^{\mathrm{m}}(\theta_\ell, c) \text{ or } T(\theta_{\text{mid}}) > c^{\mathrm{m}}(\theta_{\text{mid}}, c); \lambda \right) \right. \\ & \quad \left. P\left(T(\theta_u) > c^{\mathrm{m}}(\theta_u, c) \text{ or } T(\theta_{\text{mid}}) > c^{\mathrm{m}}(\theta_{\text{mid}}, c); \lambda \right) \right\} \\ =: & \bar{p}(c, \lambda) \end{aligned}$$

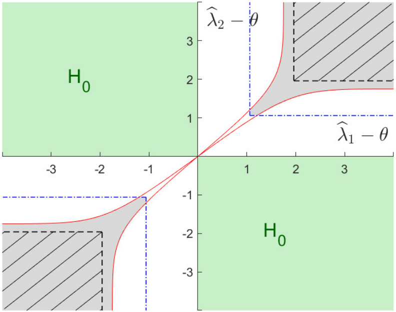
Thus it suffices to have

$$c^t = \inf \left\{ c \geq 0 : \sup_{\lambda \in \Lambda} \bar{p}(c, \lambda) \leq \alpha \right\}$$

This lower truncation guarantees uniform coverage

A Simple Example - Step 3 Modification

The rejection region of $\hat{c}^{\mathrm{m}}$ for $H_0 : \min\{\lambda_1, \lambda_2\} \le \theta \le \max\{\lambda_1, \lambda_2\}$



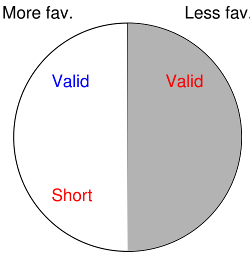
Larger power: the new CI has a larger rejection region

Valid: the lower truncation removes counter-intuitive rejection region

A Simple Example - Summary

The main idea of the modified conditional CI

1. Define less and more favorable DGPs
2. Construct a conditional cv
    - valid under less favorable DGPs
3. Modify the conditional cv
    - valid under more favorable DGPs



General Cases - Test Statistic

Construct CI by inverting

$$H_0 : \min_{b \in \mathcal{B}} \lambda_{\ell,b} \leq \theta \leq \max_{b \in \mathcal{B}} \lambda_{u,b}$$

General Cases - Test Statistic

Construct CI by inverting

$$H_0 : \min_{b \in \mathcal{B}} \lambda_{\ell,b} \leq \theta \leq \max_{b \in \mathcal{B}} \lambda_{u,b}$$

There are normal estimators $\widehat{\lambda}_\ell$, $\widehat{\lambda}_u$

$$\begin{pmatrix} \widehat{\lambda}_\ell - \lambda_\ell \\ \widehat{\lambda}_u - \lambda_u \end{pmatrix} \sim N(0, \Sigma)$$

General Cases - Test Statistic

Construct CI by inverting

$$H_0 : \min_{b \in \mathcal{B}} \lambda_{\ell,b} \leq \theta \leq \max_{b \in \mathcal{B}} \lambda_{u,b}$$

There are normal estimators $\widehat{\lambda}_\ell$, $\widehat{\lambda}_u$

$$\begin{pmatrix} \widehat{\lambda}_\ell - \lambda_\ell \\ \widehat{\lambda}_u - \lambda_u \end{pmatrix} \sim N(0, \Sigma)$$

The test statistic

$$\widehat{T}(\theta) = \max\left\{ \min_{b \in \mathcal{B}} \frac{\widehat{\lambda}_{\ell,b} - \theta}{\sigma_{\ell,b}}, \quad \min_{b \in \mathcal{B}} \frac{\theta - \widehat{\lambda}_{u,b}}{\sigma_{u,b}} \right\}$$

where $\sigma_{\ell,b} = \sqrt{\text{var}(\widehat{\lambda}_{\ell,b})}$, $\sigma_{u,b} = \sqrt{\text{var}(\widehat{\lambda}_{u,b})}$

General Cases - Test Statistic

Construct CI by inverting

$$H_0 : \min_{b \in \mathcal{B}} \lambda_{\ell,b} \leq \theta \leq \max_{b \in \mathcal{B}} \lambda_{u,b}$$

There are normal estimators $\widehat{\lambda}_\ell$, $\widehat{\lambda}_u$

$$\begin{pmatrix} \widehat{\lambda}_\ell - \lambda_\ell \\ \widehat{\lambda}_u - \lambda_u \end{pmatrix} \sim N(0, \Sigma)$$

The test statistic

$$\widehat{T}(\theta) = \max \left\{ \min_{b \in \mathcal{B}} \frac{\widehat{\lambda}_{\ell,b} - \theta}{\sigma_{\ell,b}}, \quad \min_{b \in \mathcal{B}} \frac{\theta - \widehat{\lambda}_{u,b}}{\sigma_{u,b}} \right\}$$

where $\sigma_{\ell,b} = \sqrt{\text{var}(\widehat{\lambda}_{\ell,b})}$, $\sigma_{u,b} = \sqrt{\text{var}(\widehat{\lambda}_{u,b})}$

The simple critical value $c^{\text{sim}} = \Phi^{-1}(1 - \frac{\alpha}{2})$ gives a simple CI

$$CI^{\text{sim}} = \left[\min_{b \in \mathcal{B}} \hat{\lambda}_{\ell,b} - \sigma_{\ell,b}\Phi^{-1}(1 - \frac{\alpha}{2}),\ \max_{b \in \mathcal{B}} \hat{\lambda}_{u,b} + \sigma_{u,b}\Phi^{-1}(1 - \frac{\alpha}{2}) \right]$$

General Cases - Test Statistic

Construct CI by inverting

$$H_0 : \min_{b \in \mathcal{B}} \lambda_{\ell,b} \leq \theta \leq \max_{b \in \mathcal{B}} \lambda_{u,b}$$

There are normal estimators $\widehat{\lambda}_\ell$, $\widehat{\lambda}_u$

$$\begin{pmatrix} \widehat{\lambda}_\ell - \lambda_\ell \\ \widehat{\lambda}_u - \lambda_u \end{pmatrix} \sim N(0, \Sigma)$$

The test statistic

$$\widehat{T}(\theta) = \max \left\{ \min_{b \in \mathcal{B}} \frac{\widehat{\lambda}_{\ell,b} - \theta}{\sigma_{\ell,b}}, \quad \min_{b \in \mathcal{B}} \frac{\theta - \widehat{\lambda}_{u,b}}{\sigma_{u,b}} \right\}$$

where $\sigma_{\ell,b} = \sqrt{\text{var}(\widehat{\lambda}_{\ell,b})}$, $\sigma_{u,b} = \sqrt{\text{var}(\widehat{\lambda}_{u,b})}$

The simple critical value $c^{\text{sim}} = \Phi^{-1}(1 - \frac{\alpha}{2})$ gives a simple CI

$$CI^{\text{sim}} = \left[\min_{b \in \mathcal{B}} \hat{\lambda}_{\ell,b} - \sigma_{\ell,b} \Phi^{-1}(1 - \frac{\alpha}{2}), \ \max_{b \in \mathcal{B}} \hat{\lambda}_{u,b} + \sigma_{u,b} \Phi^{-1}(1 - \frac{\alpha}{2}) \right]$$

c^{sim} is valid but can be very conservative

General Cases - A Simple Critical Value

c^{sim} is valid but can be very conservative

$$P\left(\theta \notin CI^{\text{sim}}\right) = P\left(\theta \notin \left[\min_{b\in\mathcal{B}} \hat{\lambda}_{\ell,b} - \sigma_{\ell,b}\Phi^{-1}(1-\frac{\alpha}{2}),\ \max_{b\in\mathcal{B}} \hat{\lambda}_{u,b} + \sigma_{u,b}\Phi^{-1}(1-\frac{\alpha}{2})\right]\right)$$

General Cases - A Simple Critical Value

c^{sim} is valid but can be very conservative

$$P\left(\theta \notin CI^{\text{sim}}\right) = P\left(\theta \notin \left[\min_{b\in\mathcal{B}} \hat{\lambda}_{\ell,b} - \sigma_{\ell,b}\Phi^{-1}(1-\frac{\alpha}{2}),\ \max_{b\in\mathcal{B}} \hat{\lambda}_{u,b} + \sigma_{u,b}\Phi^{-1}(1-\frac{\alpha}{2})\right]\right)$$
$$\leq P\left(\theta \notin \left[\hat{\lambda}_{\ell,b_\ell} - \sigma_{\ell,b_\ell}\Phi^{-1}(1-\frac{\alpha}{2}),\ \hat{\lambda}_{u,b_u} + \sigma_{u,b_u}\Phi^{-1}(1-\frac{\alpha}{2})\right]\right)$$

where

$$b_\ell = \arg\min_{b\in\mathcal{B}} \lambda_{\ell,b}, \quad b_u = \arg\max_{b\in\mathcal{B}} \lambda_{u,b}$$

- $\leq$ is conservative if $\lambda_{\ell,b_\ell} \approx \min_{b\neq b_\ell} \lambda_{\ell,b}$ or $\lambda_{u,b_u} \approx \min_{b\neq b_u} \lambda_{u,b}$

General Cases - A Simple Critical Value

c^{sim} is valid but can be very conservative

$$
\begin{aligned}
P\left(\theta \notin CI^{\text{sim}}\right) =& P\left(\theta \notin \left[\min_{b\in\mathcal{B}} \hat{\lambda}_{\ell,b} - \sigma_{\ell,b}\Phi^{-1}(1-\frac{\alpha}{2}),\ \max_{b\in\mathcal{B}} \hat{\lambda}_{u,b} + \sigma_{u,b}\Phi^{-1}(1-\frac{\alpha}{2})\right]\right) \\
\leq& P\left(\theta \notin \left[\hat{\lambda}_{\ell,b_\ell} - \sigma_{\ell,b_\ell}\Phi^{-1}(1-\frac{\alpha}{2}),\ \hat{\lambda}_{u,b_u} + \sigma_{u,b_u}\Phi^{-1}(1-\frac{\alpha}{2})\right]\right) \\
\leq& P\left(\theta < \hat{\lambda}_{\ell,b_\ell} - \sigma_{\ell,b_\ell}\Phi^{-1}(1-\frac{\alpha}{2})\right) + P\left(\theta > \hat{\lambda}_{u,b_u} + \sigma_{u,b_u}\Phi^{-1}(1-\frac{\alpha}{2})\right)
\end{aligned}
$$

- $\leq$ is conservative if $\lambda_{\ell,b_\ell} \approx \min_{b\neq b_\ell} \lambda_{\ell,b}$ or $\lambda_{u,b_u} \approx \min_{b\neq b_u} \lambda_{u,b}$
- $\leq$ follows from $P(A\cup B) \leq P(A) + P(B)$

General Cases - A Simple Critical Value

c^{sim} is valid but can be very conservative

$$
\begin{aligned}
P\left(\theta \notin CI^{\text{sim}}\right) =& P\left(\theta \notin \left[\min_{b\in\mathcal{B}} \hat{\lambda}_{\ell,b} - \sigma_{\ell,b}\Phi^{-1}(1-\frac{\alpha}{2}),\ \max_{b\in\mathcal{B}} \hat{\lambda}_{u,b} + \sigma_{u,b}\Phi^{-1}(1-\frac{\alpha}{2})\right]\right) \\
\leq& P\left(\theta \notin \left[\hat{\lambda}_{\ell,b_\ell} - \sigma_{\ell,b_\ell}\Phi^{-1}(1-\frac{\alpha}{2}),\ \hat{\lambda}_{u,b_u} + \sigma_{u,b_u}\Phi^{-1}(1-\frac{\alpha}{2})\right]\right) \\
\leq& P\left(\theta < \hat{\lambda}_{\ell,b_\ell} - \sigma_{\ell,b_\ell}\Phi^{-1}(1-\frac{\alpha}{2})\right) + P\left(\theta > \hat{\lambda}_{u,b_u} + \sigma_{u,b_u}\Phi^{-1}(1-\frac{\alpha}{2})\right) \\
=& P\left(\frac{\hat{\lambda}_{\ell,b_\ell} - \lambda_{\ell,b_\ell}}{\sigma_{\ell,b}} + \frac{\lambda_{\ell,b_\ell} - \theta}{\sigma_{\ell,b}} > \Phi^{-1}(1-\frac{\alpha}{2})\right) \\
& + P\left(\frac{\lambda_{u,b_u} - \hat{\lambda}_{u,b_u}}{\sigma_{u,b}} + \frac{\theta - \lambda_{u,b_u}}{\sigma_{u,b}} > \Phi^{-1}(1-\frac{\alpha}{2})\right)
\end{aligned}
$$

- $\leq$ is conservative if $\lambda_{\ell,b_\ell} \approx \min_{b\neq b_\ell} \lambda_{\ell,b}$ or $\lambda_{u,b_u} \approx \min_{b\neq b_u} \lambda_{u,b}$
- $\leq$ follows from $P(A \cup B) \leq P(A) + P(B)$

General Cases - A Simple Critical Value

c^{sim} is valid but can be very conservative

$$
\begin{aligned}
P\left(\theta \notin CI^{\text{sim}}\right) =& P\left(\theta \notin \left[\min_{b\in\mathcal{B}} \hat{\lambda}_{\ell,b} - \sigma_{\ell,b}\Phi^{-1}(1-\frac{\alpha}{2}),\ \max_{b\in\mathcal{B}} \hat{\lambda}_{u,b} + \sigma_{u,b}\Phi^{-1}(1-\frac{\alpha}{2})\right]\right) \\
\leq& P\left(\theta \notin \left[\hat{\lambda}_{\ell,b_\ell} - \sigma_{\ell,b_\ell}\Phi^{-1}(1-\frac{\alpha}{2}),\ \hat{\lambda}_{u,b_u} + \sigma_{u,b_u}\Phi^{-1}(1-\frac{\alpha}{2})\right]\right) \\
\leq& P\left(\theta < \hat{\lambda}_{\ell,b_\ell} - \sigma_{\ell,b_\ell}\Phi^{-1}(1-\frac{\alpha}{2})\right) + P\left(\theta > \hat{\lambda}_{u,b_u} + \sigma_{u,b_u}\Phi^{-1}(1-\frac{\alpha}{2})\right) \\
=& P\left(\frac{\hat{\lambda}_{\ell,b_\ell} - \lambda_{\ell,b_\ell}}{\sigma_{\ell,b}} + \frac{\lambda_{\ell,b_\ell} - \theta}{\sigma_{\ell,b}} > \Phi^{-1}(1-\frac{\alpha}{2})\right) \\
&+ P\left(\frac{\lambda_{u,b_u} - \hat{\lambda}_{u,b_u}}{\sigma_{u,b}} + \frac{\theta - \lambda_{u,b_u}}{\sigma_{u,b}} > \Phi^{-1}(1-\frac{\alpha}{2})\right) \\
\leq& \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha
\end{aligned}
$$

- $\leq$ is conservative if $\lambda_{\ell,b_\ell} \approx \min_{b\neq b_\ell} \lambda_{\ell,b}$ or $\lambda_{u,b_u} \approx \min_{b\neq b_u} \lambda_{u,b}$
- $\leq$ follows from $P(A \cup B) \leq P(A) + P(B)$
- $\leq$ is conservative if $\lambda_{u,b_u} - \lambda_{\ell,b_\ell} \gg 0$,
  see Imbens and Manski(2004), Stoye(2009)

General Cases - A Simple Critical Value

c^{sim} is valid but can be very conservative

$$
\begin{aligned}
P\left(\theta \notin CI^{\text{sim}}\right) =& P\left(\theta \notin \left[\min_{b\in\mathcal{B}} \hat{\lambda}_{\ell,b} - \sigma_{\ell,b}\Phi^{-1}(1-\frac{\alpha}{2}),\ \max_{b\in\mathcal{B}} \hat{\lambda}_{u,b} + \sigma_{u,b}\Phi^{-1}(1-\frac{\alpha}{2})\right]\right) \\
\leq& P\left(\theta \notin \left[\hat{\lambda}_{\ell,b_\ell} - \sigma_{\ell,b_\ell}\Phi^{-1}(1-\frac{\alpha}{2}),\ \hat{\lambda}_{u,b_u} + \sigma_{u,b_u}\Phi^{-1}(1-\frac{\alpha}{2})\right]\right) \\
\leq& P\left(\theta < \hat{\lambda}_{\ell,b_\ell} - \sigma_{\ell,b_\ell}\Phi^{-1}(1-\frac{\alpha}{2})\right) + P\left(\theta > \hat{\lambda}_{u,b_u} + \sigma_{u,b_u}\Phi^{-1}(1-\frac{\alpha}{2})\right) \\
=& P\left(\frac{\hat{\lambda}_{\ell,b_\ell} - \lambda_{\ell,b_\ell}}{\sigma_{\ell,b}} + \frac{\lambda_{\ell,b_\ell} - \theta}{\sigma_{\ell,b}} > \Phi^{-1}(1-\frac{\alpha}{2})\right) \\
& + P\left(\frac{\lambda_{u,b_u} - \hat{\lambda}_{u,b_u}}{\sigma_{u,b}} + \frac{\theta - \lambda_{u,b_u}}{\sigma_{u,b}} > \Phi^{-1}(1-\frac{\alpha}{2})\right) \\
\leq& \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha
\end{aligned}
$$

- $\leq$ is NOT conservative if $\lambda_{\ell,b_\ell} \ll \min_{b\neq b_\ell} \lambda_{\ell,b}$ and $\lambda_{u,b_u} \gg \min_{b\neq b_u} \lambda_{u,b}$
- $\leq$ follows from $P(A\cup B) \leq P(A) + P(B)$
- $\leq$ is NOT conservative if $\lambda_{u,b_u} = \lambda_{\ell,b_\ell}$
  see Imbens and Manski(2004), Stoye(2009)

General Cases - A Simple Critical Value

c^{sim} is valid but can be very conservative

$$
\begin{aligned}
P\left(\theta \notin CI^{\text{sim}}\right) =& P\left(\theta \notin \left[\min_{b\in\mathcal{B}} \hat{\lambda}_{\ell,b} - \sigma_{\ell,b}\Phi^{-1}(1-\frac{\alpha}{2}),\ \max_{b\in\mathcal{B}} \hat{\lambda}_{u,b} + \sigma_{u,b}\Phi^{-1}(1-\frac{\alpha}{2})\right]\right) \\
\leq& P\left(\theta \notin \left[\hat{\lambda}_{\ell,b_\ell} - \sigma_{\ell,b_\ell}\Phi^{-1}(1-\frac{\alpha}{2}),\ \hat{\lambda}_{u,b_u} + \sigma_{u,b_u}\Phi^{-1}(1-\frac{\alpha}{2})\right]\right) \\
\leq& P\left(\theta < \hat{\lambda}_{\ell,b_\ell} - \sigma_{\ell,b_\ell}\Phi^{-1}(1-\frac{\alpha}{2})\right) + P\left(\theta > \hat{\lambda}_{u,b_u} + \sigma_{u,b_u}\Phi^{-1}(1-\frac{\alpha}{2})\right) \\
=& P\left(\frac{\hat{\lambda}_{\ell,b_\ell} - \lambda_{\ell,b_\ell}}{\sigma_{\ell,b}} + \frac{\lambda_{\ell,b_\ell} - \theta}{\sigma_{\ell,b}} > \Phi^{-1}(1-\frac{\alpha}{2})\right) \\
& + P\left(\frac{\lambda_{u,b_u} - \hat{\lambda}_{u,b_u}}{\sigma_{u,b}} + \frac{\theta - \lambda_{u,b_u}}{\sigma_{u,b}} > \Phi^{-1}(1-\frac{\alpha}{2})\right) \\
\leq& \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha
\end{aligned}
$$

- c^{sim} is nearly optimal among *constant* critical values
- It is crucial to use a *data-dependent* critical value

General Cases - More/ Less fav. DGPs

$$\widehat{T}(\theta) = \max\left\{\min_{b\in\mathcal{B}} \mathcal{Z}_{\ell,b}, \quad \min_{b\in\mathcal{B}} \mathcal{Z}_{u,b}\right\}$$

where $\mathcal{Z}_{\ell,b} = \frac{\widehat{\lambda}_{\ell,b}-\theta}{\sigma_{\ell,b}}$, $\mathcal{Z}_{u,b} = \frac{\widehat{\lambda}_{u,b}-\theta}{\sigma_{u,b}}$

A DGP is less favorable if $q\left(\widehat{T}(\theta), 1-\alpha\right)$ is large

General Cases - More/ Less fav. DGPs

$$\widehat{T}(\theta) = \max\left\{\min_{b\in\mathcal{B}} \mathcal{Z}_{\ell,b}, \quad \min_{b\in\mathcal{B}} \mathcal{Z}_{u,b}\right\}$$

where $\mathcal{Z}_{\ell,b} = \frac{\widehat{\lambda}_{\ell,b}-\theta}{\sigma_{\ell,b}}$, $\mathcal{Z}_{u,b} = \frac{\widehat{\lambda}_{u,b}-\theta}{\sigma_{u,b}}$

A DGP is less favorable if $q\left(\widehat{T}(\theta), 1-\alpha\right)$ is large

The less favorable DGPs

$$\min_{b\in\mathcal{B}\setminus b_\ell} \frac{\lambda_{\ell,b}-\lambda_{\ell,b_\ell}}{\sigma_{\ell,b}} \gg 0, \quad \min_{b\in\mathcal{B}\setminus b_u} \frac{\lambda_{u,b_u}-\lambda_{u,b}}{\sigma_{u,b}} \gg 0, \quad \frac{\lambda_{u,b_u}-\lambda_{\ell,b_\ell}}{\min\left\{\sigma_{\ell,b_\ell}, \sigma_{u,b_u}\right\}} \approx 0 \qquad \text{(LF)}$$

General Cases - More/ Less fav. DGPs

$$\widehat{T}(\theta) = \max\left\{\min_{b\in\mathcal{B}} \mathcal{Z}_{\ell,b}, \quad \min_{b\in\mathcal{B}} \mathcal{Z}_{u,b}\right\}$$

where $\mathcal{Z}_{\ell,b} = \frac{\widehat{\lambda}_{\ell,b}-\theta}{\sigma_{\ell,b}}$, $\mathcal{Z}_{u,b} = \frac{\widehat{\lambda}_{u,b}-\theta}{\sigma_{u,b}}$

A DGP is less favorable if $q\left(\widehat{T}(\theta), 1-\alpha\right)$ is large

The less favorable DGPs

$$\min_{b\in\mathcal{B}\setminus b_\ell} \frac{\lambda_{\ell,b}-\lambda_{\ell,b_\ell}}{\sigma_{\ell,b}} \gg 0, \quad \min_{b\in\mathcal{B}\setminus b_u} \frac{\lambda_{u,b_u}-\lambda_{u,b}}{\sigma_{u,b}} \gg 0, \quad \frac{\lambda_{u,b_u}-\lambda_{\ell,b_\ell}}{\min\left\{\sigma_{\ell,b_\ell}, \sigma_{u,b_u}\right\}} \approx 0 \qquad \text{(LF)}$$

Under (LF)

$$P\left(\left\{\hat{T}(\theta) = \mathcal{Z}_{\ell,b_\ell}\right\} \cup \left\{\hat{T}(\theta) = \mathcal{Z}_{u,b_u}\right\}\right) \approx 1$$

General Cases - More/ Less fav. DGPs

$$\widehat{T}(\theta) = \max\left\{\min_{b\in\mathcal{B}} \mathcal{Z}_{\ell,b}, \quad \min_{b\in\mathcal{B}} \mathcal{Z}_{u,b}\right\}$$

where $\mathcal{Z}_{\ell,b} = \frac{\widehat{\lambda}_{\ell,b}-\theta}{\sigma_{\ell,b}}$, $\mathcal{Z}_{u,b} = \frac{\widehat{\lambda}_{u,b}-\theta}{\sigma_{u,b}}$

A DGP is less favorable if $q\left(\widehat{T}(\theta), 1-\alpha\right)$ is large

The less favorable DGPs

$$\min_{b\in\mathcal{B}\setminus b_\ell} \frac{\lambda_{\ell,b}-\lambda_{\ell,b_\ell}}{\sigma_{\ell,b}} \gg 0, \quad \min_{b\in\mathcal{B}\setminus b_u} \frac{\lambda_{u,b_u}-\lambda_{u,b}}{\sigma_{u,b}} \gg 0, \quad \frac{\lambda_{u,b_u}-\lambda_{\ell,b_\ell}}{\min\left\{\sigma_{\ell,b_\ell}, \sigma_{u,b_u}\right\}} \approx 0 \qquad \text{(LF)}$$

Under (LF)

$$P\left(\left\{\hat{T}(\theta) = \mathcal{Z}_{\ell,b_\ell}\right\} \cup \left\{\hat{T}(\theta) = \mathcal{Z}_{u,b_u}\right\}\right) \approx 1$$

I construct $\hat{c}^{\text{con}}$ based on

$$\hat{T}(\theta)\,\Big|\,\hat{T}(\theta) = \mathcal{Z}_{\ell,b_\ell} \text{ and } \hat{T}(\theta)\,\Big|\,\hat{T}(\theta) = \mathcal{Z}_{u,b_u}$$

General Cases - Conditional CV

Loosely speaking, ▸ Details

$$\hat{T}(\theta) \,\big|\, \hat{T}(\theta) = \mathcal{Z}_{\ell,b_\ell} \overset{\text{FOSD}}{\preceq} \mathcal{TN}\left(0, [t_{\ell,1}(\theta,b),\ t_{\ell,2}(\theta,b)]\right)$$

$$\hat{T}(\theta) \,\big|\, \hat{T}(\theta) = \mathcal{Z}_{u,b_u} \overset{\text{FOSD}}{\preceq} \mathcal{TN}\left(0, [t_{u,1}(\theta,b),\ t_{u,2}(\theta,b)]\right)$$

where

$$t_{\ell,1}(\theta,b) = \min_{\tilde{b}\in\mathcal{B}} \left(1+\rho_{\ell u}(b,\tilde{b})\right)^{-1}\left(\mathcal{Z}_{u,\tilde{b}} + \rho_{\ell u}(b,\tilde{b})\mathcal{Z}_{\ell,b}\right)$$

$$t_{u,1}(\theta,b) = \min_{\tilde{b}\in\mathcal{B}} \left(1-\rho_{\ell}(b,\tilde{b})\right)^{-1}\left(\mathcal{Z}_{\ell,\tilde{b}} - \rho_{\ell}(b,\tilde{b})\mathcal{Z}_{\ell,b}\right)$$

$$t_{\ell,2}(\theta,b) = \min_{\tilde{b}\in\mathcal{B}:\rho_\ell(b,\tilde{b})<1} \left(1-\rho_{\ell}(b,\tilde{b})\right)^{-1}\left(\mathcal{Z}_{\ell,\tilde{b}} - \rho_{\ell}(b,\tilde{b})\mathcal{Z}_{\ell,b}\right)$$

$$t_{u,2}(\theta,b) = \min_{\tilde{b}\in\mathcal{B}:\rho_u(\tilde{b},b)<1} \left(1-\rho_{u}(\tilde{b},b)\right)^{-1}\left(\mathcal{Z}_{u,\tilde{b}} - \rho_{u}(\tilde{b},b)\mathcal{Z}_{u,b}\right)$$

$$\rho_\ell(b_1,b_2) = \frac{\Sigma_{\ell,b_1 b_2}}{\sigma_{\ell,b_1}\sigma_{\ell,b_2}}, \quad \rho_u(b_1,b_2) = \frac{\Sigma_{u,b_1 b_2}}{\sigma_{u,b_1}\sigma_{u,b_2}}, \quad \rho_{\ell u}(b_1,b_2) = \frac{\Sigma_{\ell u,b_1 b_2}}{\sigma_{\ell,b_1}\sigma_{u,b_2}}$$

General Cases - Conditional CV

Define $\hat{c}^{\text{con}}(\theta)$ as $1-\alpha^{\text{c}}$ quantile of $\mathcal{TN}\left(0, \left[t_{\ell,1}(\theta, \hat{b}_\ell), t_{\ell,2}(\theta, \hat{b}_u)\right]\right)$

$$\hat{c}^{\text{c}}(\theta) = \begin{cases} \Phi^{-1}\Big(\alpha^{\text{c}}\Phi\big(t_{\ell,1}(\theta, \hat{b}_\ell)\big) + (1-\alpha^{\text{c}})\Phi\big(t_{\ell,2}(\theta, \hat{b}_\ell)\big)\Big) & \text{if } \hat{T}(\theta) = \mathcal{Z}_{\ell,\hat{b}_\ell} \\ \Phi^{-1}\Big(\alpha^{\text{c}}\Phi\big(t_{u,1}(\theta, \hat{b}_u)\big) + (1-\alpha^{\text{c}})\Phi\big(t_{u,2}(\theta, \hat{b}_u)\big)\Big) & \text{if } \hat{T}(\theta) = \mathcal{Z}_{u,\hat{b}_u} \end{cases}$$

- $\alpha^{\text{c}} \in (\frac{1}{2}\alpha, \alpha)$ is a user chosen tuning parameter, suggested value $\frac{4}{5}\alpha$

General Cases - Conditional CV

Define $\hat{c}^{\text{con}}(\theta)$ as $1-\alpha^{\text{c}}$ quantile of $\mathcal{TN}\left(0, \left[t_{\ell,1}(\theta,\hat{b}_\ell), t_{\ell,2}(\theta,\hat{b}_u)\right]\right)$

$$\hat{c}^{\text{c}}(\theta) = \begin{cases} \Phi^{-1}\Big(\alpha^{\text{c}}\Phi\big(t_{\ell,1}(\theta,\hat{b}_\ell)\big) + (1-\alpha^{\text{c}})\Phi\big(t_{\ell,2}(\theta,\hat{b}_\ell)\big)\Big) & \text{if } \hat{T}(\theta) = \mathcal{Z}_{\ell,\hat{b}_\ell} \\ \Phi^{-1}\Big(\alpha^{\text{c}}\Phi\big(t_{u,1}(\theta,\hat{b}_u)\big) + (1-\alpha^{\text{c}})\Phi\big(t_{u,2}(\theta,\hat{b}_u)\big)\Big) & \text{if } \hat{T}(\theta) = \mathcal{Z}_{u,\hat{b}_u} \end{cases}$$

- $\alpha^{\text{c}} \in \left(\frac{1}{2}\alpha, \alpha\right)$ is a user chosen tuning parameter, suggested value $\frac{4}{5}\alpha$

Proposition

Assume that $P\left(\mathcal{Z}_{\ell,b_\ell} = \mathcal{Z}_{u,b_u}\right) = 0$. *It holds that*

$$P\left(\hat{T}(\theta) > \hat{c}^{con}(\theta) \,\middle|\, E_\ell \cup E_u\right) \leq \alpha^c$$

General Cases - Conditional CV

Define $\hat{c}^{\text{con}}(\theta)$ as $1-\alpha^{\text{c}}$ quantile of $\mathcal{TN}\left(0, \left[t_{\ell,1}(\theta,\hat{b}_\ell), t_{\ell,2}(\theta,\hat{b}_u)\right]\right)$

$$\hat{c}^{\text{c}}(\theta) = \begin{cases} \Phi^{-1}\Big(\alpha^{\text{c}}\Phi\big(t_{\ell,1}(\theta,\hat{b}_\ell)\big) + (1-\alpha^{\text{c}})\Phi\big(t_{\ell,2}(\theta,\hat{b}_\ell)\big)\Big) & \text{if } \hat{T}(\theta) = \mathcal{Z}_{\ell,\hat{b}_\ell} \\ \Phi^{-1}\Big(\alpha^{\text{c}}\Phi\big(t_{u,1}(\theta,\hat{b}_u)\big) + (1-\alpha^{\text{c}})\Phi\big(t_{u,2}(\theta,\hat{b}_u)\big)\Big) & \text{if } \hat{T}(\theta) = \mathcal{Z}_{u,\hat{b}_u} \end{cases}$$

- $\alpha^{\text{c}} \in (\frac{1}{2}\alpha, \alpha)$ is a user chosen tuning parameter, suggested value $\frac{4}{5}\alpha$

Proposition

Assume that $P\left(\mathcal{Z}_{\ell,b_\ell} = \mathcal{Z}_{u,b_u}\right) = 0$. *It holds that*

$$P\left(\hat{T}(\theta) > \hat{c}^{con}(\theta)\middle| E_\ell \cup E_u\right) \le \alpha^c$$

$\hat{c}^{\text{con}}(\theta)$ is **valid** under less favorable DGPs

$$P\left(\hat{T}(\theta) > \hat{c}^{\text{con}}(\theta)\right) \le \underbrace{P\left(\hat{T}(\theta) > \hat{c}^{\text{con}}(\theta)\,|E_\ell \cup E_u\right)}_{\le \alpha^{\text{con}}} \underbrace{P(E_\ell \cup E_u)}_{\approx 1} + \underbrace{P\left(\overline{E_\ell \cup E_u}\right)}_{\approx 0} \le \alpha$$

General Cases - Conditional CV

$\hat{c}^{\text{con}}(\theta)$ can be significantly **smaller** than c^{sim} under more favorable DGPs

General Cases - Conditional CV

$\hat{c}^{\text{con}}(\theta)$ can be significantly **smaller** than c^{sim} under more favorable DGPs

Let $\theta = \lambda_{\ell, b_\ell}$

$$\hat{c}^{\text{con}}(\theta) = \Phi^{-1}\Big(\alpha^{\text{c}}\Phi\big(t_1\big) + (1-\alpha^{\text{c}})\Phi\big(t_2\big)\Big)$$

General Cases - Conditional CV

$\hat{c}^{\text{con}}(\theta)$ can be significantly **smaller** than c^{sim} under more favorable DGPs

Let $\theta = \lambda_{\ell,b_\ell}$

$$\hat{c}^{\text{con}}(\theta) = \Phi^{-1}\Big(\alpha^{c}\Phi\big(t_1\big) + (1-\alpha^{c})\Phi\big(t_2\big)\Big)$$

If the identified set is large

$$\begin{aligned} t_1 &= \min_{b \in \mathcal{B}} \big(1 + \rho_{\ell u}(\hat{b}_\ell, b)\big)^{-1} \Big(\mathcal{Z}_{u,b} + \rho_{\ell u}(\hat{b}_\ell, b)\mathcal{Z}_{\ell,\hat{b}_\ell}\Big) \\ &\leq \big(1 + \rho_{\ell u}(\hat{b}_\ell, b_u)\big)^{-1} \left(\frac{\hat{\lambda}_{\ell,\hat{b}_\ell} - \hat{\lambda}_{u,b_u}}{\sigma_{u,b_u}} + \left(\rho_{\ell u}(\hat{b}_\ell, b_u) - \frac{\sigma_{\ell,\hat{b}_\ell}}{\sigma_{u,b_u}}\right)\mathcal{Z}_{\ell,\hat{b}_\ell}\right) \approx -\infty \end{aligned}$$

General Cases - Conditional CV

$\hat{c}^{\text{con}}(\theta)$ can be significantly **smaller** than c^{sim} under more favorable DGPs

Let $\theta = \lambda_{\ell,b_\ell}$

$$\hat{c}^{\text{con}}(\theta) = \Phi^{-1}\Big(\alpha^{\text{c}}\Phi\big(t_1\big) + (1-\alpha^{\text{c}})\Phi\big(t_2\big)\Big)$$

If the identified set is large

$$t_1 = \min_{b\in\mathcal{B}} \big(1+\rho_{\ell u}(\hat{b}_\ell, b)\big)^{-1}\left(\mathcal{Z}_{u,b} + \rho_{\ell u}(\hat{b}_\ell, b)\mathcal{Z}_{\ell,\hat{b}_\ell}\right)$$
$$\leq \big(1+\rho_{\ell u}(\hat{b}_\ell, b_u)\big)^{-1}\left(\frac{\hat{\lambda}_{\ell,\hat{b}_\ell} - \hat{\lambda}_{u,b_u}}{\sigma_{u,b_u}} + \left(\rho_{\ell u}(\hat{b}_\ell, b_u) - \frac{\sigma_{\ell,\hat{b}_\ell}}{\sigma_{u,b_u}}\right)\mathcal{Z}_{\ell,\hat{b}_\ell}\right) \approx -\infty$$

Therefore,

$$\hat{c}^{\text{con}}(\theta) \approx \Phi^{-1}\Big(0 + (1-\alpha^{\text{c}})\Phi\big(t_{\ell,2}\big)\Big) \leq \Phi^{-1}\Big(1-\alpha^{\text{c}}\Big) < \Phi^{-1}\Big(1-\frac{\alpha}{2}\Big)$$

General Cases - Conditional CV

$\hat{c}^{\text{con}}(\theta)$ can be significantly **smaller** than c^{sim} under more favorable DGPs

Let $\theta = \lambda_{\ell, b_\ell}$

$$\hat{c}^{\text{con}}(\theta) = \Phi^{-1}\Big(\alpha^{\text{c}}\Phi\big(t_1\big) + (1-\alpha^{\text{c}})\Phi\big(t_2\big)\Big)$$

If the identified set is large

$$t_1 = \min_{b \in \mathcal{B}} \left(1 + \rho_{\ell u}(\hat{b}_\ell, b)\right)^{-1} \left(\mathcal{Z}_{u,b} + \rho_{\ell u}(\hat{b}_\ell, b)\mathcal{Z}_{\ell, \hat{b}_\ell}\right)$$
$$\leq \left(1 + \rho_{\ell u}(\hat{b}_\ell, b_u)\right)^{-1} \left(\frac{\hat{\lambda}_{\ell, \hat{b}_\ell} - \hat{\lambda}_{u, b_u}}{\sigma_{u, b_u}} + \left(\rho_{\ell u}(\hat{b}_\ell, b_u) - \frac{\sigma_{\ell, \hat{b}_\ell}}{\sigma_{u, b_u}}\right)\mathcal{Z}_{\ell, \hat{b}_\ell}\right) \approx -\infty$$

Therefore,

$$\hat{c}^{\text{con}}(\theta) \approx \Phi^{-1}\Big(0 + (1-\alpha^{\text{c}})\Phi\big(t_{\ell,2}\big)\Big) \leq \Phi^{-1}\Big(1-\alpha^{\text{c}}\Big) < \Phi^{-1}\Big(1 - \frac{\alpha}{2}\Big)$$

General Cases - Conditional CV

$\hat{c}^{\text{con}}(\theta)$ can be significantly **smaller** than c^{sim} under more favorable DGPs

Let $\theta = \lambda_{\ell,b_\ell}$

$$\hat{c}^{\text{con}}(\theta) = \Phi^{-1}\Big(\alpha^{\text{c}}\Phi\big(t_1\big) + (1-\alpha^{\text{c}})\Phi\big(t_2\big)\Big)$$

If the identified set is large

$$t_1 \approx -\infty$$

Moreover, if λ_ℓ is not well separated

$$t_2 = \min_{b \in \mathcal{B}} \big(1 - \rho_\ell(\hat{b}_\ell, b)\big)^{-1} \left(\frac{\hat{\lambda}_{\ell,b} - \lambda_{\ell,b}}{\sigma_{\ell,b}} + \frac{\lambda_{\ell,b} - \lambda_{\ell,b_\ell}}{\sigma_{\ell,b}} - \mathcal{Z}_{\ell,\hat{b}_\ell} \right) + \mathcal{Z}_{\ell,\hat{b}_\ell} < \infty$$

which further reduces $\hat{c}^{\text{con}}(\theta)$

General Cases - Modification

The modified conditional critical value

$$c^{\mathrm{m}}(\theta, c^{\mathrm{t}}) = \begin{cases} \widehat{c}^{\mathrm{con}}(\theta) & \text{if } \widehat{c}^{\mathrm{con}}(\theta) \geq c^{\mathrm{t}} \\ c^{\mathrm{t}} & \text{if } \widehat{c}^{\mathrm{con}}(\theta) < c^{\mathrm{t}} \end{cases}$$

The lower truncation c^{t} is

$$c^{\mathrm{t}} = \inf_{c} \left\{ c \geq 0 : \sup_{\lambda \in \Lambda} \bar{p}(c, \lambda) \leq \alpha \right\}$$

where $\theta_\ell = \min_{b \in \mathcal{B}} \lambda_{\ell,b}$, $\theta_u = \max_{b \in \mathcal{B}} \lambda_{u,b}$, $\theta_m = (\theta_\ell + \theta_u)/2$

$$\begin{aligned} \bar{p}(c, \lambda) = \max \Big\{ & P\left(\widehat{T}(\theta_\ell) > c^{\mathrm{m}}(\theta_\ell, c) \text{ or } \widehat{T}(\theta_m) > c^{\mathrm{m}}(\theta_m, c); (\lambda, \Sigma) \right), \\ & P\left(\widehat{T}(\theta_m) > c^{\mathrm{m}}(\theta_m, c) \text{ or } \widehat{T}(\theta_u) > c^{\mathrm{m}}(\theta_u, c); (\lambda, \Sigma) \right) \Big\} \end{aligned}$$

General Cases - Modification

The modified conditional critical value

$$c^{\mathrm{m}}(\theta, c^{\mathrm{t}}) = \begin{cases} \widehat{c}^{\mathrm{con}}(\theta) & \text{if } \widehat{c}^{\mathrm{con}}(\theta) \geq c^{\mathrm{t}} \\ c^{\mathrm{t}} & \text{if } \widehat{c}^{\mathrm{con}}(\theta) < c^{\mathrm{t}} \end{cases}$$

The lower truncation c^{t} is

$$c^{\mathrm{t}} = \inf_{c} \left\{ c \geq 0 : \sup_{\lambda \in \Lambda_\eta} \bar{p}(c, \lambda) \leq \alpha - \eta \right\}$$

where $\theta_\ell = \min_{b \in \mathcal{B}} \lambda_{\ell,b}$, $\theta_u = \max_{b \in \mathcal{B}} \lambda_{u,b}$, $\theta_m = (\theta_\ell + \theta_u)/2$

$$\bar{p}(c, \lambda) = \max \Big\{ P\left(\widehat{T}(\theta_\ell) > c^{\mathrm{m}}(\theta_\ell, c) \text{ or } \widehat{T}(\theta_m) > c^{\mathrm{m}}(\theta_m, c); (\lambda, \Sigma) \right), \\ P\left(\widehat{T}(\theta_m) > c^{\mathrm{m}}(\theta_m, c) \text{ or } \widehat{T}(\theta_u) > c^{\mathrm{m}}(\theta_u, c); (\lambda, \Sigma) \right) \Big\}$$

Size & Power: Assumptions

Assumption (Known Singularity (KS))

There are known $|\mathcal{B}| \times J$ *matrices* A_ℓ, A_u *such that for some* $(\delta_P, \widehat{\delta}_n)$

$$\lambda_\ell = A_\ell \delta_P,\ \lambda_u = A_u \delta_P,\ \widehat{\lambda}_\ell = A_\ell \widehat{\delta}_n,\ \widehat{\lambda}_u = A_u \widehat{\delta}_n$$

Size & Power: Assumptions

Assumption (Known Singularity (KS))

There are known $|\mathcal{B}| \times J$ *matrices* A_ℓ, A_u *such that for some* $(\delta_P, \widehat{\delta}_n)$

$$\lambda_\ell = A_\ell \delta_P,\ \lambda_u = A_u \delta_P,\ \widehat{\lambda}_\ell = A_\ell \widehat{\delta}_n,\ \widehat{\lambda}_u = A_u \widehat{\delta}_n$$

Assumption (Asymptotic Normality (AN))

Let $\xi_P \sim \mathcal{N}(0, \Omega_P)$*. Assume*

$$\lim_{n\to\infty} \sup_{P\in\mathcal{P}} \sup_{f\in BL_1} \left| E_P\left[f\left(\sqrt{n}\left(\widehat{\delta}_n - \delta_P\right)\right)\right] - E\left[f\left(\xi_P\right)\right]\right| = 0$$

Assumption (Full Rank (FR))

For all $P \in \mathcal{P}$, $0 < \underline{e} \leq eig(\Omega_P) \leq \bar{e} < \infty$

Size & Power: Assumptions

Assumption (Known Singularity (KS))

There are known $|\mathcal{B}| \times J$ *matrices* A_ℓ, A_u *such that for some* $(\delta_P, \widehat{\delta}_n)$

$$\lambda_\ell = A_\ell \delta_P,\ \lambda_u = A_u \delta_P,\ \widehat{\lambda}_\ell = A_\ell \widehat{\delta}_n,\ \widehat{\lambda}_u = A_u \widehat{\delta}_n$$

Assumption (Asymptotic Normality (AN))

Let $\xi_P \sim \mathcal{N}(0, \Omega_P)$. *Assume*

$$\lim_{n\to\infty} \sup_{P\in\mathcal{P}} \sup_{f\in BL_1} \left| E_P\left[f\left(\sqrt{n}\left(\widehat{\delta}_n - \delta_P \right)\right)\right] - E\left[f(\xi_P)\right] \right| = 0$$

Assumption (Full Rank (FR))

For all $P \in \mathcal{P}$, $0 < \underline{e} \leq eig(\Omega_P) \leq \bar{e} < \infty$

Assumption (Consistent Covariance Estimator (CE))

For all $\varepsilon > 0$, $\lim_{n\to\infty} \sup_{P\in\mathcal{P}} P\left(\left\| \widehat{\Omega}_n - \Omega_P \right\| > \varepsilon \right) = 0$

Size & Power: Asymptotic Size Properties

Theorem (Uniform Coverage)

Suppose Assumptions KS, AN, FR, CE hold, for any $\alpha \in (0, 0.5)$*,*

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} \inf_{\theta\in\left[\lambda_{\ell,b_\ell},\lambda_{u,b_u}\right]} P\left(\theta \in CI^m\right) \geq 1-\alpha$$

The modified conditional CI has proper asymptotic coverage

Size & Power: Comparison with Simple CI

Theorem (Symmetric Or Large Bounds)

Suppose Assumptions KS, AN, FR, CE hold, $\alpha \in (0, \frac{1}{2})$

(Symmetric Bounds) *If* $corr(\widehat{\lambda}_{\ell,b_1}, \widehat{\lambda}_{\ell,b_2}) < \rho_1^*(\alpha, \alpha^c)$, $corr(\widehat{\lambda}_{\ell,b_\ell}, \widehat{\lambda}_{\ell,b_u}) < \rho_2^*(\alpha)$

$$\widehat{\lambda}_\ell = \widehat{\lambda}_u$$

Then

1. *My CI is Strictly Shorter:*
   There is $\alpha' > \alpha$ *such that*
   $$\liminf_n P\left(CI^m\left(\hat{\lambda}_n, \hat{\Sigma}_n/n; \alpha\right) \subseteq CI^{sim}\left(\hat{\lambda}_n, \hat{\Sigma}_n/n; \alpha'\right)\right) = 1$$
2. *My CI has Higher Power:*
   There is $\kappa \in (0, +\infty)$ *such that*
   $$\liminf_n P\left(\theta_n \notin CI^m\left(\hat{\lambda}, \hat{\Sigma}/n; \alpha\right)\right) - P\left(\theta_n \notin CI^{sim}\left(\hat{\lambda}, \hat{\Sigma}/n; \alpha\right)\right) > 0$$
   for some $\theta_n = \theta_\ell - \frac{\kappa}{\sqrt{n}}$.

▸ Uniform

Size & Power: Comparison with Simple CI

Theorem (Symmetric Or Large Bounds)

Suppose Assumptions KS, AN, FR, CE hold, $\alpha \in (0, \frac{1}{2})$

(Large Bounds)

$$\max_{b \in \mathcal{B}} \lambda_{u,b} - \min_{b \in \mathcal{B}} \lambda_{\ell,b} \geq \varepsilon > 0$$

Then

1. *My CI is Strictly Shorter:*
   There is $\alpha' > \alpha$ *such that*
   $$\liminf_n P\left(CI^m\left(\hat{\lambda}_n, \hat{\Sigma}_n / n; \alpha\right) \subseteq CI^{sim}\left(\hat{\lambda}_n, \hat{\Sigma}_n / n; \alpha'\right)\right) = 1$$
2. *My CI has Higher Power:*
   There is $\kappa \in (0, +\infty)$ *such that*
   $$\liminf_n P\left(\theta_n \notin CI^m\left(\hat{\lambda}, \hat{\Sigma}/n; \alpha\right)\right) - P\left(\theta_n \notin CI^{sim}\left(\hat{\lambda}, \hat{\Sigma}/n; \alpha\right)\right) > 0$$
   for some $\theta_n = \theta_\ell - \frac{\kappa}{\sqrt{n}}$.

▸ Uniform

Size & Power: Comparison with Adj Boot CI

Let CI^{AdjBoot} be the adjusted bootstrap procedure proposed in YKHS23

- use a subsample with size $m = \frac{n}{\kappa_n}$, $\kappa_n \to \infty$ and $\kappa_n = o(n)$
- use the full sample to estimate the nuisance para., e.g. $\sqrt{m}(\lambda_{\ell,b} - \min \lambda_\ell)$

Theorem (Larger Local Power)

Let $a > 0$, $\kappa'_n = o(\sqrt{\kappa_n})$, $\kappa'_n \to \infty$. *Then*

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}_n} \left\{ P\left(\theta_n \notin CI^m\right) - P\left(\theta_n \notin CI^{AdjBoot}\right) \right\} \geq 1 - \alpha$$

for local alternatives

$$\theta_n = \min_{b\in\mathcal{B}} \lambda_{\ell,b} - \frac{\kappa'_n}{\sqrt{n}} a \text{ or } \theta_n = \max_{b\in\mathcal{B}} \lambda_{u,b} + \frac{\kappa'_n}{\sqrt{n}} a$$

Outline

Setting

Relaxation of the parallel trend assumptions, where

$$\theta = ATT \in \left[\min_{b \in \mathcal{B}} \lambda_{\ell,b},\ \max_{b \in \mathcal{B}} \lambda_{u,b} \right]$$

where $\mathcal{B} = \{-(T-1), ..., T\}$,

$$\lambda_\ell = \lambda_u = \begin{cases} \gamma + \Delta_\beta & \text{if } \beta = -(T-1), ..., 0 \\ \gamma - \Delta_{-(\beta-1)} & \text{if } \beta = 1, ..., T \end{cases}$$

Setting

I conduct inference based on $\left(\widehat{\Delta}, \widehat{\gamma}, \Omega\right)$ where $(\widehat{\Delta}, \widehat{\gamma})$ is simulated from

$$\begin{pmatrix} \widehat{\Delta} \\ \widehat{\gamma} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \Delta \\ \gamma \end{pmatrix}, \Omega\right)$$

I use four Ω calibrated from

1. Benzarti and Carloni (2019): consumption tax cuts
2. Dustmann, Lindner, Schönberg, Umkehrer, Vom Berge (2022): minimum wage
3. Lovenheim and Willén (2019): teacher collective bargaining
4. Christensen, Keiser, Lade (2023): environmental crises

I normalized $\gamma = 0$ and use three Δ

1. Parallel trends assumption holds, i.e. $\Delta = 0_T$
2. Small pre-trends: Δ is calibrated
3. One large pre-trend: $\Delta = (10\omega_M, 0_{T-1})$

In sum, I use $4 \times 3 = 12$ empirically motivated DGPs ▸ Details

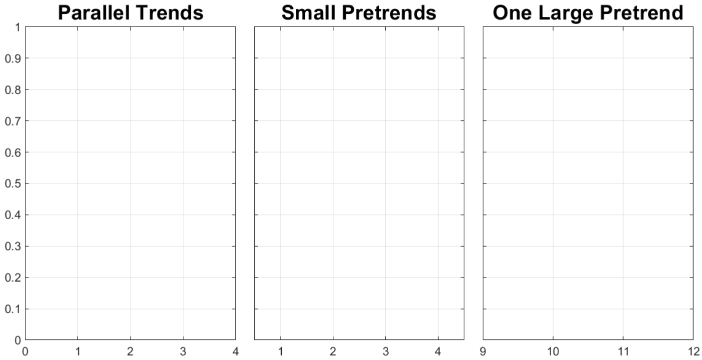
Alternative Methods

I compare my CI with

1. Simple CI in, e.g., Kolesár and Rothe (2018, AER)
2. Hybrid CI in Rambachan and Roth (2023, RES)
3. Adjusted bootstrap in Ye, Keele, Hasegawa and Small (2023, JASA)

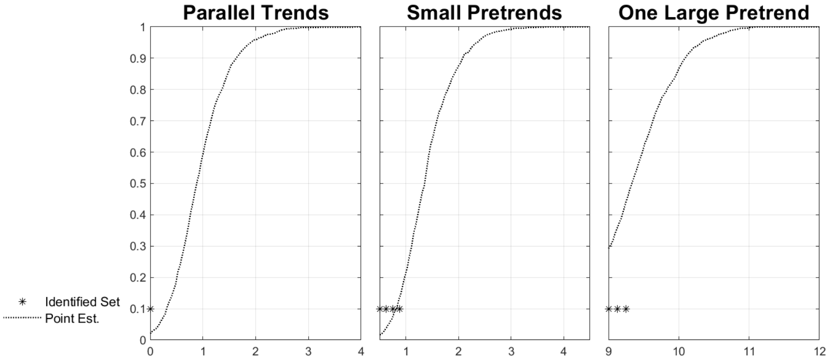
CDF of Upper Bounds of CIs

DGP: Ω from Lovenheim and Willén (2019)



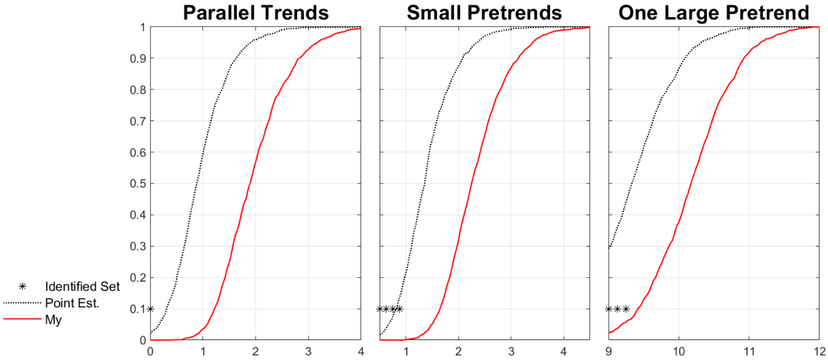
CDF of Upper Bounds of CIs

DGP: Ω from Lovenheim and Willén (2019)



CDF of Upper Bounds of CIs

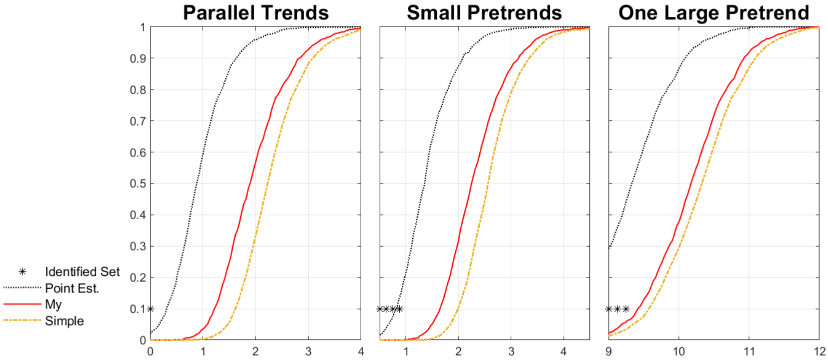
DGP: Ω from Lovenheim and Willén (2019)



Modified conditional CI has proper coverage

CDF of Upper Bounds of CIs

DGP: Ω from Lovenheim and Willén (2019)

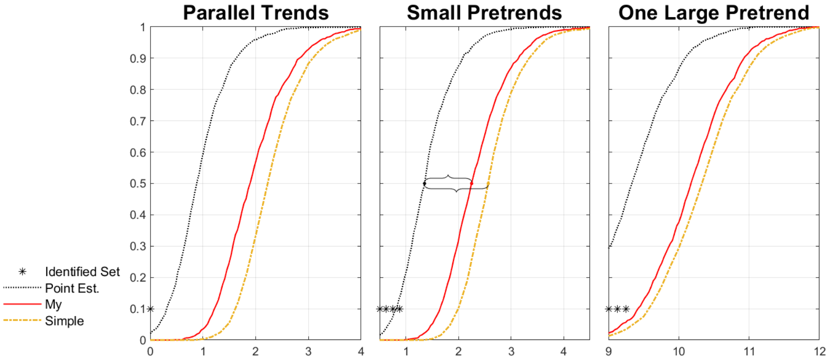


Modified conditional CI outperforms simple CI in all DGPs

- Reduces median simple CI (net of point est.) by **31%** under small violation

CDF of Upper Bounds of CIs

DGP: Ω from Lovenheim and Willén (2019)

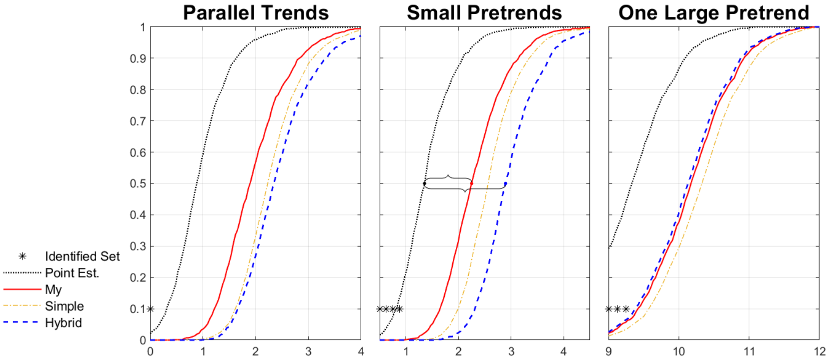


Modified conditional CI outperforms simple CI in all DGPs

- Reduces median simple CI (net of point est.) by **31%** under small violation

CDF of Upper Bounds of CIs

DGP: Ω from Lovenheim and Willén (2019)

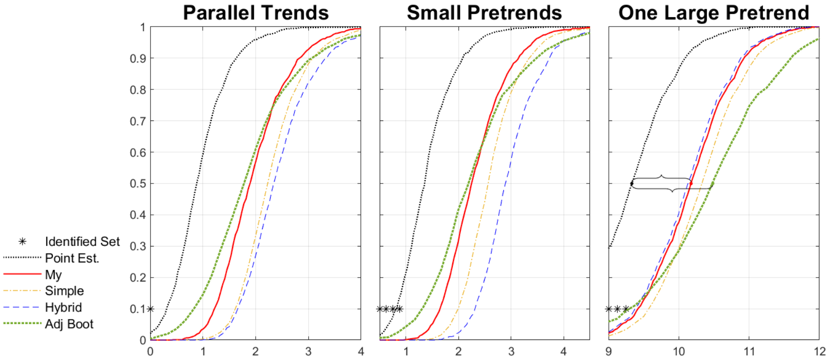


Modified conditional CI outperforms Hybrid CI under no or small violations

- Reduces median Hybrid CI (net of point est.) by **43%** under small violation
- Hybrid CI is efficient with large violation, but my CI is close

CDF of Upper Bounds of CIs

DGP: Ω from Lovenheim and Willén (2019)

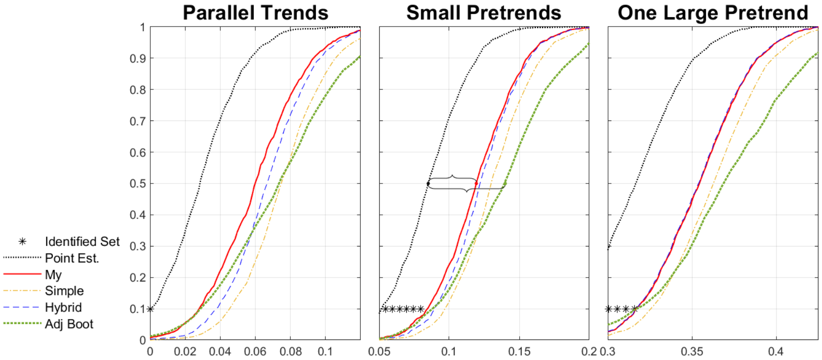


Modified conditional CI outperforms Adj Boot CI for relatively large alternatives

- Reduces median Adj Boot CI (net of point est.) by **27%** under large violation
- Power plot of Adj Boot will be flatter with larger *n*

CDF of Upper Bounds of CIs

DGP: Ω from Benzarti and Carloni (2019)



Modified conditional CI outperforms Adj Boot CI for relatively large alternatives

- Reduces median Adj Boot CI (net of point est.) by **38%** under small violation
- Power plot of Adj Boot CI will be flatter with larger n

Outline

Empirical Illustration

RR23 in Dustmann, Lindner, Schönberg, Umkehrer, Vom Berge (2022, QJE)

What are the effects of the minimum wage?

- Addresses wage inequality
- Potential disemployment

Empirical Illustration

RR23 in Dustmann, Lindner, Schönberg, Umkehrer, Vom Berge (2022, QJE)

What are the effects of the minimum wage?

- Addresses wage inequality $\Leftarrow$ Significant Wage Effect
- Potential disemployment $\Leftarrow$ Insignificant Employment Effect

Empirical Illustration

RR23 in Dustmann, Lindner, Schönberg, Umkehrer, Vom Berge (2022, QJE)

What are the effects of the minimum wage?

- Is employment effect ≥ -0.6 (wage effect) w/o parallel trends? DiD

The authors relax parallel trends as in RR23 SDRM

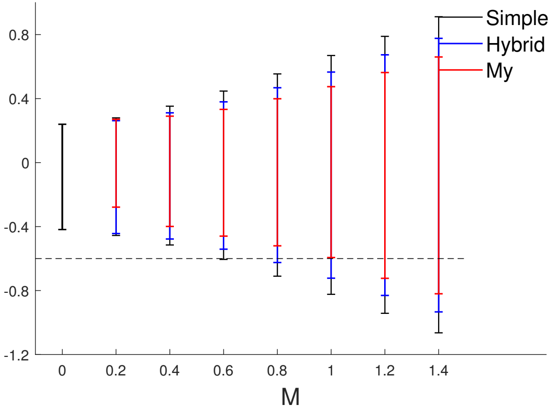
Empirical Illustration

RR23 in Dustmann, Lindner, Schönberg, Umkehrer, Vom Berge (2022, QJE)

What are the effects of the minimum wage?

- Is employment effect ≥ -0.6 (wage effect) w/o parallel trends? DiD

The authors relax parallel trends as in RR23 SDRM



- The breakdown M: $M^{\text{My}} = 1$, $M^{\text{Hybrid}} = 0.75$, $M^{\text{sim}} = 0.6$

Outline

Conclusion

This paper studies inference on **union bounds**

$$\theta \in \left[\min_{b\in\mathcal{B}} \lambda_{\ell,b},\ \max_{b\in\mathcal{B}} \lambda_{u,b}\right]$$

I propose a CI based on *modified conditional* inference which

- Theory & Simulation: has shorter CI & larger local power under a large set of DGPs
- Empirical illustration: gives statistically significant results while the pre-existing alternatives do not

Conclusion

This paper studies inference on **union bounds**

$$\theta \in \left[\min_{b \in \mathcal{B}} \lambda_{\ell,b}, \ \max_{b \in \mathcal{B}} \lambda_{u,b} \right]$$

I propose a CI based on *modified conditional* inference which

- Theory & Simulation: has shorter CI & larger local power under a large set of DGPs
- Empirical illustration: gives statistically significant results while the pre-existing alternatives do not

Future Extension

- Optimal choice of α^c
- Application to other non-standard inference problems
- Misspecification robust inference

Finite $\mathcal{B}$: Empirical Illustration

To study the employment and wage effect, run

$$\log(\text{emp}_{rt}) = \sum_{\tau=2011,\tau\neq 2014}^{2016} \gamma_\tau^{\text{e}} \overline{GAP}_r 1\left[\tau = t\right] + \alpha_r^{\text{e}} + \xi_t^{\text{e}} + \varepsilon_{rt}^{\text{e}}$$

$$\log(\text{wage}_{rt}) = \sum_{\tau=2011,\tau\neq 2014}^{2016} \gamma_\tau^{\text{w}} \overline{GAP}_r 1\left[\tau = t\right] + \alpha_r^{\text{w}} + \xi_t^{\text{w}} + \varepsilon_{rt}^{\text{w}}$$

- $\log(\text{emp}_{rt})$ is the log employment in district r time t; $\log(\text{wage}_{rt})$ is log wage
- $\overline{GAP}_r$ is a measure of the exposure to the minimum wage
- Pre-policy year 2011-2014

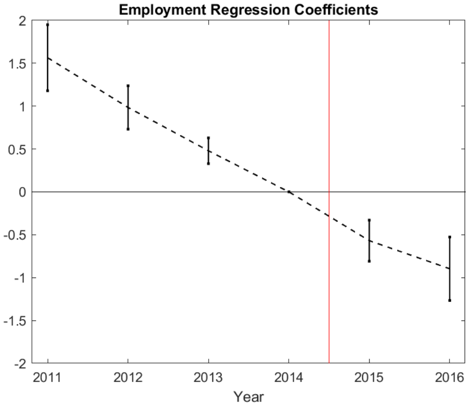
We are interested in the employment and wage effect at $t = 2015$ ▸ Back

Finite $\mathcal{B}$: Empirical Illustration

RR23 in Dustmann, Lindner, Schönberg, Umkehrer, Vom Berge (2022, QJE)

What are the effects of the minimum wage? ▸ Back

- Is employment effect ≥ -0.6 (wage effect) without parallel trends?



Finite $\mathcal{B}$: Empirical Illustration

RR23 in Dustmann, Lindner, Schönberg, Umkehrer, Vom Berge (2022, QJE)

What are the effects of the minimum wage? Back

- Is employment effect ≥ -0.6 (wage effect) without parallel trends?

Finite $\mathcal{B}$: Empirical Illustration

RR23 in Dustmann, Lindner, Schönberg, Umkehrer, Vom Berge (2022, QJE)

What are the effects of the minimum wage? Back

- Is employment effect ≥ -0.6 (wage effect) without parallel trends?

Relax the parallel trends assumption: $|\Delta_{2015} - \Delta_{2014}| \leqslant M \times \max_{t=2013,2014} |\Delta_t - \Delta_{t-1}|$

Simulation - Setting

I set $n = 5000$

Each sample $\{W_i\}_{i=1}^n$ and estimator is generated by

$$\begin{pmatrix} W_{\Delta,i} \\ W_{\gamma,i} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \Delta \\ \gamma \end{pmatrix}, n\Omega\right)$$

The estimator is calculated by

$$\begin{pmatrix} \widehat{\Delta} \\ \widehat{\gamma} \end{pmatrix} = \begin{pmatrix} \frac{1}{n}\sum W_{\Delta,i} \\ \frac{1}{n}\sum W_{\gamma,i} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \Delta \\ \gamma \end{pmatrix}, \Omega\right)$$

I conduct inference using pair $\left(\widehat{\Delta}, \widehat{\gamma}, \Omega\right)$

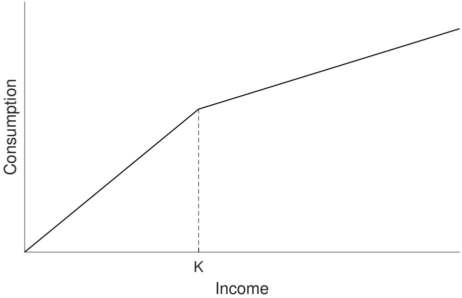
Back

Bunching and Taxable Income Elasticity

Blomquist, Newey, Kumar and Liang (2021, JPE) Intro Conclusion

Bunching and Taxable Income Elasticity

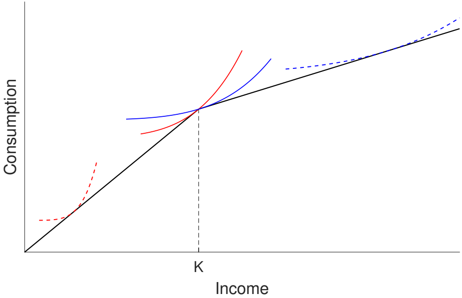
Blomquist, Newey, Kumar and Liang (2021, JPE) ▸ Intro ▸ Conclusion



The consumption has two linear segments with slopes $\rho_1 > \rho_2$ and a kink at K

Bunching and Taxable Income Elasticity

Blomquist, Newey, Kumar and Liang (2021, JPE) ▸ Intro ▸ Conclusion



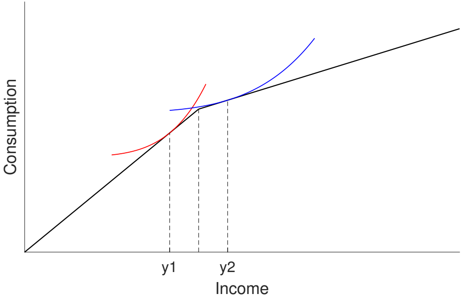
Assume isoelastic utility function $U(c, y, \eta) = c - \frac{\eta}{1+1/\theta}\left(\frac{y}{\eta}\right)^{1+1/\theta}$

- $\eta > 0$: unobservable heterogeneity
- $\theta > 0$: income elasticity
- $\phi(\eta)$: be the density of η

No restriction on $\phi(\eta) \Rightarrow \theta$ is not identified

Bunching and Taxable Income Elasticity

Blomquist, Newey, Kumar and Liang (2021, JPE) ▸ Intro ▸ Conclusion



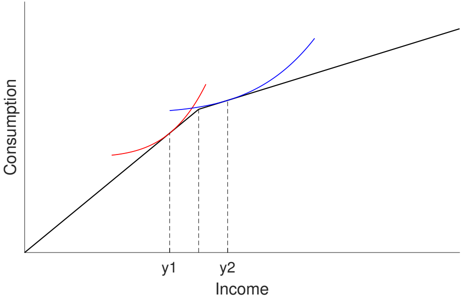
Assume there are $\sigma_2 \geq 1 \geq \sigma_1 > 0$ such that

$$\sigma_1 \min\left\{\phi\left(\eta_1\right), \phi\left(\eta_2\right)\right\} \leq \phi(\eta) \leq \sigma_2 \max\left\{\phi\left(\eta_1\right), \phi\left(\eta_2\right)\right\}$$

for $\eta \in [\eta_1, \eta_2]$

Bunching and Taxable Income Elasticity

Blomquist, Newey, Kumar and Liang (2021, JPE) ▸ Intro ▸ Conclusion



The identified set is $\theta \in \left[\min_{b=1,2} \lambda_{l,b},\ \max_{\beta=1,2} \lambda_{u,b}\right]$ where

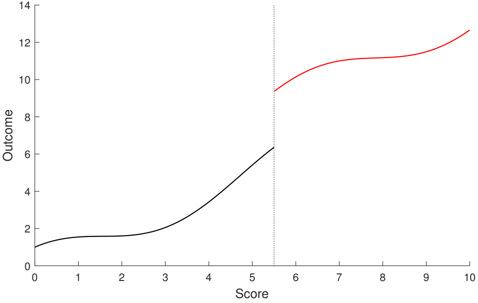
$$\lambda_{l,1} = \frac{\log\left(\frac{y_1}{y_2} + \frac{P(y_1 \le Y \le y_2)}{f^-(y_1)\sigma_2 y_2}\right)}{\log \rho_1 - \log \rho_2},\ \lambda_{l,2} = \frac{-\log\left(\frac{y_2}{y_1} - \frac{P(y_1 \le Y \le y_2)}{f^-(y_2)\sigma_2 y_1}\right)}{\log \rho_1 - \log \rho_2}$$

Regression Discontinuity Design

Kolesár and Rothe (2018, AER) DiD Intro Conclusion

Regression Discontinuity Design

Kolesár and Rothe (2018, AER) DiD Intro Conclusion



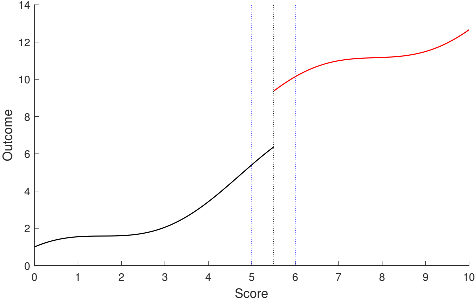
Treatment $D = 1\{X \geq k\}$ with running variable X

Let $\mu(X) = E[Y \mid X]$. The ATE at the threshold is

$$\theta = E\Big[Y(1) - Y(0) \mid X = 0\Big] = \lim_{x \downarrow k} \mu(x) - \lim_{x \uparrow k} \mu(x)$$

Regression Discontinuity Design

Kolesár and Rothe (2018, AER) ▸ DiD ▸ Intro ▸ Conclusion

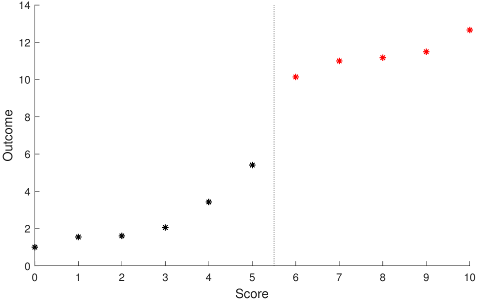


Estimate θ: local OLS of Y on $m(X)$ with $X \in [k-h, k+h]$ where

$$m(x) = (1\{x \geq k\}, 1\{x \geq k\}(x-k), \ldots, 1\{x \geq k\}(x-k)^p, \\ 1, x-k, \ldots, (x-k)^p)$$

Regression Discontinuity Design

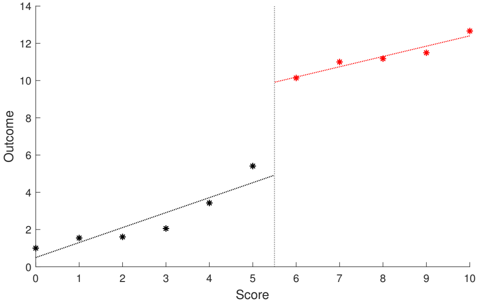
Kolesár and Rothe (2018, AER) DiD Intro Conclusion



Regression Discontinuity Design

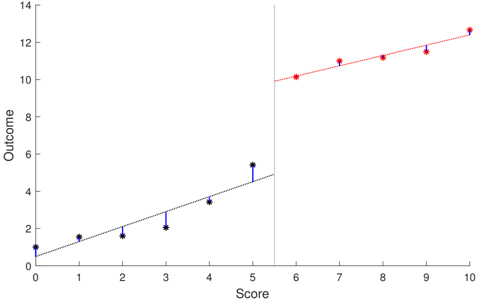
Kolesár and Rothe (2018, AER)

DiD Intro Conclusion



Regression Discontinuity Design

Kolesár and Rothe (2018, AER) ▸ DiD ▸ Intro ▸ Conclusion



Assumption: bounds on specification errors at the threshold

$$|\lim_{x\uparrow k}\Delta(x)| \leq \max_{x'<k}|\Delta(x')|, \; |\lim_{x\downarrow k}\Delta(x)| \leq \max_{x'>k}|\Delta(x')|$$

Size & Power: Comparison with Simple CI

Theorem (Symmetric Or Large Bounds)

Suppose Assumptions KS, AN, FR, CE hold, $\alpha \in (0, \frac{1}{2})$
(Symmetric Bounds) *If* $corr(\widehat{\lambda}_{\ell,b_1}, \widehat{\lambda}_{\ell,b_2}) < \rho_1^*(\alpha, \alpha^c)$, $corr(\widehat{\lambda}_{\ell,b_\ell}, \widehat{\lambda}_{\ell,b_u}) < \rho_2^*(\alpha)$

$$\widehat{\lambda}_\ell = \widehat{\lambda}_u$$

1. *My CI is Strictly Shorter:*
   There is $\alpha' > \alpha$ *such that*
   $$\liminf_n \inf_{P \in \mathcal{P}_n} P\left(CI^m\left(\hat{\lambda}_n, \hat{\Sigma}_n/n; \alpha\right) \subseteq CI^{sim}\left(\hat{\lambda}_n, \hat{\Sigma}_n/n; \alpha'\right)\right) = 1$$
2. *My CI has Higher Power:*
   For all $P_n \in \mathcal{P}_n$, *there is a subsequence* P_{τ_n} *and* $\kappa \in (0, +\infty)$ *such that*
   $$\liminf_n P_{\tau_n}\left(\theta_{\tau_n} \notin CI^m\left(\hat{\lambda}, \hat{\Sigma}/\tau_n; \alpha\right)\right) - P_{\tau_n}\left(\theta_{\tau_n} \notin CI^{sim}\left(\hat{\lambda}, \hat{\Sigma}/\tau_n; \alpha\right)\right) > 0$$
   for some $\theta_{\tau_n} = \theta_\ell - \frac{\kappa}{\sqrt{\tau_n}}$.

Back

Size & Power: Comparison with Simple CI

Theorem (Symmetric Or Large Bounds)

Suppose Assumptions KS, AN, FR, CE hold, $\alpha \in (0, \frac{1}{2})$
(Large Bounds) *Let* $\kappa_n = o(\sqrt{n})$ *and* $\kappa_n \to \infty$, *and*

$$\mathcal{P}_n = \left\{ P \in \mathcal{P} : \lambda_{u,b_u} - \lambda_{\ell,b_\ell} \geq \frac{\kappa_n}{\sqrt{n}} \right\}$$

1. *My CI is Strictly Shorter:*
   There is $\alpha' > \alpha$ *such that*

$$\liminf_n \inf_{P \in \mathcal{P}_n} P\left(CI^m\left(\hat{\lambda}_n, \hat{\Sigma}_n / n; \alpha\right) \subseteq CI^{sim}\left(\hat{\lambda}_n, \hat{\Sigma}_n / n; \alpha'\right)\right) = 1$$

2. *My CI has Higher Power:*
   For all $P_n \in \mathcal{P}_n$, *there is a subsequence* P_{τ_n} *and* $\kappa \in (0, +\infty)$ *such that*

$$\liminf_n P_{\tau_n}\left(\theta_{\tau_n} \notin CI^m\left(\hat{\lambda}, \hat{\Sigma}/\tau_n; \alpha\right)\right) - P_{\tau_n}\left(\theta_{\tau_n} \notin CI^{sim}\left(\hat{\lambda}, \hat{\Sigma}/\tau_n; \alpha\right)\right) > 0$$

for some $\theta_{\tau_n} = \theta_\ell - \frac{\kappa}{\sqrt{\tau_n}}$.

Back

Intersection & Union

Intersection bounds

$$H_0 : \max_{b \in \mathcal{B}} \lambda_{\ell,b} \leq \theta$$

An intuitive test statistic

$$\hat{T}(\theta) = \max_{b \in \mathcal{B}} \frac{\sqrt{n}(\hat{\lambda}_{\ell,b} - \theta)}{\sigma_{\ell,b}}$$

Intersection & Union

Intersection bounds

$$H_0 : \max_{b \in \mathcal{B}} \lambda_{\ell,b} \leq \theta$$

An intuitive test statistic

$$\hat{T}(\theta) = \max_{b \in \mathcal{B}} \frac{\sqrt{n}(\hat{\lambda}_{\ell,b} - \theta)}{\sigma_{\ell,b}} = \max_{b \in \mathcal{B}} \frac{\sqrt{n}(\hat{\lambda}_{\ell,b} - \lambda_{\ell,b})}{\sigma_{\ell,b}} + \frac{\sqrt{n}(\lambda_{\ell,b} - \theta)}{\sigma_{\ell,b}}$$

Intersection & Union

Intersection bounds

$$H_0 : \max_{b \in \mathcal{B}} \lambda_{\ell,b} \leq \theta$$

An intuitive test statistic

$$\hat{T}(\theta) = \max_{b \in \mathcal{B}} \frac{\sqrt{n}(\hat{\lambda}_{\ell,b} - \theta)}{\sigma_{\ell,b}} = \max_{b \in \mathcal{B}} \frac{\sqrt{n}(\hat{\lambda}_{\ell,b} - \lambda_{\ell,b})}{\sigma_{\ell,b}} + \frac{\sqrt{n}(\lambda_{\ell,b} - \theta)}{\sigma_{\ell,b}}$$

Key difficulty: $\sqrt{n}(\lambda_{\ell,b} - \theta)$ cannot be consistently estimated

Solution: $\sqrt{n}(\lambda_{\ell,b} - \theta) \leq 0 \Rightarrow$ get an upper bound $\sqrt{n}(\lambda_{\ell,b} - \theta)/\sqrt{\ln n}$

Intersection & Union

Intersection bounds

$$H_0 : \max_{b \in \mathcal{B}} \lambda_{\ell,b} \leq \theta$$

An intuitive test statistic

$$\hat{T}(\theta) = \max_{b \in \mathcal{B}} \frac{\sqrt{n}(\hat{\lambda}_{\ell,b} - \theta)}{\sigma_{\ell,b}} = \max_{b \in \mathcal{B}} \frac{\sqrt{n}(\hat{\lambda}_{\ell,b} - \lambda_{\ell,b})}{\sigma_{\ell,b}} + \frac{\sqrt{n}(\lambda_{\ell,b} - \theta)}{\sigma_{\ell,b}}$$

Key difficulty: $\sqrt{n}(\lambda_{\ell,b} - \theta)$ cannot be consistently estimated

Solution: $\sqrt{n}(\lambda_{\ell,b} - \theta) \leq 0 \Rightarrow$ get an upper bound $\sqrt{n}(\lambda_{\ell,b} - \theta)/\sqrt{\ln n}$

Union bounds

$$H_0 : \min_{b \in \mathcal{B}} \lambda_{\ell,b} \leq \theta$$

An intuitive test statistic

$$\hat{T}(\theta) = \min_{b \in \mathcal{B}} \frac{\sqrt{n}(\hat{\lambda}_{\ell,b} - \theta)}{\sigma_{\ell,b}} = \min_{b \in \mathcal{B}} \frac{\sqrt{n}(\hat{\lambda}_{\ell,b} - \lambda_{\ell,b})}{\sigma_{\ell,b}} + \frac{\sqrt{n}(\lambda_{\ell,b} - \theta)}{\sigma_{\ell,b}}$$

Intersection & Union

Intersection bounds

$$H_0 : \max_{b \in \mathcal{B}} \lambda_{\ell,b} \leq \theta$$

An intuitive test statistic

$$\hat{T}(\theta) = \max_{b \in \mathcal{B}} \frac{\sqrt{n}(\hat{\lambda}_{\ell,b} - \theta)}{\sigma_{\ell,b}} = \max_{b \in \mathcal{B}} \frac{\sqrt{n}(\hat{\lambda}_{\ell,b} - \lambda_{\ell,b})}{\sigma_{\ell,b}} + \frac{\sqrt{n}(\lambda_{\ell,b} - \theta)}{\sigma_{\ell,b}}$$

Key difficulty: $\sqrt{n}(\lambda_{\ell,b} - \theta)$ cannot be consistently estimated

Solution: $\sqrt{n}(\lambda_{\ell,b} - \theta) \leq 0 \Rightarrow$ get an upper bound $\sqrt{n}(\lambda_{\ell,b} - \theta)/\sqrt{\ln n}$

Union bounds

$$H_0 : \min_{b \in \mathcal{B}} \lambda_{\ell,b} \leq \theta$$

An intuitive test statistic

$$\hat{T}(\theta) = \min_{b \in \mathcal{B}} \frac{\sqrt{n}(\hat{\lambda}_{\ell,b} - \theta)}{\sigma_{\ell,b}} = \min_{b \in \mathcal{B}} \frac{\sqrt{n}(\hat{\lambda}_{\ell,b} - \lambda_{\ell,b})}{\sigma_{\ell,b}} + \frac{\sqrt{n}(\lambda_{\ell,b} - \theta)}{\sigma_{\ell,b}}$$

Key difficulty: $\sqrt{n}(\lambda_{\ell,b} - \theta)$ cannot be consistently estimated & unknown sign

Back

Step 2: Conditional CV

Lemma

$$\frac{\Phi\left(\hat{T}(\theta)\right) - \Phi\left(t_{\ell,1}(\theta, b_\ell)\right)}{\Phi\left(t_{\ell,2}(\theta, b_\ell)\right) - \Phi\left(t_{\ell,1}(\theta, b_\ell)\right)} \,\Big|\, \left\{\hat{T}(\theta) = \mathcal{Z}_{\ell, b_\ell}\right\} \overset{\text{FOSD}}{\preceq} \text{Unif}(0,1)$$

A Simple Example - Step 3 Conditional CV

I illustrate with ▸ Back

$$\hat{T}(\theta) \,\Big|\, \hat{T}(\theta) = \hat{\lambda}_1 - \theta, \hat{s} = s$$

A Simple Example - Step 3 Conditional CV

I illustrate with ▸ Back

$$\hat{T}(\theta) \,\Big|\, \hat{T}(\theta) = \hat{\lambda}_1 - \theta, \hat{s} = s$$
$$\sim \hat{\lambda}_1 - \theta \,\Big|\, \hat{T}(\theta) = \hat{\lambda}_1 - \theta, \hat{s} = s$$

A Simple Example - Step 3 Conditional CV

I illustrate with ▸ Back

$$
\begin{aligned}
&\hat{T}(\theta) \,\big|\, \hat{T}(\theta) = \hat{\lambda}_1 - \theta, \hat{s} = s \\
\sim &\hat{\lambda}_1 - \theta \,\big|\, \hat{T}(\theta) = \hat{\lambda}_1 - \theta, \hat{s} = s \\
\sim &\hat{\lambda}_1 - \theta \,\big|\, \theta - \hat{\lambda}_2 \leq \hat{\lambda}_1 - \theta \leq \hat{\lambda}_2 - \theta, \hat{s} = s
\end{aligned}
$$

A Simple Example - Step 3 Conditional CV

I illustrate with ▸ Back

$$\begin{aligned} &\hat{T}(\theta) \,\Big|\, \hat{T}(\theta) = \hat{\lambda}_1 - \theta, \hat{s} = s \\ \sim &\hat{\lambda}_1 - \theta \,\Big|\, \hat{T}(\theta) = \hat{\lambda}_1 - \theta, \hat{s} = s \\ \sim &\hat{\lambda}_1 - \theta \,\Big|\theta - \hat{\lambda}_2 \leq \hat{\lambda}_1 - \theta \leq \hat{\lambda}_2 - \theta, \hat{s} = s \\ \sim &\hat{\lambda}_1 - \theta \,\Big|\theta - \hat{\lambda}_2 \leq \hat{\lambda}_1 - \theta \leq \hat{\lambda}_2 - \theta, \text{ and } s = 1 \end{aligned}$$

A Simple Example - Step 3 Conditional CV

I illustrate with ▸ Back

$$
\begin{aligned}
&\hat{T}(\theta) \,\big|\, \hat{T}(\theta) = \hat{\lambda}_1 - \theta, \hat{s} = s \\
\sim &\hat{\lambda}_1 - \theta \,\big|\, \hat{T}(\theta) = \hat{\lambda}_1 - \theta, \hat{s} = s \\
\sim &\hat{\lambda}_1 - \theta \,\big|\theta - \hat{\lambda}_2 \leq \hat{\lambda}_1 - \theta \leq \hat{\lambda}_2 - \theta, \hat{s} = s \\
\sim &\hat{\lambda}_1 - \theta \,\big|\theta - \hat{\lambda}_2 \leq \hat{\lambda}_1 - \theta \leq \hat{\lambda}_2 - \theta, \text{ and } s = 1 \\
\sim &\mathcal{TN}\left(\lambda_1 - \theta, \left[\theta - \hat{\lambda}_2, \hat{\lambda}_2 - \theta\right]\right)
\end{aligned}
$$

A Simple Example - Step 3 Conditional CV

I illustrate with ▸ Back

$$
\begin{aligned}
& \hat{T}(\theta) \,\middle|\, \hat{T}(\theta) = \hat{\lambda}_1 - \theta, \hat{s} = s \\
\sim & \hat{\lambda}_1 - \theta \,\middle|\, \hat{T}(\theta) = \hat{\lambda}_1 - \theta, \hat{s} = s \\
\sim & \hat{\lambda}_1 - \theta \,\middle|\, \theta - \hat{\lambda}_2 \le \hat{\lambda}_1 - \theta \le \hat{\lambda}_2 - \theta, \hat{s} = s \\
\sim & \hat{\lambda}_1 - \theta \,\middle|\, \theta - \hat{\lambda}_2 \le \hat{\lambda}_1 - \theta \le \hat{\lambda}_2 - \theta, \text{ and } s = 1 \\
\sim & \mathcal{TN}\left(\lambda_1 - \theta, \left[\theta - \hat{\lambda}_2, \hat{\lambda}_2 - \theta\right]\right) \\
\preceq & \mathcal{TN}\left(0, \left[\theta - \hat{\lambda}_2, \hat{\lambda}_2 - \theta\right]\right)
\end{aligned}
$$

A Simple Example - Step 3 Conditional CV

Lemma

$$\frac{\Phi\left(\hat{T}(\theta)\right) - \Phi\left(t_{\ell,1}(\theta, b_\ell)\right)}{\Phi\left(t_{\ell,2}(\theta, b_\ell)\right) - \Phi\left(t_{\ell,1}(\theta, b_\ell)\right)} \Big| \left\{ \hat{T}(\theta) = \mathcal{Z}_{\ell,b_\ell} \right\} \overset{\text{FOSD}}{\preceq} \text{Unif}(0,1)$$

$$\frac{\Phi\left(\hat{T}(\theta)\right) - \Phi\left(t_{u,1}(\theta, b_u)\right)}{\Phi\left(t_{u,2}(\theta, b_u)\right) - \Phi\left(t_{u,1}(\theta, b_u)\right)} \Big| \left\{ \hat{T}(\theta) = \mathcal{Z}_{u,b_u} \right\} \overset{\text{FOSD}}{\preceq} \text{Unif}(0,1)$$

where

$$t_{\ell,1}(\theta, b) = \begin{cases} \min_{\tilde{b}\in\mathcal{B}} \left(1 + \rho_{\ell u}(b, \tilde{b})\right)^{-1} \left(\mathcal{Z}_{u,\tilde{b}} + \rho_{\ell u}(b, \tilde{b})\mathcal{Z}_{\ell,b}\right), & \text{if } \min_{\tilde{b}\in\mathcal{B}} \rho_{\ell u}(b, \tilde{b}) > -1 \\ -\infty & \text{otherwise} \end{cases}$$

$$t_{u,1}(\theta, b) = \begin{cases} \min_{\tilde{b}\in\mathcal{B}} \left(1 + \rho_{\ell u}(\tilde{b}, b)\right)^{-1} \left(\mathcal{Z}_{\ell,\tilde{b}} + \rho_{\ell u}(\tilde{b}, b)\mathcal{Z}_{u,b}\right), & \text{if } \min_{\tilde{b}\in\mathcal{B}} \rho_{\ell u}(\tilde{b}, b) > -1 \\ -\infty & \text{otherwise} \end{cases}$$

$$t_{\ell,2}(\theta, b) = \begin{cases} \min_{\tilde{b}\in\mathcal{B}:\rho_\ell(b,\tilde{b})<1} \left(1 - \rho_{\ell}(b, \tilde{b})\right)^{-1} \left(\mathcal{Z}_{\ell,\tilde{b}} - \rho_{\ell}(b, \tilde{b})\mathcal{Z}_{\ell,b}\right) & \text{if } \min_{\tilde{b}\in\mathcal{B}} \rho_{\ell}(b, \tilde{b}) < 1 \\ +\infty & \text{otherwise} \end{cases}$$

$$t_{u,2}(\theta, b) = \begin{cases} \min_{\tilde{b}\in\mathcal{B}:\rho_u(\tilde{b},b)<1} \left(1 - \rho_{u}(\tilde{b}, b)\right)^{-1} \left(\mathcal{Z}_{u,\tilde{b}} - \rho_{u}(\tilde{b}, b)\mathcal{Z}_{u,b}\right) & \text{if } \min_{\tilde{b}\in\mathcal{B}} \rho_{u}(b, \tilde{b}) < 1 \\ +\infty & \text{otherwise} \end{cases}$$