(Visualizing) Plausible Dynamic Treatment Effect Paths

Simon Freyaldenhoven[1] Christian Hansen[2]

[1]Federal Reserve Bank of Philadelphia

[2]University of Chicago

The goal

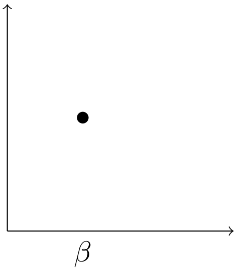
- We are interested in the dynamic treatment effect path of a policy.
- Examples include
  1. Dynamic treatment effects in microeconomics (distributed lag models),
  2. Impulse response functions in macroeconomics (local projections),
  3. Event study paths in finance (event studies).

Notation

- Let $\beta = \{\beta_h\}_{h=1}^{H}$ be the parameter of interest.
- β_h corresponds to treatment effect at horizon h.
- We have access to jointly normal estimates $\hat{\beta} = \{\hat{\beta}_h\}_{h=1}^{H}$.

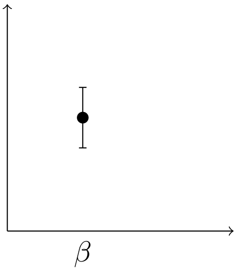
A primer on uncertainty quantification

Quantifying uncertainty



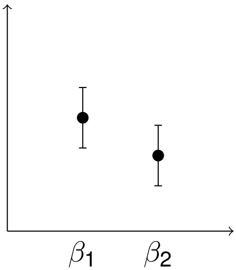
- Single parameter of interest β.
- Standard approach is to construct confidence interval.
- Coverage is $(1-\alpha)$: $\mathbb{P}(\ell(X) < \beta < u(X)) = 1-\alpha$.
- Intuitively: values inside CI appear “plausible”

Quantifying uncertainty



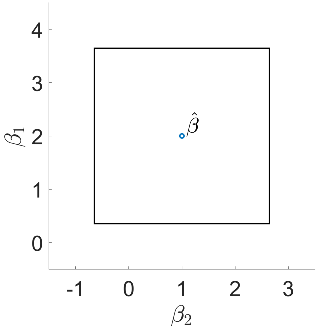
- Single parameter of interest β.
- Standard approach is to construct confidence interval.
- Coverage is $(1-\alpha)$: $\mathbb{P}(\ell(X) < \beta < u(X)) = 1-\alpha$.
- Intuitively: values inside CI appear "plausible"

Quantifying uncertainty



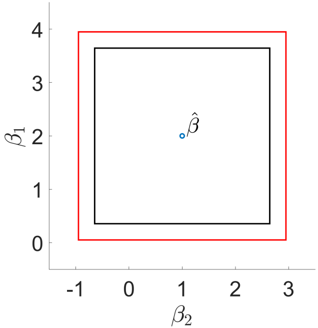
- Standard confidence intervals are pointwise valid.
- $(\ell(X), u(X)) : \mathbb{P}(\ell_h(X) < \beta_h < u_h(X)) = 1 - \alpha$.
- NOT uniformly valid.
- For example, with $Cov(\hat{\beta}_k, \hat{\beta}_l) = 0$:
  $\mathbb{P}(\ell_h(X) < \beta_h < u_h(X)\ \forall h) = (1 - \alpha)^H$.

In two dimensions



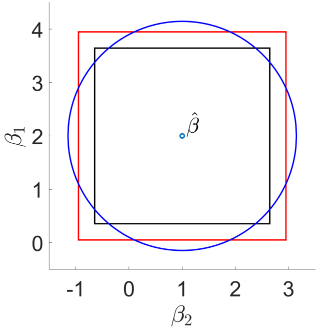
- Fix $\alpha = 0.1$, $Cov(\hat{\beta}_1, \hat{\beta}_2) = 0$.
- Pointwise CIs:
  $(\ell(X), u(X)) : \mathbb{P}(\ell_h(X) < \beta_h < u_h(X)) = 0.9$.
- $\mathbb{P}(\ell(X) < \beta < u(X)) = 0.9^2 = 0.81$.
- $\mathbb{P}(\beta \in CR^{pointwise}) = 0.81$.

In two dimensions



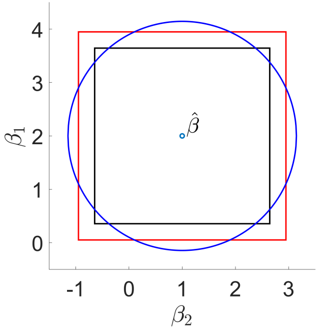
- Fix $\alpha = 0.1$, $Cov(\hat{\beta}_1, \hat{\beta}_2) = 0$.
- sup-t CIs: $(\ell(X), u(X)) : \mathbb{P}(\ell_h(X) < \beta_h < u_h(X)) \approx 0.949$.
- $\mathbb{P}(\ell(X) < \beta < u(X)) = 0.9$.
- $\mathbb{P}(\beta \in CR^{sup-t}) = 0.9$.

In two dimensions



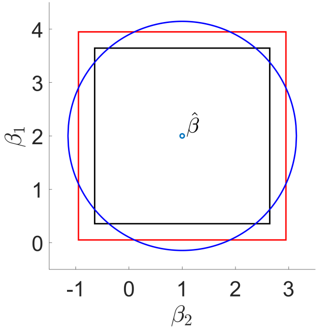
- Fix $\alpha = 0.1$, $Cov(\hat{\beta}_1, \hat{\beta}_2) = 0$.
- Wald confidence region: the set of β for which a joint F-test of the observed point estimates is not rejected.
- $\mathbb{P}(\beta \in CR^{Wald}) = 0.9$.

Comments



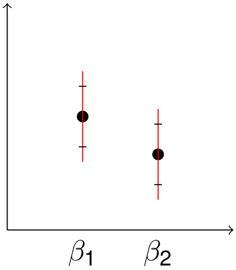
- Many other confidence regions exist.
- Both CR^{sup-t} and CR^{Wald} depend on off-diagonal entries in $Var(\beta)$. Example
- Power against different alternatives.
- CR^{Wald} infeasible to visualize in higher dimensions.

Comments



- Many other confidence regions exist.
- Both CR^{sup-t} and CR^{Wald} depend on off-diagonal entries in $Var(\beta)$. Example
- Power against different alternatives.
- CR^{Wald} infeasible to visualize in higher dimensions.

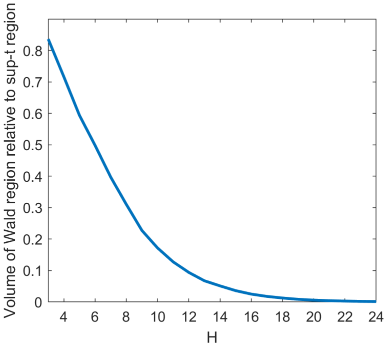
Quantifying uncertainty



- CR^{sup-t}: $(\ell_k(X), u_k(X))$.
- $(\ell_h(X), u_h(X)) : \beta_h \pm c^{sup-t}\sigma_{\beta_h}$.
- $\mathbb{P}(\beta \in CR^{sup-t}) = 1 - \alpha$.

cf. Freyberger and Rai (2018); Montiel Olea and Plagborg-Moller (2019); Callaway and Sant'Anna (2021); Jorda (2023); Boxell, Gentzkow, and Shapiro (2024); Mogstad, Romano, Shaikh, and Wilhelm (2024)

In higher dimensions



- Volume of CR^{sup-t} explodes relative to CR^{Wald}.
- Implication: vast majority of paths inside CR^{sup-t} rejected by a joint test.
- Suppose we uniformly draw paths from CR^{sup-t} for $\alpha = 0.05$.
  At $H = 24$, 99.9% of paths rejected by a joint test!

This paper

- We propose two types of plausible bounds that are
  a) feasible to add to a standard plot.
  b) (in general) narrower than existing confidence bands.

1. Restricting functional forms in data-driven way to ”reasonable shapes” (Restricted Plausible Bounds).

2. Relaxing uniformity requirement (Averaged Plausible Bounds).

This paper

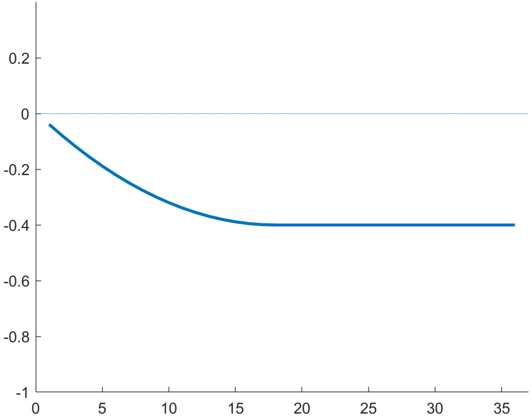
- We propose two types of plausible bounds that are
  a) feasible to add to a standard plot.
  b) (in general) narrower than existing confidence bands.

1. Restricting functional forms in data-driven way to ”reasonable shapes” (Restricted Plausible Bounds).

2. Relaxing uniformity requirement (Averaged Plausible Bounds).

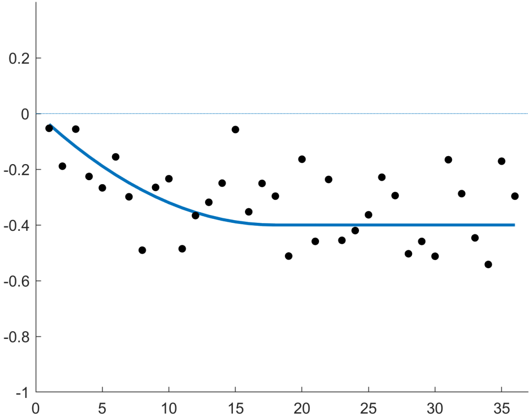
Restricted Plausible Bounds

An example



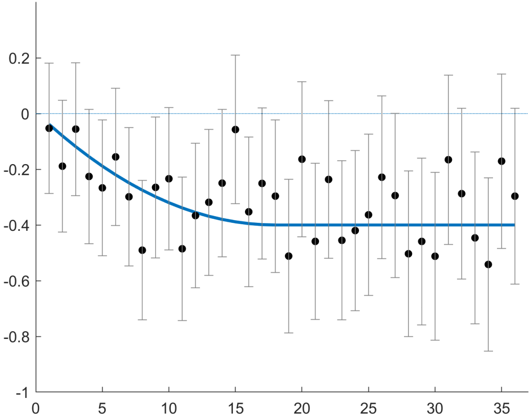
- Blue : true treatment effect β.

An example



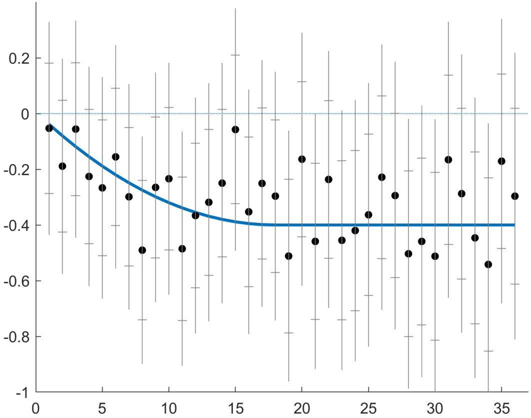
- Black dots: estimates $\hat{\beta}$.
- $\hat{\beta}_h \overset{i.i.d.}{\sim} N(\beta_h, \sigma_h)$.

An example



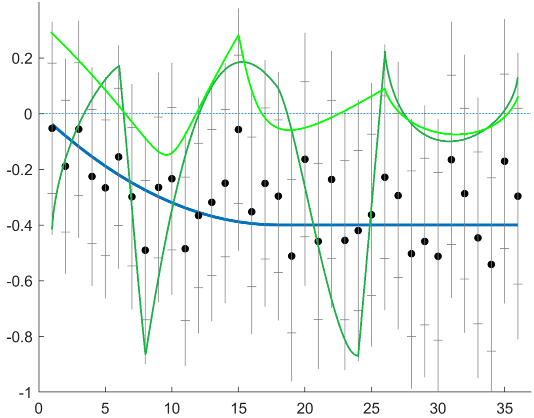
- Intervals: pointwise confidence intervals for β_h.

An example



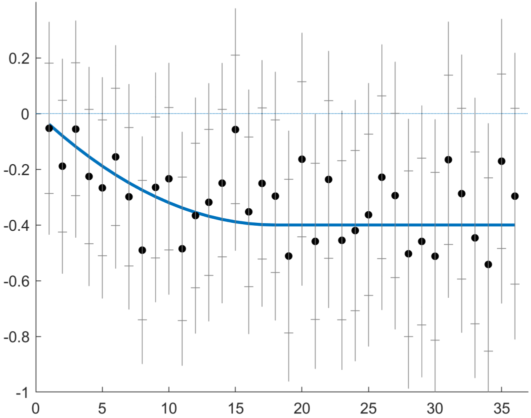
- Set of outer lines: sup-t confidence bands.

Uninformative data?



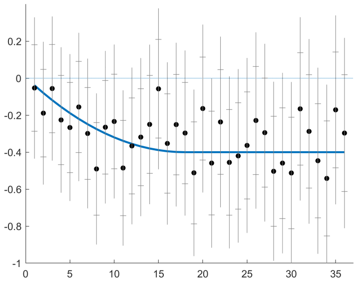
- Many paths with different overall effect or different shape appear “plausible”.

Uninformative data?



- Note: 1 out of 100,000 uniformly drawn paths in CR^{sup-t} not rejected by joint test.

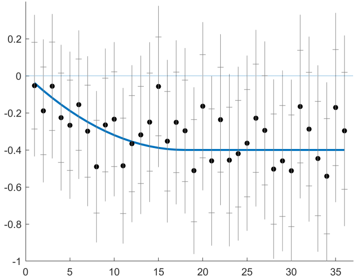
Uninformative data?



Restricting degrees of freedom:

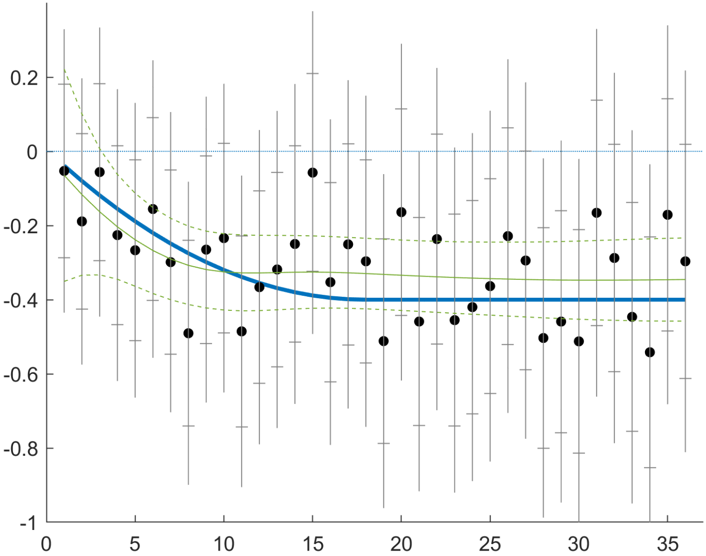
- Parametric model for β (e.g. effect grows linearly)
- time aggregation (e.g. monthly to quarterly). Restricts β to “step function”.
- estimate via VAR: Restricting to functional forms compatible with chosen VAR.
  - $AR(1) \Rightarrow \beta_h = \rho^h$.

Uninformative data?

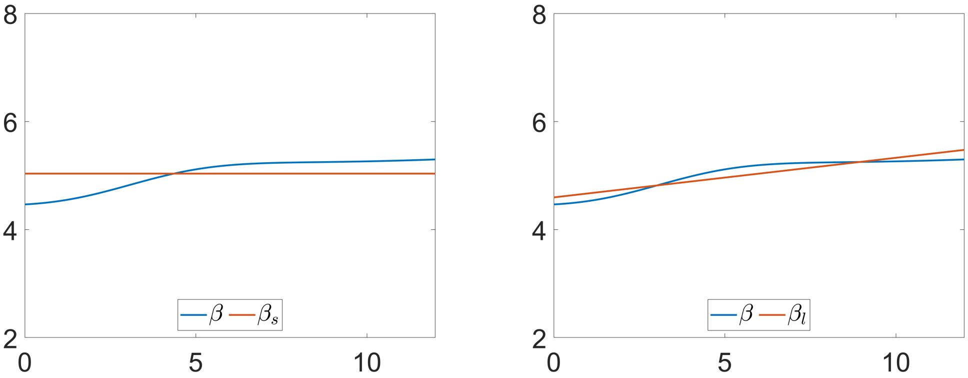


- Our proposal: transparent, data-driven restrictions on β.
- Economic intuition: smooth + eventually flat.

Restricted plausible bounds



A toy example (two surrogate models)



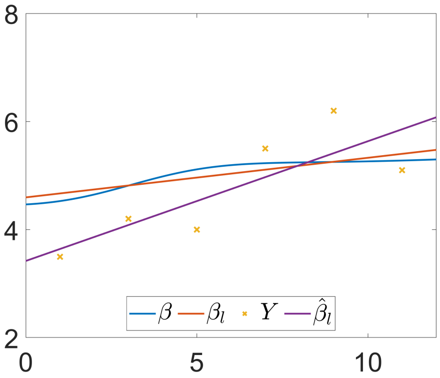
- $\beta_s = \arg\min_b \|\beta - b\|$ s.t. $\Delta b = 0$
- model with constant treatment effect
- $\beta_l = \arg\min_b \|\beta - b\|$ s.t. $\Delta^2 b = 0$
- model with linear treatment effect

A toy example

- If model M is fixed, inference for surrogate β_M is easy.
- E.g., model with linear treatment effect:

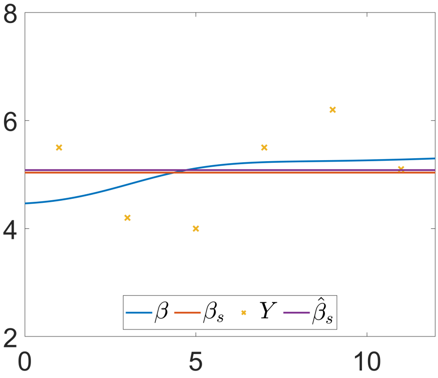
$$\hat{\beta}_l = \arg\min_b \|Y - b\| \text{ s.t. } \Delta^2 b = 0$$

- CIs for β_l follow (not β!).

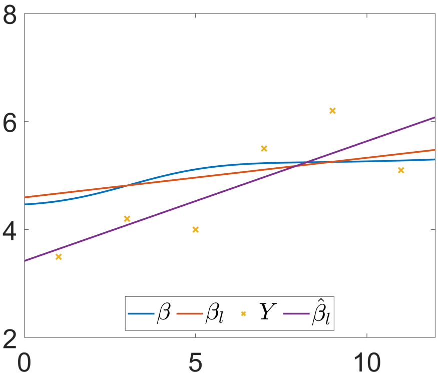


A toy example

If model $M(Y)$ depends on the data, this creates a problem.



$M = s$ if realization of Y_1 is large



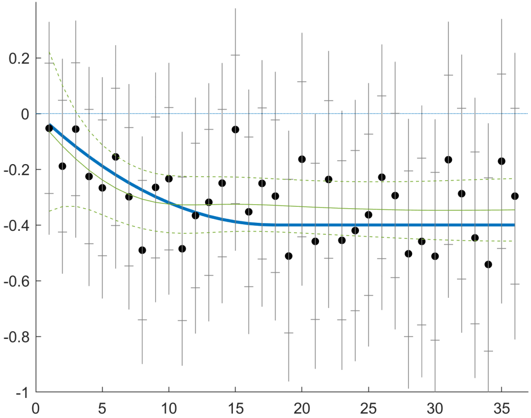
$M = l$ if realization of Y_1 is small

Restricted plausible bounds (toy example)

Our proposal: CR for $\beta_{M(Y)}$: M explicitly random, function of data

- Use data to select degrees of freedom/degree of smoothness (e.g. β_l or β_s).
- Take into account model uncertainty to construct uniformly valid CR for selected surrogate.

An example



- three degrees of freedom
- chosen using the data

Restricted plausible bounds

Let $\hat{\beta} \sim N(\beta, V_\beta)$, where β is a $H \times 1$ vector.

Define More details

$$\beta^* = \arg\min_b \underbrace{(\hat{\beta} - b)' V_\beta^{-1} (\hat{\beta} - b)}_{\text{distance from } \hat{\beta}} \tag{1}$$

$$\text{such that} \quad \underbrace{b' D_1' W_1(K) D_1 b < c_1}_{\substack{\text{small first difference,} \\ \text{after horizon K}}} \quad \text{and} \quad \underbrace{b' D_3' W_2 D_3 b < c_2}_{\text{small third difference}}.$$

- "treatment path is eventually flat."
- "treatment path is smooth."

Restricted plausible bounds

Let $\hat{\beta} \sim N(\beta, V_\beta)$, where β is a $H \times 1$ vector.

Define More details

$$\beta^* = \arg\min_b \underbrace{(\hat{\beta} - b)' V_\beta^{-1} (\hat{\beta} - b)}_{\text{distance from } \hat{\beta}}$$

$$\text{such that} \quad \underbrace{b' D_1' W_1(K) D_1 b < c_1}_{\substack{\text{small first difference,} \\ \text{after horizon K}}} \quad \text{and} \quad \underbrace{b' D_3' W_2 D_3 b < c_2}_{\text{small third difference}} . \tag{1}$$

- "treatment path is eventually flat."
- "treatment path is smooth."

Restricted plausible bounds

Equivalently,

$$\beta^* = \arg\min_b \; Q(b, \lambda_1, \lambda_2, K)$$

$$= \arg\min_b \; \underbrace{(\hat{\beta} - b)' V_\beta^{-1} (\hat{\beta} - b)}_{\text{distance from } \hat{\beta}} + \lambda_1 \underbrace{b' D_1' W_1(K) D_1 b}_{\substack{\text{penalty on first difference} \\ \text{after horizon K}}} + \lambda_2 \underbrace{b' D_3' W_2 D_3 b}_{\text{penalty on third difference}}$$

- "treatment path is eventually flat."
- "treatment path is smooth."

Restricted plausible bounds

$$\beta^* = \arg\min_b \; Q(b, \lambda_1, \lambda_2, K)$$
$$= \arg\min_b \; \underbrace{(\hat{\beta} - b)' V_{\beta}^{-1} (\hat{\beta} - b)}_{\text{distance from } \hat{\beta}} + \lambda_1 \underbrace{b' D_1' W_1(K) D_1 b}_{\substack{\text{penalty on first difference} \\ \text{after horizon K}}} + \lambda_2 \underbrace{b' D_3' W_2 D_3 b}_{\text{penalty on third difference}}$$

- Closed form solution for β^*:

$$\beta^* = \left(V_{\beta}^{-1} + \lambda_1 D_1' W_1(K) D_1 + \lambda_2 D_3' W_3 D_3 \right)^{-1} V_{\beta}^{-1} \hat{\beta}$$
$$= P\hat{\beta}$$

Restricted plausible bounds

For fixed λ_1, λ_2, K (fixed M):

- Distribution for β^*:

$$\beta^* - P\beta \sim N(0, PV_\beta P')$$

- Can construct confidence region for $\{P\beta_h\}_{h=1}^H$.
- What is $P\beta = \beta(\lambda_1, \lambda_2, K)$? "Surrogate of β" [Genovese and Wasserman, 2008] (cf. linear approximation)

Our proposal

Step 1: Use data to select Model $M(Y)$ (to select λ_1, λ_2, and K).

- Tie Researchers' hands by prespecifying $\mathcal{M}$, the universe of models considered
- Choice set includes:
  - constant treatment effect model ($df = 1$)
  - unrestricted estimates $\hat{\beta}$ ($df = H$)
- Select model (e.g. degrees of freedom $df(\lambda_1, \lambda_2, K)$) using information criterion.

More details

Our proposal

Step 2: Construct CR for $\beta_{M(Y)}$

- Take into account that model M is random, function of data
- Use Valid Post-Selection Inference (Berk et al. [2013]) to create CR^{POSI}.
  More details
- CR^{POSI} is a valid CR for selected surrogate.

Restricted plausible bounds

Proposition 1

For any treatment path β, we obtain valid coverage for its surrogate $\beta(\lambda_1, \lambda_2, K)$:

$$\mathbb{P}[\beta(\lambda_1, \lambda_2, K) \in CR^{POSI}] \geq 1 - \alpha. \tag{2}$$

Proof.

This follows immediately from the POSI guarantees:
$\mathbb{P}(\beta_M \in CR^{POSI} | M(Y) = M) \geq 1 - \alpha)$. □

Restricted plausible bounds

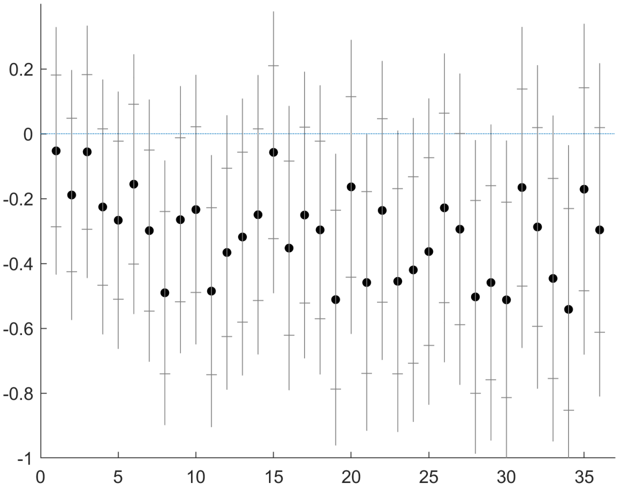
Proposition 2

Suppose $\mathbb{P}\left(\beta_{M(Y)} = \beta_M = \beta\right) = 1$. *Then,*

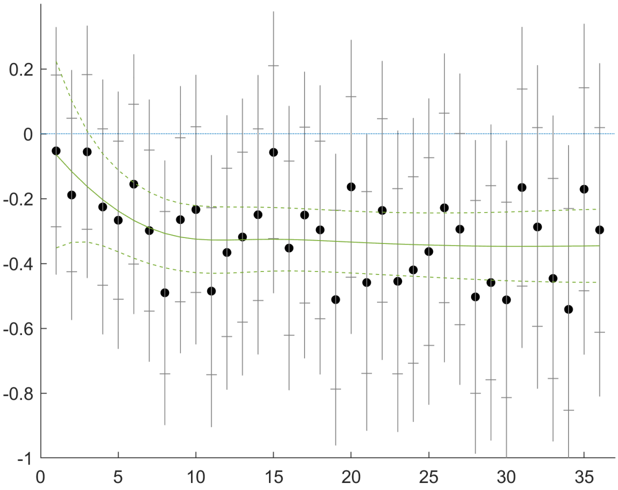
$$\mathbb{P}(\beta \in CR^{POSI}) \geq 1 - \alpha. \tag{3}$$

Note: Under some regularity conditions, $\lim_{V_\beta \to 0} \mathbb{P}\left(\beta_M(Y) = \beta_M = \beta\right) = 1$.

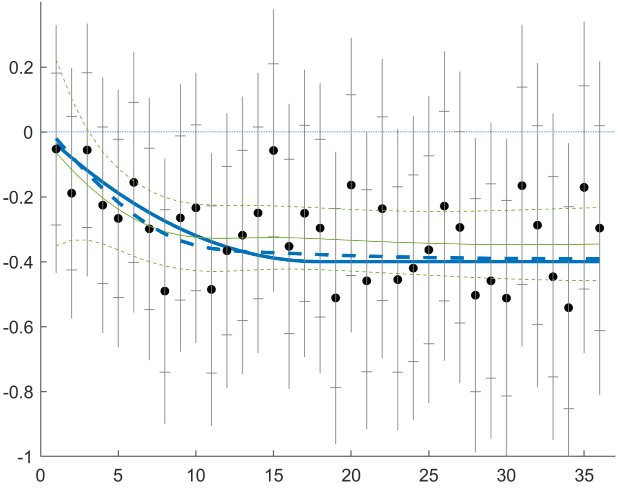
Example



Example



Example



Additional example

Averaged plausible bounds

Averaged plausible bounds

- Do not require any functional form restrictions.
- Relax uniform coverage to a weaker notion of "average coverage".
- True treatment paths will be on average within our bounds for $(1 - \alpha)$ of all realizations.

Averaged plausible bounds

- Correspond to largest and smallest total treatment effect that is consistent with the data.
- Boundary paths for testing the cumulative effect of the policy.
- Denoted by $\{\tilde{lb}_h, \tilde{ub}_h\}_{h=1}^{H}$.

Averaged plausible bounds

Proposition 3

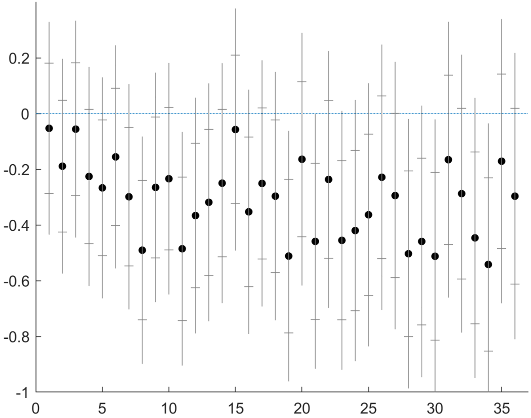
A Wald test on the cumulative treatment effect with significance level $(1-\alpha)$ *will reject*

a) *any treatment path* $\tilde{\beta}_h$ *with* $\sum_{h=1}^{H} \tilde{\beta}_h > \sum_{h=1}^{H} \tilde{ub}_h^{1-\alpha}$.

b) *any treatment path* $\tilde{\beta}_h$ *with* $\sum_{h=1}^{H} \tilde{\beta}_h < \sum_{h=1}^{H} \tilde{lb}_h^{1-\alpha}$.

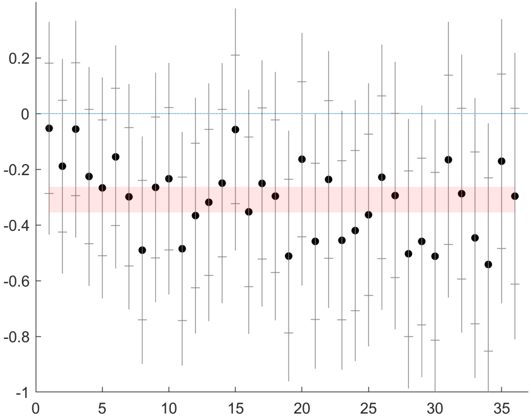
- Any path that is not, on average, inside our averaged plausible bounds, implies a cumulative treatment effect that is rejected by the corresponding hypothesis test. More details

Estimates too noisy?



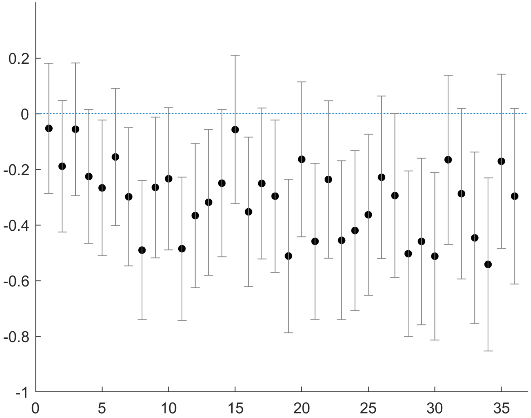
- Estimates can appear uninformative based on uniform confidence bands.

Averaged plausible bounds



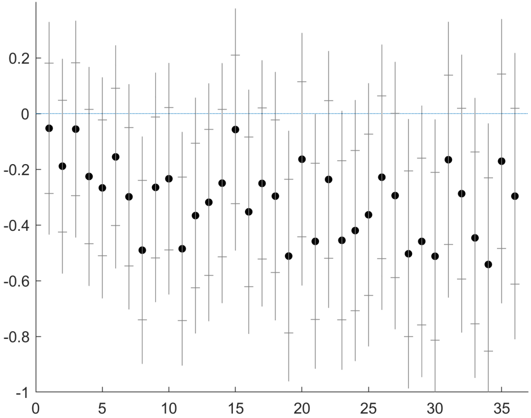
- Averaged plausible bounds: shaded red area
- Tight bounds on overall treatment effect

Summary: traditional figure



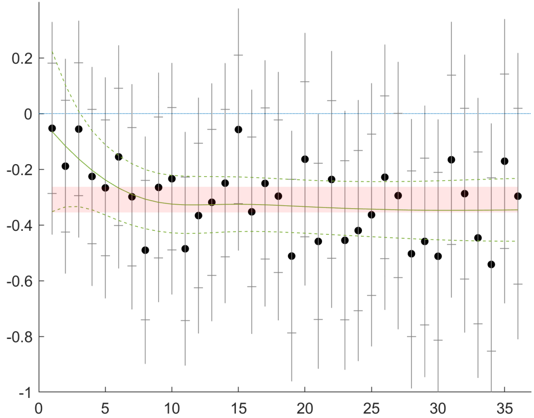
- Includes pointwise valid CIs

Summary: modern figure



- Includes uniform CR

Summary: our figure

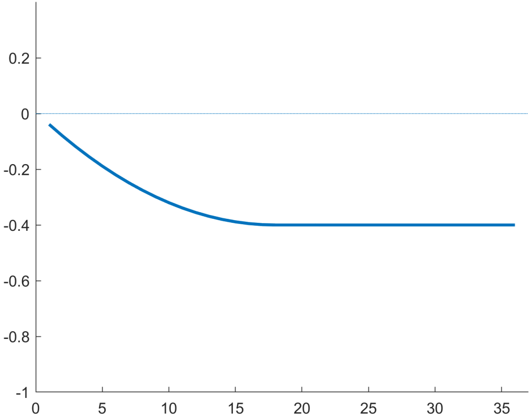


Includes three additional objects:

- Shaded red area: averaged plausible bounds:
- Green lines:
  1. Restricted plausible bounds (dashed)
  2. Restricted estimates (solid)

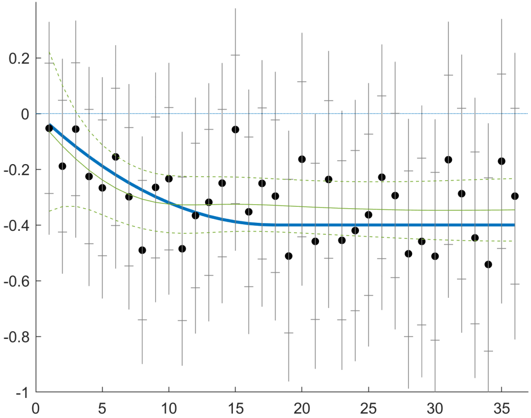
Numerical Illustrations

The setup



- Blue : true treatment effect β.
- Repeated simulations of $\hat{\beta} \sim N(\beta, V_\beta)$

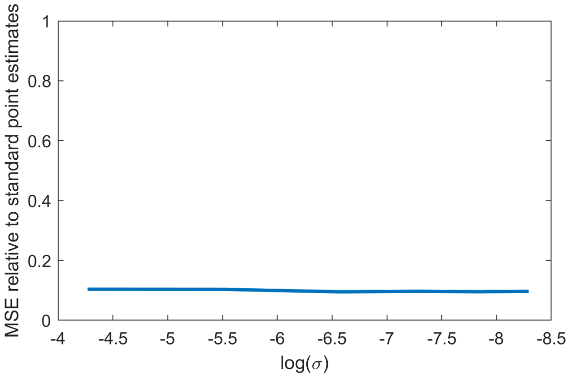
Estimation (restricted estimator $\hat{\beta}_{M(Y)}$)



Compare

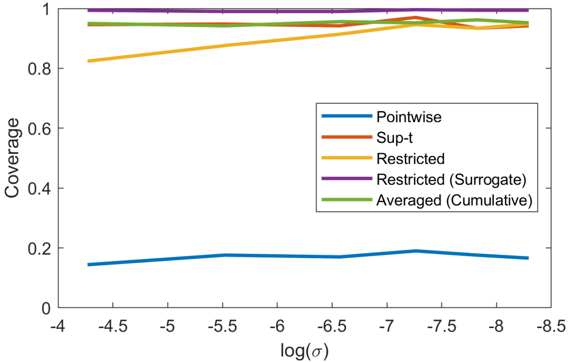
- Unrestricted estimates $\hat{\beta}$
- Restricted estimates $\hat{\beta}_{M(Y)}$

Estimation (restricted estimator $\hat{\beta}_{M(Y)}$)



- Depicted: $MSE_{\hat{\beta}_{M(Y)}} / MSE_{\hat{\beta}}$
- Good point estimation properties
- Large improvement in MSE (cf. James-Stein estimator)

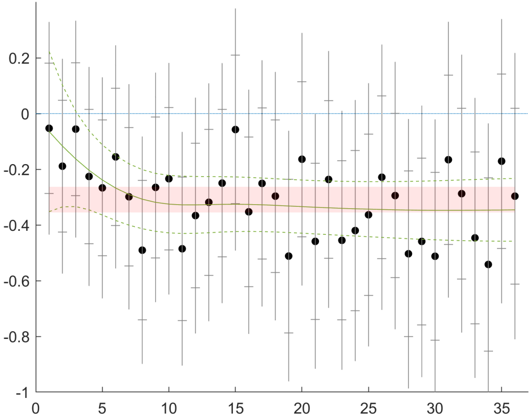
Inference



Restricted Bounds (CR^{POSI}):

- Coverage of β approaches 95% as $n \to \infty$.
- Coverage guarantee for β_M.
- Much better than pointwise (Even though tighter!).

Inference



- Restricted much narrower than pointwise
- Much better coverage

Conclusion

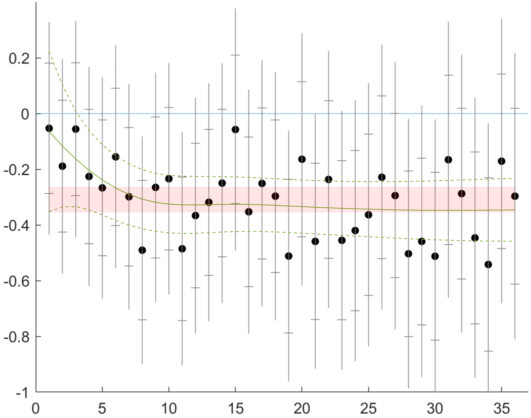
- We are interested in (joint inference on) the dynamic treatment effect of a policy.
- We propose two novel types of bounds to include in standard visualizations.
  - Both bounds can be substantially tighter than standard confidence bands.
  - Can provide useful insights when traditional confidence bands appear uninformative.
- Improved point estimation through data-driven smoothing

Next steps

- Write the paper.
- Your thoughts:
  a) Things you liked?
  b) Things you didn't like?
  c) Additional things you'd like to see/discussed?
  d) Thoughts on naming/framing?
- Papers that come to mind we could replicate?

Thank you!

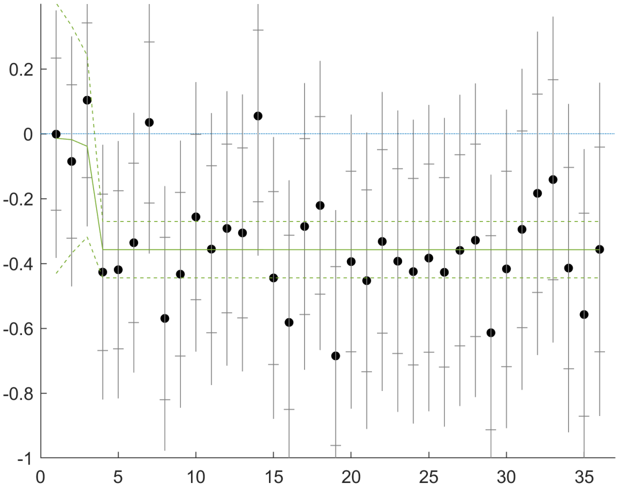
Summary: our figure



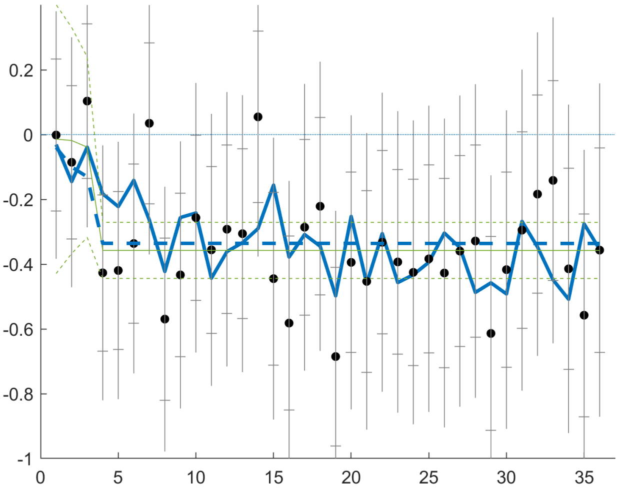
Includes three additional objects:

- Shaded red area: averaged plausible bounds:
- Green lines:
  1. Restricted plausible bounds (dashed)
  2. Restricted estimates (solid)

Example 2



Example 2



back

Averaged plausible bounds

Let

$$ub^{1-\alpha} = \max \sum_{h=1}^{H} \beta_h^* \qquad \text{s.t. } (\beta^* - \hat{\beta})' \Sigma_{\hat{\beta}} (\beta^* - \hat{\beta}) = c^{(1-\alpha)}, \tag{4}$$

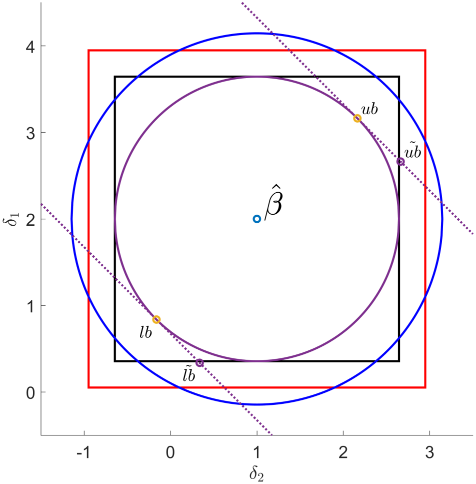
where $c^{(1-\alpha)}$ denotes denotes the inverse of the chi-square cdf with 1 degree of freedom at chosen significance level $(1-\alpha)$.

- Define $lb^{1-\alpha}$ analogously, replacing the $\max$ in (4) with $\min$.
- Closed form solution.

Averaged plausible bounds

- Note: Only scalar corresponding to upper and lower limit of cumulative effect is identified
- Many other ways to visualize. E.g.:
  - shift point estimates ($\dot{ub}, \dot{lb}$)
  - shift moving average ($\breve{ub}, \breve{lb}$)
  - shift constant effects estimate ($\tilde{ub}, \tilde{lb}$)

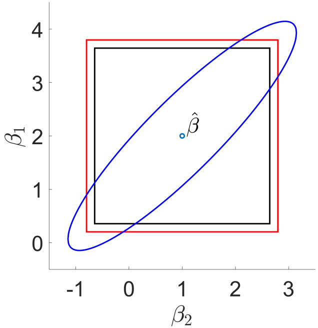
In two dimensions



- purple dotted lines give bounds on cumulative effect
- paths inside these bounds corresponds to paths inside

back

Comments



- Suppose $Cov(\hat{\beta}_1, \hat{\beta}_2) = 0.9$.
- Both CR^{sup-t} and CR^{Wald} depend on off-diagonal entries in $Var(\beta)$. back

Notation

Let $\hat{\beta} \sim N(\beta, V_\beta)$, where β is a $H \times 1$ vector.

- D_1 is first difference maker, D_3 is third difference maker
- $V_1 = D_1 V_\beta D_1'$, $V_3 = D_3 V_\beta D_3'$ are variance matrices for first and third differences
- $W_1(K) = diag(zeros(K), V_1(K+1:end, K+1:end)/mean(diag(V_1(K+1:end, K+1:end))))$
- $W_3 = diag(V_3)/mean(diag(V_3))$

back

Model selection

- Residuals $\hat{\beta} - \beta^* = \hat{\beta} - P(\lambda_1, \lambda_2, K)\hat{\beta} = (I - P(\lambda_1, \lambda_2, K))\hat{\beta}$
- Analog of residual degrees of freedom: trace$(I - P(\lambda_1, \lambda_2, K))$
- Analog of model degrees of freedom: df$(\lambda_1, \lambda_2, K)$ = trace$(P(\lambda_1, \lambda_2, K))$
- Choose β^* that minimizes BIC analog (over pre-specified grid) : $Q(\beta^*, \lambda_1, \lambda_2, K) + \log(H)\text{df}(\lambda_1, \lambda_2, K)$

back

Post-selection Inference

Consider CIs of the form $\{\ell_h(X), u_h(X)\}_{h=1}^{H} = \hat{\beta}_M \pm C^\alpha \sigma_\beta$.

For $\alpha = 0.05$:

- pointwise CIs: $\hat{\beta}_M = \hat{\beta}$, $C^\alpha = 1.96$
- sup-t CIs: $\hat{\beta}_M = \hat{\beta}$, C^α "sup-t constant"
- POSI CIs:$\hat{\beta}_M = \hat{\beta}_{M(Y)}$, C^α "POSI constant"

Post-selection Inference

POSI CIs: $\{\ell_h(X), u_h(X)\}_{h=1}^{H} = \hat{\beta}_{M(Y)} \pm C^{\alpha}\sigma_{\beta}$.

For $\alpha = 0.05$, let C^{α} be the minimal value that satisfies

$$\mathbb{P}\left(\max_{M \in \mathcal{M}} \max_{h} |t_{h \cdot M}| \leq C^{\alpha}\right) \geq 0.95$$

- $t_{j \cdot M}$ denotes the t-ratio for the proxy $\beta_{M(Y)}$ at horizon h
- $C^{0.05}$ depends on $\mathcal{M}$, the universe of models considered
- $C^{0.05}$ does not depend on the model selection procedure

back