Improved inference for nonparametric regression and regression-discontinuity designs

Giuseppe Cavaliere[1,2], Sílvia Gonçalves[3], Morten Ørregaard Nielsen[4], Edoardo Zanelli[1]

[1]University of Bologna, [2]Exeter Business School, [3]McGill University, [4]ACE, Aarhus University

Venice Time Series Workshop
April 22, 2026

Introduction

- Given a random sample $\{(y_i, x_i) : i = 1, \ldots, n\}$, the parameter of interest is

$$g(\mathsf{x}) := E(y_i | x_i = \mathsf{x}),$$

  where x is a fixed evaluation point.
- We estimate $g(\mathsf{x})$ using nonparametric (local linear) kernel regression.
- Main goal: to propose a method of inference on $g(\mathsf{x})$ that
  - is easy to implement,
  - works well in finite samples,
  - is valid whether x is in the interior or on the boundary of its domain.
- An important application of nonparametric regression at the boundary is regression-discontinuity design (RDD).

Main challenges

- Although $\hat{g}(\mathsf{x})$ is consistent, the test statistic is biased:

$$T_n := \sqrt{nh}(\hat{g}(\mathsf{x}) - g(\mathsf{x})) \xrightarrow{\mathrm{d}} N(B, v_1^2).$$

- How to build CIs for $g(\mathsf{x})$?

Main challenges

- Although $\hat{g}(\mathsf{x})$ is consistent, the test statistic is biased:

$$T_n := \sqrt{nh}(\hat{g}(\mathsf{x}) - g(\mathsf{x})) \xrightarrow{\mathrm{d}} N(B, v_1^2).$$

- How to build CIs for $g(\mathsf{x})$?

1. Undersmoothing: If $h \downarrow$, then $B \downarrow$, but $v_1 \uparrow$ (power loss).

Main challenges

- Although $\hat{g}(\mathsf{x})$ is consistent, the test statistic is biased:

$$T_n := \sqrt{nh}(\hat{g}(\mathsf{x}) - g(\mathsf{x})) \xrightarrow{\mathrm{d}} N(B, v_1^2).$$

- How to build CIs for $g(\mathsf{x})$?

1. Undersmoothing: If $h \downarrow$, then $B \downarrow$, but $v_1 \uparrow$ (power loss).
2. Bias Correction: Estimate B with $\hat{B}_n$ and use

$$\mathrm{CI}_{\mathsf{BC},n} = [(\hat{g}(\mathsf{x}) - (nh)^{-1/2}\hat{B}_n) \pm (nh)^{-1/2}\hat{v}_{1n} z_{1-\alpha/2}],$$

but variance estimate is wrong (incorrect asymptotic coverage).

Main challenges

- Although $\hat{g}(\mathsf{x})$ is consistent, the test statistic is biased:

$$T_n := \sqrt{nh}(\hat{g}(\mathsf{x}) - g(\mathsf{x})) \xrightarrow{\mathrm{d}} N(B, v_1^2).$$

- How to build CIs for $g(\mathsf{x})$?

1. Undersmoothing: If $h \downarrow$, then $B \downarrow$, but $v_1 \uparrow$ (power loss).
2. Bias Correction: Estimate B with $\hat{B}_n$ and use

$$\mathrm{CI}_{\mathsf{BC},n} = [(\hat{g}(\mathsf{x}) - (nh)^{-1/2}\hat{B}_n) \pm (nh)^{-1/2}\hat{v}_{1n} z_{1-\alpha/2}],$$

   but variance estimate is wrong (incorrect asymptotic coverage).
3. Robust Bias Correction (Calonico et al., 2014, 2018): Variability of $\hat{B}_n$ is crucial,

$$\mathrm{CI}_{\mathsf{RBC},n} = [(\hat{g}(\mathsf{x}) - (nh)^{-1/2}\hat{B}_n) \pm (nh)^{-1/2}\hat{v}_{\mathsf{RBC},n} z_{1-\alpha/2}],$$

   where $\hat{v}^2_{\mathsf{RBC},n} = \hat{\mathbb{V}}(T_n - \hat{B}_n)$.

Bootstrap inference

- Bootstrap is widely used for bias correction, but not in nonparametric regression.
- Letting $T_n^* := \sqrt{nh}(\hat{g}^*(\mathsf{x}) - \hat{g}(\mathsf{x}))$, then

$$T_n^* - \hat{B}_n \xrightarrow{\mathrm{d}^*}_{\mathrm{p}} N(0, v_1^2).$$

- However, $\hat{B}_n$ does not converge to B because it contains a stochastic component in the limit:

$$\hat{B}_n \xrightarrow{\mathrm{d}} B + \xi_2,$$

  where ξ_2 is a normal r.v. (which may or not be centered at 0).
- Thus, the bootstrap does not consistently estimate the distribution of T_n.

Bootstrap inference

- Bootstrap is widely used for bias correction, but not in nonparametric regression.
- Letting $T_n^* := \sqrt{nh}(\hat{g}^*(\mathsf{x}) - \hat{g}(\mathsf{x}))$, then

$$T_n^* - \hat{B}_n \xrightarrow{\mathrm{d}^*}_{\mathrm{p}} N(0, v_1^2).$$

- However, $\hat{B}_n$ does not converge to B because it contains a stochastic component in the limit:

$$\hat{B}_n \xrightarrow{\mathrm{d}} B + \xi_2,$$

  where ξ_2 is a normal r.v. (which may or not be centered at 0).
- Thus, the bootstrap does not consistently estimate the distribution of T_n.
- Existing solutions either:
  - explicitly correct for the bias using analytical plug-in estimates; or
  - use a different bandwidth to generate the bootstrap data (making $\hat{B}_n \xrightarrow{p} B$); or
  - use "small" bandwidths, thus removing all biases in the limit.

Bootstrap inference

- Bootstrap is widely used for bias correction, but not in nonparametric regression.
- Letting $T_n^* := \sqrt{nh}(\hat{g}^*(\mathsf{x}) - \hat{g}(\mathsf{x}))$, then

$$T_n^* - \hat{B}_n \xrightarrow{\mathrm{d}^*}_{\mathrm{p}} N(0, v_1^2).$$

- However, $\hat{B}_n$ does not converge to B because it contains a stochastic component in the limit:

$$\hat{B}_n \xrightarrow{\mathrm{d}} B + \xi_2,$$

  where ξ_2 is a normal r.v. (which may or not be centered at 0).
- Thus, the bootstrap does not consistently estimate the distribution of T_n.
- Existing solutions either:
  - explicitly correct for the bias using analytical plug-in estimates; or
  - use a different bandwidth to generate the bootstrap data (making $\hat{B}_n \xrightarrow{p} B$); or
  - use "small" bandwidths, thus removing all biases in the limit.
- **Our solution**: based on prepivoting idea.
  - Beran (1987,1988), Cavaliere, Gonçalves, Nielsen, Zanelli (2024).

Prepivoting: Cavaliere et al. (2024, JASA)

- Consider an invalid bootstrap p-value

$$\hat{p}_n := \mathbb{P}^*(T_n^* \leq T_n) \overset{\text{d}}{\not\to} U,$$

  where $U \sim \text{Uni}[0,1]$.
- Main idea: $\hat{p}_n$ is not uniform, but its cdf transform is:
  - If H_n is the cdf of $\hat{p}_n$, then

$$H_n(\hat{p}_n) \overset{\text{d}}{\to} U.$$

  - In practice, H_n is unknown but can be estimated, yielding a prepivoted p-value:

$$\tilde{p}_n = \hat{H}_n(\hat{p}_n) \overset{\text{d}}{\to} U.$$

- We consider prepivoted confidence intervals derived from $\tilde{p}_n$.

Main contributions

- We consider two residual-based bootstrap methods: $y_i^* = \tilde{g}(x_i) + \varepsilon_i^*$.
- The methods differ in the choice of $\tilde{g}(x_i)$:
  1. prepivoted global-polynomial bootstrap (PGP*),
  2. prepivoted local-polynomial bootstrap (PLP*).

Main contributions

- We consider two residual-based bootstrap methods: $y_i^* = \tilde{g}(x_i) + \varepsilon_i^*$.
- The methods differ in the choice of $\tilde{g}(x_i)$:
  1. prepivoted global-polynomial bootstrap (PGP*),
  2. prepivoted local-polynomial bootstrap (PLP*).
- Results:
  1. PGP* and PLP* both have asymptotically correct coverage.

Main contributions

- We consider two residual-based bootstrap methods: $y_i^* = \tilde{g}(x_i) + \varepsilon_i^*$.
- The methods differ in the choice of $\tilde{g}(x_i)$:
  1. prepivoted global-polynomial bootstrap (PGP*),
  2. prepivoted local-polynomial bootstrap (PLP*).
- Results:
  1. PGP* and PLP* both have asymptotically correct coverage.
  2. PGP* is equivalent to existing RBC.

Main contributions

- We consider two residual-based bootstrap methods: $y_i^* = \tilde{g}(x_i) + \varepsilon_i^*$.
- The methods differ in the choice of $\tilde{g}(x_i)$:
  1. prepivoted global-polynomial bootstrap (PGP*),
  2. prepivoted local-polynomial bootstrap (PLP*).
- Results:
  1. PGP* and PLP* both have asymptotically correct coverage.
  2. PGP* is equivalent to existing RBC.
  3. PLP* is equivalent to a new RBC-type CI based on a new bias correction (and variance adjustment).

Main contributions

- We consider two residual-based bootstrap methods: $y_i^* = \tilde{g}(x_i) + \varepsilon_i^*$.
- The methods differ in the choice of $\tilde{g}(x_i)$:
  1. prepivoted global-polynomial bootstrap (PGP*),
  2. prepivoted local-polynomial bootstrap (PLP*).
- Results:
  1. PGP* and PLP* both have asymptotically correct coverage.
  2. PGP* is equivalent to existing RBC.
  3. PLP* is equivalent to a new RBC-type CI based on a new bias correction (and variance adjustment).
  4. PLP* is more efficient than PGP* and RBC (17% shorter intervals).

Main contributions

- We consider two residual-based bootstrap methods: $y_i^* = \tilde{g}(x_i) + \varepsilon_i^*$.
- The methods differ in the choice of $\tilde{g}(x_i)$:
  1. prepivoted global-polynomial bootstrap (PGP*),
  2. prepivoted local-polynomial bootstrap (PLP*).
- Results:
  1. PGP* and PLP* both have asymptotically correct coverage.
  2. PGP* is equivalent to existing RBC.
  3. PLP* is equivalent to a new RBC-type CI based on a new bias correction (and variance adjustment).
  4. PLP* is more efficient than PGP* and RBC (17% shorter intervals).
  5. The new RBC-type CI can be implemented as a Gaussian interval with no resampling needed.

Outline

Setup

- Given a random sample $\{(y_i, x_i) : i = 1, \ldots, n\}$, we consider

$$g(\mathsf{x}) := \mathbb{E}(y_i | x_i = \mathsf{x})$$

  for a fixed evaluation point x that can be interior or on the boundary.
- We estimate $g(\mathsf{x})$ using the local linear estimator

$$\hat{g}(\mathsf{x}) = \frac{1}{nh} \sum_{i=1}^{n} w_i(\mathsf{x}) y_i,$$

  where $w_i(\cdot)$ is the equivalent kernel.
- We are interested in the statistic

$$T_n := \sqrt{nh}(\hat{g}(\mathsf{x}) - g(\mathsf{x})).$$

Assumptions

Assumption 1

(y_i, x_i) *are i.i.d. and* x_i *has bounded support* $\mathbb{S}_x$ *and continuous density* f.
For all x *in an open neighborhood of* x, *it holds that*
(i) $f(x) > 0$;
(ii) $\sigma^2(x) := \mathbb{V}[y_i | x_i = x] > 0$ *is continuous and* $\sup_x \mathbb{E}[y_i^4 | x_i = x] < \infty$;
(iii) g'' *is Hölder continuous with exponent* $\eta > 0$.

Assumption 2

(i) The kernel K *is a positive, bounded, even function with compact support* $[-1, 1]$;
(ii) The bandwidth h *is such that* $h \to 0$ *as* $n \to \infty$ *and* $\sqrt{nh^5} \to \kappa$ *for some* $\kappa \in [0, +\infty)$.

Decomposition

- Under these assumptions, we have the standard decomposition

$$T_n := \sqrt{nh}(\hat{g}(\mathsf{x}) - g(\mathsf{x})) = B_n + \xi_{1n},$$

  - B_n: "bias" component
  - $\xi_{1n} \xrightarrow{\mathrm{d}} N(0, v_1^2)$: "variance component"
- Asymptotic bias expansion:

$$B_n = \underbrace{\sqrt{nh^5}\frac{1}{2}g''(\mathsf{x})C_n(\mathsf{x})}_{B_{\mathsf{AT},n}} + \mathrm{o_p}(1), \quad C_n(\mathsf{x}) := \frac{1}{nh}\sum_{i=1}^{n} w_i(\mathsf{x})\left(\frac{x_i - \mathsf{x}}{h}\right)^2,$$

where $C_n(\mathsf{x})$ is fully observed, but $g''(x)$ is unknown.

Outline

Bootstrap inference - methods

1. Generate bootstrap data as

$$y_i^* = \tilde{g}(x_i) + \varepsilon_i^*, \qquad \varepsilon_i^* = e_i^* \hat{\varepsilon}_i, \qquad e_i^* \sim iid(0,1).$$

- PLP* and PGP* differ in the bootstrap conditional expectation, $\tilde{g}(\mathsf{x})$.

Bootstrap inference - methods

1. Generate bootstrap data as

$$y_i^* = \tilde{g}(x_i) + \varepsilon_i^*, \qquad \varepsilon_i^* = e_i^* \hat{\varepsilon}_i, \qquad e_i^* \sim iid(0,1).$$

- PLP* and PGP* differ in the bootstrap conditional expectation, $\tilde{g}(\mathsf{x})$.
- Conditional expectations in bootstrap DGP:

$$\begin{aligned} \text{PLP*}: &\quad \tilde{g}_{\mathsf{LP}}(x_i) = \hat{g}(x_i), \\ \text{PGP*}: &\quad \tilde{g}_{\mathsf{GP}}(x_i) = \tilde{\beta}_0(\mathsf{x}) + \tilde{\beta}_1(\mathsf{x})(x_i - \mathsf{x}) + \tilde{\beta}_2(\mathsf{x})(x_i - \mathsf{x})^2. \end{aligned}$$

Bootstrap inference - methods

1. Generate bootstrap data as

$$y_i^* = \tilde{g}(x_i) + \varepsilon_i^*, \qquad \varepsilon_i^* = e_i^* \hat{\varepsilon}_i, \qquad e_i^* \sim iid(0,1).$$

- PLP* and PGP* differ in the bootstrap conditional expectation, $\tilde{g}(\mathsf{x})$.
- Conditional expectations in bootstrap DGP:

$$\begin{aligned} \text{PLP*}: &\quad \tilde{g}_{\mathsf{LP}}(x_i) = \hat{g}(x_i), \\ \text{PGP*}: &\quad \tilde{g}_{\mathsf{GP}}(x_i) = \tilde{\beta}_0(\mathsf{x}) + \tilde{\beta}_1(\mathsf{x})(x_i - \mathsf{x}) + \tilde{\beta}_2(\mathsf{x})(x_i - \mathsf{x})^2. \end{aligned}$$

- PLP* proposed by early statistics literature, but known to be invalid with 'large' bandwidths:
  - Härdle and Bowman (1988), Härdle and Marron (1991)
- GLP* suggested in the RDD literature but without prepivoting:
  - Bartalotti et al. (2017) and He and Bartalotti (2020)

Bootstrap inference - methods

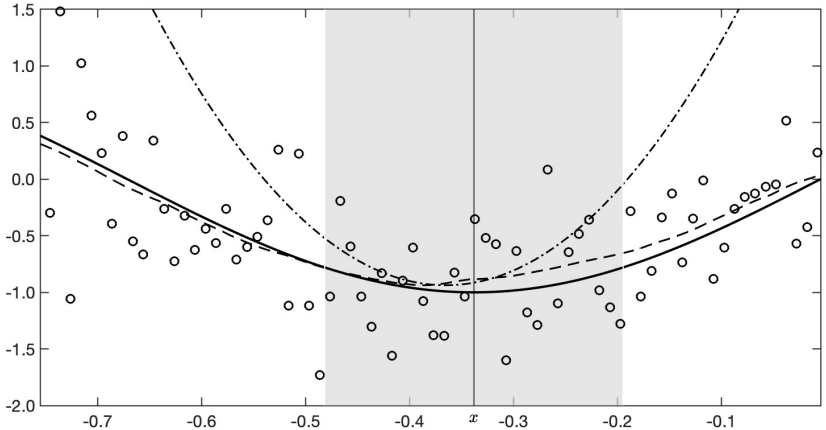


Figure 1: Bootstrap conditional expectations: GP (dot-dashed) and LP (dashed)

Bootstrap inference - methods

2. Calculate the bootstrap local linear estimator $\hat{g}^*(\mathsf{x})$ and test statistic

$$T_n^* = \sqrt{nh}(\hat{g}^*(\mathsf{x}) - \tilde{g}(\mathsf{x})),$$

where $\hat{g}^*(\mathsf{x})$ is defined as

$$\hat{g}^*(\mathsf{x}) := \frac{1}{nh}\sum_{i=1}^{n} w_i(\mathsf{x}) y_i^*.$$

- In both cases (LP and GP), we have the decomposition

$$T_n^* = \underbrace{\hat{B}_n}_{\text{bias}} + \underbrace{\xi_{1n}^*}_{\text{variance}}, \qquad \xi_{1n}^* \xrightarrow{\mathrm{d}^*}_{\mathrm{p}} N(0, v_1^2),$$

but the bias (and variance) components differ.

Bootstrap bias components

- Recall the bias expansion

$$B_n = \sqrt{nh}\left(\frac{1}{nh}\sum_{i=1}^{n} w_i(\mathsf{x})g(x_i) - g(\mathsf{x})\right) = \underbrace{\sqrt{nh^5}\frac{1}{2}g''(\mathsf{x})C_n(\mathsf{x})}_{=B_{\mathsf{AT},n}} + \mathrm{o_p}(1).$$

Bootstrap bias components

- Recall the bias expansion

$$B_n = \sqrt{nh}\left(\frac{1}{nh}\sum_{i=1}^{n} w_i(\mathsf{x})g(x_i) - g(\mathsf{x})\right) = \underbrace{\sqrt{nh^5}\frac{1}{2}g''(\mathsf{x})C_n(\mathsf{x})}_{=B_{\mathsf{AT},n}} + \mathrm{o_p}(1).$$

- The implicit GP bootstrap bias estimator is

$$\hat{B}_{\mathsf{GP},n} = \sqrt{nh^5}\frac{1}{2}\tilde{g}''(\mathsf{x})C_n(\mathsf{x}).$$

- The implicit LP bootstrap bias estimator is

$$\hat{B}_{\mathsf{LP},n} = \sqrt{nh}\left(\frac{1}{nh}\sum_{i=1}^{n} w_i(\mathsf{x})\hat{g}(x_i) - \hat{g}(\mathsf{x})\right).$$

Bootstrap bias expansions

- For interior x,

$$\hat{B}_{\mathsf{GP},n} = B_{\mathsf{AT},n} + \underbrace{\xi_{\mathsf{GP},2n}}_{\xrightarrow{d} N(0,v^2_{\mathsf{GP}})} + o_p(1),$$

and

$$\hat{B}_{\mathsf{LP},n} = B_{\mathsf{AT},n} + \underbrace{\xi_{\mathsf{LP},2n}}_{\xrightarrow{d} N(0,v^2_{\mathsf{LP}})} + o_p(1).$$

  - The presence of $\xi_{\mathsf{LP},2n}$ and $\xi_{\mathsf{GP},2n}$ (both asymptotically Gaussian with mean zero) implies that none of these methods estimate B_n consistently (although they are centered at $B_{\mathsf{AT},n}$).
  - Crucially, their variances are different, which will result in different efficiency properties.

Bootstrap bias expansions

- For interior x,

$$\hat{B}_{\mathsf{GP},n} = B_{\mathsf{AT},n} + \underbrace{\xi_{\mathsf{GP},2n}}_{\xrightarrow{\mathrm{d}} N(0,v^2_{\mathsf{GP}})} + \mathrm{o_p}(1),$$

and

$$\hat{B}_{\mathsf{LP},n} = B_{\mathsf{AT},n} + \underbrace{\xi_{\mathsf{LP},2n}}_{\xrightarrow{\mathrm{d}} N(0,v^2_{\mathsf{LP}})} + \mathrm{o_p}(1).$$

  - The presence of $\xi_{\mathsf{LP},2n}$ and $\xi_{\mathsf{GP},2n}$ (both asymptotically Gaussian with mean zero) implies that none of these methods estimate B_n consistently (although they are centered at $B_{\mathsf{AT},n}$).
  - Crucially, their variances are different, which will result in different efficiency properties.
- For x on the boundary, $\hat{B}_{\mathsf{LP},n}$ is not centered at $B_{\mathsf{AT},n}$.
- We propose a simple modification of T^*_n that adapts to the nature of x.

Bootstrap inference - boundary

- Very briefly...
- For x on the boundary, $\hat{B}_{\mathsf{LP},n}$ is not centered at $B_{\mathsf{AT},n}$.
- However, there is a simple factor $Q_n := C_n(\mathsf{x})/C_{2n}(\mathsf{x})$, which is a function only of K, h, and the data, $\mathscr{X}_n$, such that

$$Q_n \hat{B}_{\mathsf{LP},n} = B_{\mathsf{AT},n} + Q_n \xi_{\mathsf{LP},2n} + \mathrm{o_p}(1).$$

- Thus, we consider the modified statistic $T^*_{m,n} = Q_n T^*_n$, whose bias component is correctly centered (and variance is changed by factor Q_n^2).
- For interior x we show that $Q_n \to_{\mathrm{p}} 1$, so the modified statistic is adaptive.
- For the slides we consider an interior point from now on, but this modification allows everything to be generalized to a boundary point – and hence RDD.

Outline

Invalidity of the standard bootstrap p-value

- We can show that

$$\begin{aligned}\hat{p}_n &:= \mathbb{P}^*(T_n^* \leq T_n) \\ &= \mathbb{P}^*\left(T_n^* - \hat{B}_n \leq T_n - \hat{B}_n\right) \\ &= \Phi\left(\frac{T_n - \hat{B}_n}{v_1}\right) + o_p(1).\end{aligned}$$

Invalidity of the standard bootstrap p-value

- We can show that

$$\begin{aligned}\hat{p}_n &:= \mathbb{P}^*(T_n^* \leq T_n) \\ &= \mathbb{P}^*\left(T_n^* - \hat{B}_n \leq T_n - \hat{B}_n\right) \\ &= \Phi\left(\frac{T_n - \hat{B}_n}{v_1}\right) + o_p(1).\end{aligned}$$

- But

$$T_n - \hat{B}_n = \underbrace{(T_n - B_n)}_{\xrightarrow{d} N(0,v_1^2)} - \underbrace{(\hat{B}_n - B_n)}_{\xrightarrow{d} N(0,v_2^2)} \xrightarrow{d} N(0, v_{\mathsf{P}}^2) \neq N(0, v_1^2),$$

  - $v_{\mathsf{P}}^2 = v_1^2 + v_2^2 - 2v_{12} \neq v_1^2$ since $v_2^2 > 0$ (v_2^2 is v_{GP}^2 or v_{LP}^2 depending on the bootstrap used).

Invalidity of the standard bootstrap p-value

- We can show that

$$\begin{aligned}\hat{p}_n &:= \mathbb{P}^*(T_n^* \leq T_n) \\ &= \mathbb{P}^*\left(T_n^* - \hat{B}_n \leq T_n - \hat{B}_n\right) \\ &= \Phi\left(\frac{T_n - \hat{B}_n}{v_1}\right) + o_p(1).\end{aligned}$$

- But

$$T_n - \hat{B}_n = \underbrace{(T_n - B_n)}_{\xrightarrow{d} N(0,v_1^2)} - \underbrace{(\hat{B}_n - B_n)}_{\xrightarrow{d} N(0,v_2^2)} \xrightarrow{d} N(0, v_{\mathsf{P}}^2) \neq N(0, v_1^2),$$

  - $v_{\mathsf{P}}^2 = v_1^2 + v_2^2 - 2v_{12} \neq v_1^2$ since $v_2^2 > 0$ (v_2^2 is v_{GP}^2 or v_{LP}^2 depending on the bootstrap used).

- We obtain

$$\hat{p}_n \xrightarrow{d} \Phi(m_{\mathsf{P}}\Phi^{-1}(U)) \neq U, \quad \text{because} \quad m_{\mathsf{P}} := v_{\mathsf{P}}/v_1 \neq 1.$$

Prepivoting bootstrap p-values

- For any u,
$$H_n(u) := \mathbb{P}(\hat{p}_n \leq u) \to \Phi(m_{\mathsf{P}}^{-1}\Phi^{-1}(u)), \quad m_{\mathsf{P}} := v_{\mathsf{P}}/v_1.$$

- A uniformly consistent plug-in estimator of $H_n(u)$ is
$$\hat{H}_n(u) = \Phi(\hat{m}_{\mathsf{P}}^{-1}\ \Phi^{-1}(u)), \quad \hat{m}_{\mathsf{P}} := \hat{v}_{\mathsf{P}}/\hat{v}_1.$$

- Prepivoted p-value:
$$\tilde{p}_n = \Phi(\hat{m}_{\mathsf{P}}^{-1}\Phi^{-1}(\hat{p}_n)) \xrightarrow{\mathrm{d}} U.$$

- We can obtain valid prepivoted confidence intervals derived from $\tilde{p}_n$.

Bootstrap confidence intervals

- Let $\hat{L}_n(u) := \mathbb{P}^*(T_n^* \leq u)$ and $\hat{L}_n^{-1}(\alpha)$ be its αth quantile. Then

$$CI_{\texttt{naive}} := \left[\hat{g}(\mathsf{x}) - (nh)^{-1/2}\hat{L}_n^{-1}(1-\alpha/2), \hat{g}(\mathsf{x}) - (nh)^{-1/2}\hat{L}_n^{-1}(\alpha/2)\right].$$

- This interval is invalid:

$$\mathbb{P}(g(\mathsf{x}) \in CI_{\texttt{naive}}) = \mathbb{P}(\alpha/2 \leq \hat{p}_n \leq 1-\alpha/2) \not\to 1-\alpha \quad \text{because} \quad \hat{p}_n \xrightarrow{\text{d}} \not\to U.$$

Bootstrap confidence intervals

- Let $\hat{L}_n(u) := \mathbb{P}^*(T_n^* \leq u)$ and $\hat{L}_n^{-1}(\alpha)$ be its αth quantile. Then

$$CI_{\texttt{naive}} := \left[\hat{g}(\mathsf{x}) - (nh)^{-1/2}\hat{L}_n^{-1}(1-\alpha/2), \hat{g}(\mathsf{x}) - (nh)^{-1/2}\hat{L}_n^{-1}(\alpha/2)\right].$$

- This interval is invalid:

$$\mathbb{P}(g(\mathsf{x}) \in CI_{\texttt{naive}}) = \mathbb{P}(\alpha/2 \leq \hat{p}_n \leq 1-\alpha/2) \not\to 1-\alpha \quad \text{because} \quad \hat{p}_n \xrightarrow{\text{d}} \not\to U.$$

- A valid interval can instead be derived from

$$\mathbb{P}(\alpha/2 \leq \underbrace{\hat{H}_n(\hat{p}_n)}_{=\tilde{p}_n} \leq 1-\alpha/2) \to 1-\alpha.$$

Bootstrap confidence intervals

- Let $\hat{L}_n(u) := \mathbb{P}^*(T_n^* \leq u)$ and $\hat{L}_n^{-1}(\alpha)$ be its αth quantile. Then

$$CI_{\texttt{naive}} := \left[\hat{g}(\mathsf{x}) - (nh)^{-1/2}\hat{L}_n^{-1}(1-\alpha/2), \hat{g}(\mathsf{x}) - (nh)^{-1/2}\hat{L}_n^{-1}(\alpha/2)\right].$$

- This interval is invalid:

$$\mathbb{P}(g(\mathsf{x}) \in CI_{\texttt{naive}}) = \mathbb{P}(\alpha/2 \leq \hat{p}_n \leq 1-\alpha/2) \not\to 1-\alpha \quad \text{because} \quad \hat{p}_n \xrightarrow{\text{d}} \not\to U.$$

- A valid interval can instead be derived from

$$\mathbb{P}(\alpha/2 \leq \underbrace{\hat{H}_n(\hat{p}_n)}_{=\tilde{p}_n} \leq 1-\alpha/2) \to 1-\alpha.$$

- This yields

$$CI_{\texttt{P}} := \left[\hat{g}(\mathsf{x}) - (nh)^{-1/2}\hat{L}_n^{-1}(\hat{H}_n^{-1}(1-\alpha/2)), \hat{g}(\mathsf{x}) - (nh)^{-1/2}\hat{L}_n^{-1}(\hat{H}_n^{-1}(\alpha/2))\right],$$

where $\hat{H}_n^{-1}(\alpha)$ is the αth quantile of $\hat{H}_n(u)$, the plug-in estimate of the cdf of $\hat{p}_n$.

Bootstrap inference - interval

Result 1

Under Assumptions 1 and 2, we have that

$$\mathrm{CI}_{\mathsf{P}} := \left[\hat{g}(\mathsf{x}) - (nh)^{-1/2} \underbrace{\hat{L}_n^{-1}(\hat{H}_n^{-1}(1-\alpha/2))}_{\approx \hat{B}_n + \hat{v}_{\mathsf{P}}\Phi^{-1}(1-\alpha/2)}, \hat{g}(\mathsf{x}) - (nh)^{-1/2} \underbrace{\hat{L}_n^{-1}(\hat{H}_n^{-1}(\alpha/2))}_{\approx \hat{B}_n + \hat{v}_{\mathsf{P}}\Phi^{-1}(\alpha/2)}\right]$$

is asymptotically equivalent to the RBC-type interval

$$\mathrm{CI}_{\mathsf{P-RBC}} := \left[(\hat{g}(\mathsf{x}) - (nh)^{-1/2}\hat{B}_n) \pm (nh)^{-1/2} z_{1-\alpha/2}\hat{v}_{\mathsf{P}}\right].$$

If the bootstrap is Gaussian then this result is exact.

- The result follows because for any α, $\hat{L}_n^{-1}(\hat{H}_n^{-1}(\alpha)) = \hat{v}_{\mathsf{P}}\Phi^{-1}(\alpha) + \hat{B}_n + \mathrm{o_p}(1)$.
- Prepivoting implicitly performs both a bias correction and an adjustment of the standard error for the added uncertainty.

Equivalence between GP bootstrap and existing RBC

- For the GP bootstrap we have that

$$\hat{B}_{\mathsf{GP},n} = \sqrt{nh^5}\frac{1}{2}\tilde{g}''(\mathsf{x})C_n(\mathsf{x}) = \hat{B}_{\mathsf{RBC},n} \quad \text{and} \quad \hat{v}_{\mathsf{PGP},n} = \hat{v}_{\mathsf{RBC},n}.$$

- Consequently, the prepivoted GP interval satisfies

$$\mathrm{CI}_{\mathsf{PGP}} = \mathrm{CI}_{\mathsf{RBC}} + \mathrm{o_p}((nh)^{-1/2}),$$

  and there is no $\mathrm{o_p}$-term if the bootstrap is Gaussian.
- $\mathrm{CI}_{\mathsf{PLP}}$ is instead based on a different bias estimator and variance adjustment.
- Different, but better?

Outline

Efficiency results

- Asymptotic length of CI_{PLP} is less than that of CI_{RBC}.
- Compare the asymptotic variances, $\operatorname{plim} \hat{v}^2_{\mathsf{PLP},n}$ and $\operatorname{plim} \hat{v}^2_{\mathsf{RBC},n}$:

$$\operatorname{plim} \begin{bmatrix} \hat{v}^2_{\mathsf{LP},n} \\ \hat{v}^2_{\mathsf{GP},n} \end{bmatrix} = \begin{bmatrix} v^2_{\mathsf{PLP}} \\ v^2_{\mathsf{RBC}} \end{bmatrix} = \frac{\sigma^2(\mathsf{x})}{f(\mathsf{x})} \begin{bmatrix} \mathcal{K}_{\mathsf{PLP}} \\ \mathcal{K}_{\mathsf{RBC}} \end{bmatrix},$$

where $\mathcal{K}_{\mathsf{PLP}}$ and $\mathcal{K}_{\mathsf{RBC}}$ depend only on kernel K and nature of x (int or bnd).

	Interior		Boundary	
Kernel	$\mathcal{K}_{\mathsf{PLP}}$	$\mathcal{K}_{\mathsf{RBC}}$	$\mathcal{K}_{\mathsf{PLP}}$	$\mathcal{K}_{\mathsf{RBC}}$
Triangular	0.95	1.33	7.18	10.29
Epanechnikov	0.85	1.25	6.80	9.82
Biweight	1.01	1.41	7.67	10.87
Triweight	1.15	1.55	8.54	11.87

Efficiency results

- Length of asymptotic confidence interval is proportional to $\mathcal{K}^{1/2}$.
- Relative length of intervals can be found as ratios of different $\mathcal{K}^{1/2}$.

Table 1: Relative (asymptotic) interval lengths, PLP* and RBC

	Triangular	Uniform	Epanechnikov	Biweight	Triweight
Interior point	0.85	0.86	0.83	0.85	0.86
Boundary point	0.83	0.85	0.83	0.84	0.85

- Table 1 holds regardless of the distribution of errors, distribution of regressors, and evaluation point.
- LP method relies on an estimate of the conditional mean function at each value of the regressor — additional smoothing induces a more efficient bias-correction in the sense that the overall variance of the debiased statistic is smaller.

Efficiency results

- Intuitively, there is an asymptotic equivalent kernel: for some implied bias correction, $\hat{B}_n$,

$$T_n - \hat{B}_n = \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} w_{\mathsf{bc}}\Big(\frac{x_i - \mathsf{x}}{h}\Big) y_i (1 + \mathrm{o_p}(1)).$$

- Then $\mathcal{K} = f^2(\mathsf{x}) \int w^2_{\mathsf{bc}}(u) du$, and setting $f(\mathsf{x}) = 1$ we plot $w_{\mathsf{bc}}(u)$.

Efficiency results

Figure 2: Asymptotic equivalent kernels for interior points

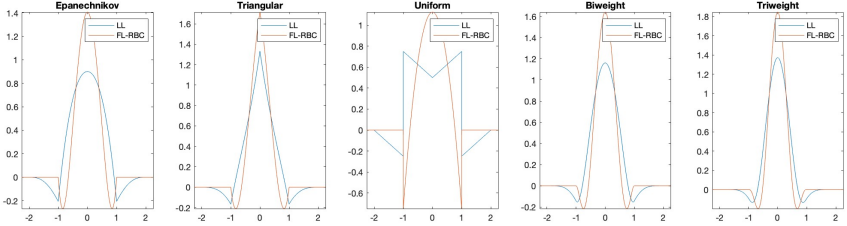
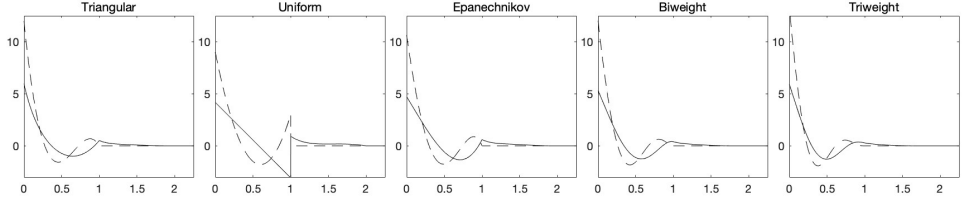


Figure 3: Asymptotic equivalent kernels for boundary points



Monte Carlo DGP1 (nonparametric regression)

- We report CIs based on 5000 Monte Carlo iterations.
- Wild bootstrap with HC3 residuals and Gaussian shocks (no resampling).
- DGP1:

$$
\begin{aligned}
y_i &= g(x_i) + \varepsilon_i, \\
x_i &\sim U[-1,1] \quad \text{and} \quad \varepsilon_i \sim N(0,1), \\
g(x) &= \frac{\sin(3\pi x/2)}{1 + 18x^2(\operatorname{sgn}(x)+1)}.
\end{aligned}
$$

- Same as Model 5 in Calonico et al. (2018), based on Hall and Horowitz (2013).
- Bandwidths:
  - h is the infeasible MSE-optimal bandwidth,
  - $\hat{h}$ is the feasible CE-optimal bandwidth.
- Evaluation points: an interior point $\mathsf{x} = -1/3$ and a boundary point $\mathsf{x} = -1$.

Simulation results: DGP1 interior point

Table 2: Coverage and length of 95% confidence intervals - Nonparametric regression

			Coverage				Length			
n	h	$\bar{h}$	GP	LP	RBC	PLP	GP	LP	RBC	PLP
250	h	0.189	81.5	88.9	93.6	94.2	0.635	0.635	0.915	0.765
	$\hat{h}$	0.359	80.3	79.3	93.1	86.5	0.461	0.461	0.664	0.554
500	h	0.165	82.1	89.2	94.2	94.4	0.479	0.479	0.690	0.573
	$\hat{h}$	0.303	81.6	85.4	94.1	91.1	0.353	0.353	0.510	0.422
1000	h	0.143	81.9	90.0	94.4	94.9	0.361	0.361	0.521	0.431
	$\hat{h}$	0.256	81.9	87.7	94.6	93.8	0.270	0.270	0.390	0.322
2000	h	0.125	83.3	90.7	95.1	95.4	0.273	0.273	0.394	0.326
	$\hat{h}$	0.217	82.5	89.4	94.9	94.5	0.207	0.207	0.299	0.247

Simulation results: DGP1 boundary point

Table 3: Coverage and length of 95% confidence intervals - Nonparametric regression

			Coverage				Length			
n	h	$\bar{h}$	GP	LP	RBC	PLP	GP	LP	RBC	PLP
250	h	0.353	77.7	85.6	90.9	92.2	1.277	1.277	1.901	1.596
	$\hat{h}$	0.373	77.8	83.2	91.0	91.2	1.271	1.271	1.895	1.602
500	h	0.307	79.9	87.6	93.3	93.8	0.961	0.961	1.426	1.191
	$\hat{h}$	0.327	80.6	85.6	93.2	92.8	0.946	0.946	1.403	1.174
1000	h	0.267	80.6	87.8	93.8	94.0	0.725	0.725	1.071	0.894
	$\hat{h}$	0.285	80.7	86.7	93.9	93.3	0.711	0.711	1.051	0.878
2000	h	0.233	81.4	89.5	94.7	94.8	0.549	0.549	0.812	0.676
	$\hat{h}$	0.257	81.5	88.1	94.5	94.3	0.526	0.526	0.779	0.649

Monte Carlo DGP2 (RDD)

- We report CIs based on 5000 Monte Carlo iterations.
- Wild bootstrap with HC3 residuals and Gaussian shocks (no resampling).
- DGP2:

$$
\begin{aligned}
y_i &= g(x_i) + \varepsilon_i, \\
x_i &\sim 2\mathcal{B}(2,4) - 1 \quad \text{and} \quad \varepsilon_i \sim N(0, .1295^2), \\
g(x) &= \begin{cases} 3.71 + 2.30x + 3.28x^2 + 1.45x^3 + 0.23x^4 + 0.03x^5 & \text{if } x < 0 \\ 0.26 + 18.49x - 54.81x^2 + 74.30x^3 - 45.02x^4 + 9.83x^5 & \text{if } x \geq 0 \end{cases}
\end{aligned}
$$

- Same as Model 2 in Calonico et al. (2014), based on the Ludwig and Miller (2007) dataset.
- Bandwidths:
  - h is the infeasible MSE-optimal bandwidth,
  - $\hat{h}$ is the feasible CE-optimal bandwidth.

Simulation results: DGP2

Table 4: Coverage and length of 95% confidence intervals - RDD

			Coverage				Length			
n	h	$\bar{h}$	GP	LP	RBC	PLP	GP	LP	RBC	PLP
500	h	0.082	80.6	86.7	93.3	93.2	0.345	0.345	0.589	0.470
	$\hat{h}$	0.073	79.6	86.0	93.5	94.0	0.373	0.373	0.674	0.539
1000	h	0.072	81.1	87.8	94.3	94.2	0.249	0.249	0.388	0.318
	$\hat{h}$	0.060	81.0	86.8	93.8	93.9	0.276	0.276	0.441	0.358
2000	h	0.063	81.2	88.0	94.5	94.7	0.185	0.185	0.279	0.231
	$\hat{h}$	0.049	80.4	87.8	94.4	94.4	0.210	0.210	0.321	0.265
4000	h	0.054	81.3	88.3	94.8	95.1	0.138	0.138	0.205	0.170
	$\hat{h}$	0.041	81.2	88.2	94.8	94.8	0.160	0.160	0.240	0.199

Conclusions

- We propose prepivoting inference for nonparametric regression and RDD.
- Our results show that:

1. Prepivoted intervals bear a close resemblance to RBC intervals.
2. RBC intervals are equivalent to prepivoted intervals based on a bootstrap DGP that estimates $g(x_i)$ from a local quadratic nonparametric regression of g at $\times$.
3. Prepivoting based on a different (well-known) bootstrap DGP in nonparametric statistics yields a different RBC-type interval:
   - Requires a modification of prepivoting to work with boundary points (RDD).
   - Has improved length properties while maintaining correct coverage — regardless of distribution of the errors, distribution of regressors, and evaluation point.
   - Can be implemented as a Gaussian interval without resampling.

Real world		LP bootstrap world
$y_i = g(x_i) + \varepsilon_i$	$\longrightarrow$	$y_i^* = \hat{g}(x_i) + \varepsilon_i^*$
$\downarrow$		$\downarrow$
$T_n := \sqrt{nh}(\hat{g}(\mathsf{x}) - g(\mathsf{x}))$		$T_n^* := \sqrt{nh}(\hat{g}^*(\mathsf{x}) - \hat{g}(\mathsf{x}))$
$\downarrow$		$\downarrow$
$\frac{T_n - B_n}{v_1} \xrightarrow{\mathrm{d}} N(0,1)$		$\frac{T_n^* - \hat{B}_{\mathsf{LP},n}}{v_1} \xrightarrow{\mathrm{d}^*_{\mathrm{p}}} N(0,1)$
$B_n = \underbrace{\sqrt{nh^5}\frac{1}{2}g''(x)C_n}_{B_{\mathsf{AT},n}} + o_p(1)$		$\hat{B}_{\mathsf{LP},n} = \underbrace{\sqrt{nh^5}\frac{1}{2}g''(x)C_{2n}}_{B_{\mathsf{LP},n}} + \xi_{\mathsf{LP},2n} + o_p(1)$

- $B_{\mathsf{LP},n} - B_{\mathsf{AT},n} = o_p(1)$ when x is interior but not otherwise.
- Even for interior points, the presence of $\xi_{\mathsf{LP},2n}$ invalidates the bootstrap.

Invalidity of LP bootstrap p-value

- As before,

$$\hat{p}_n := \mathbb{P}^*(T_n^* \leq T_n) = \Phi\left(\frac{T_n - \hat{B}_{\mathsf{LP},n}}{v_1}\right) + o_p(1),$$

but now

$$\begin{aligned} T_n - \hat{B}_{\mathsf{LP},n} &= (T_n - B_n) - (\hat{B}_{\mathsf{LP},n} - B_n) \\ &= \underbrace{\xi_{1n} - \xi_{\mathsf{LP},2n}}_{\xrightarrow{\mathrm{d}} N(0, v_{\mathsf{P}}^2)} - \underbrace{(B_{\mathsf{LP},n} - B_{\mathsf{AT,n}})}_{=A+o_p(1)} \xrightarrow{\mathrm{d}} N(-A, v_{\mathsf{P}}^2). \end{aligned}$$

Invalidity of LP bootstrap p-value

- As before,

$$\hat{p}_n := \mathbb{P}^*(T_n^* \leq T_n) = \Phi\left(\frac{T_n - \hat{B}_{\mathsf{LP},n}}{v_1}\right) + o_p(1),$$

  but now

$$\begin{aligned} T_n - \hat{B}_{\mathsf{LP},n} &= (T_n - B_n) - (\hat{B}_{\mathsf{LP},n} - B_n) \\ &= \underbrace{\xi_{1n} - \xi_{\mathsf{LP},2n}}_{\xrightarrow{\mathrm{d}} N(0, v_{\mathsf{P}}^2)} - \underbrace{(B_{\mathsf{LP},n} - B_{\mathsf{AT,n}})}_{=A+o_p(1)} \xrightarrow{\mathrm{d}} N(-A, v_{\mathsf{P}}^2). \end{aligned}$$

- For any u,

$$H_n(u) := \mathbb{P}(\hat{p}_n \leq u) \to \Phi\left(-a/m_{\mathsf{LP}} + m_{\mathsf{LP}}^{-1}\Phi^{-1}(u)\right)$$

  where $a := -A/v_1 = 0$ only if x is an interior point.

- Standard prepivoting does not work.

Modified prepivoting

- We consider

$$\hat{p}_{m,n} := \mathbb{P}^*(T^*_{m,n} \leq T_n), \quad \text{where} \quad T^*_{m,n} := Q_n T^*_n \quad \text{and} \quad Q_n := B_{\mathsf{AT},n}/B_{\mathsf{LP},n} = C_n/C_{2n}.$$

Modified prepivoting

- We consider

$$\hat{p}_{m,n} := \mathbb{P}^*(T^*_{m,n} \leq T_n), \quad \text{where} \quad T^*_{m,n} := Q_n T^*_n \quad \text{and} \quad Q_n := B_{\mathsf{AT},n}/B_{\mathsf{LP},n} = C_n/C_{2n}.$$

- We can show that the cdf of $\hat{p}_{m,n}$, although not U, no longer depends on A:

$$\hat{p}_{m,n} = \Phi\Big(\underbrace{\frac{T_n - Q_n \hat{B}_{\mathsf{LP},n}}{Q_n v_1}}_{\xrightarrow{\mathrm{d}} N(0, m^2_{\mathsf{mLP}}), \quad m^2_{\mathsf{mLP}} \neq 1} \Big) + o_p(1).$$

Modified prepivoting

- We consider

$$\hat{p}_{m,n} := \mathbb{P}^*(T^*_{m,n} \leq T_n), \quad \text{where} \quad T^*_{m,n} := Q_n T^*_n \quad \text{and} \quad Q_n := B_{\mathsf{AT},n}/B_{\mathsf{LP},n} = C_n/C_{2n}.$$

- We can show that the cdf of $\hat{p}_{m,n}$, although not U, no longer depends on A:

$$\hat{p}_{m,n} = \Phi\Big(\underbrace{\frac{T_n - Q_n \hat{B}_{\mathsf{LP},n}}{Q_n v_1}}_{\xrightarrow{\mathrm{d}} N(0, m^2_{\mathsf{mLP}}), \quad m^2_{\mathsf{mLP}} \neq 1} \Big) + o_p(1).$$

- Hence, we can modify $\hat{p}_{m,n}$ by standard prepivoting:

$$\tilde{p}_{m,n} := \Phi((\hat{m}_{\mathsf{mLP},n})^{-1} \Phi^{-1}(\hat{p}_{m,n})).$$

- We can also obtain prepivoted intervals based on $\tilde{p}_{m,n}$.

Modified prepivoted LP intervals

- We can show

$$CI_{\mathsf{mPLP}} = \underbrace{[\hat{g}(\mathsf{x}) - (nh)^{-1/2}\hat{B}_{\mathsf{mPLP},n} \pm \Phi^{-1}_{1-\alpha/2}(nh)^{-1/2}\hat{v}_{\mathsf{mPLP},n}]}_{=CI_{\mathsf{RBC}}} + o_p((nh)^{-1/2})$$

- $\hat{B}_{\mathsf{mPLP},n}$ and $\hat{v}_{\mathsf{mPLP},n}$ are known in closed form.
- Hence, we can implement the prepivoted intervals without resampling.
- We show that the asymptotic length of CI_{mPLP} is smaller than that of $CI_{\mathsf{GP}} = CI_{\mathsf{RBC}}$.
- This result is obtained by comparing $\operatorname{plim} \hat{v}_{\mathsf{mPLP},n}$ with $\operatorname{plim} \hat{v}_{\mathsf{RBC},n}$:

$$\begin{pmatrix} v^2_{\mathsf{mPLP}} \\ v^2_{\mathsf{RBC}} \end{pmatrix} = \frac{\sigma^2(x)}{f(x)} \begin{pmatrix} \mathcal{K}_{\mathsf{mPLP}} \\ \mathcal{K}_{\mathsf{RBC}} \end{pmatrix},$$

where $\mathcal{K}_{\mathsf{mPLP}}$ and $\mathcal{K}_{\mathsf{RBC}}$ are constants that depend only on the kernel K.