The HulC

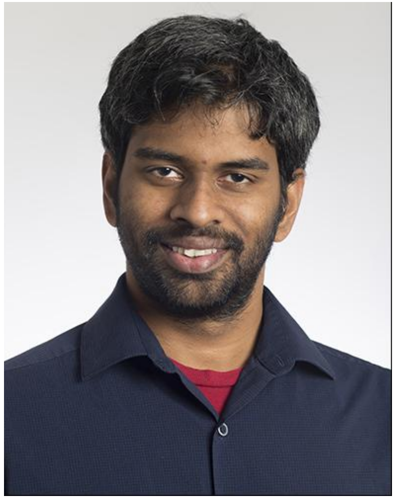
Hull based Confidence Regions

Arun Kumar Kuchibhotla

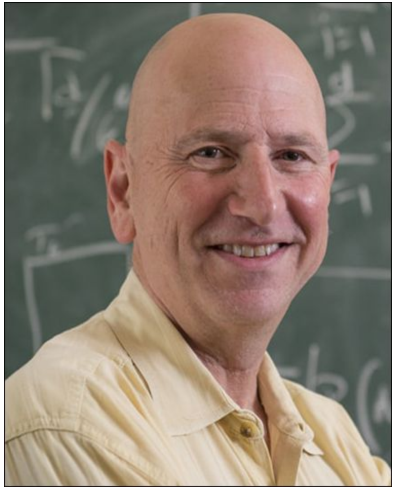
Carnegie Mellon University

Collaborators

Sivaraman Balakrishnan



Larry Wasserman



Outline

- ❖ Inference in isotonic regression
- ❖ HulC
- ❖ Adaptive HulC
- ❖ Numerical examples

Isotonic Regression

Univariate isotonic regression

Consider bivariate data $(X_i, Y_i), i \leq n$, from the model

$$Y_i = f_0(X_i) + \xi_i, \quad \mathbb{E}[\xi_i | X_i] = 0,$$

where $f_0(\cdot)$ is a non-decreasing function.

The covariates can either be fixed on a grid or be random.

Problem: suppose $x^\star$ is a (fixed) point in the support of X_i's.

How do we do inference for $f_0(x^\star)$?

Pointwise asymptotics

The least squares estimator (LSE) is given by

$$\widehat{f} := \underset{f:\text{increasing}}{\text{argmin}} \frac{1}{n} \sum_{i=1}^{n} (Y_i - f(X_i))^2.$$

If f_0 is continuously differentiable at $x^\star$ and $f_0'(x^\star) > 0$, then

$$n^{1/3}(\widehat{f}(x^\star) - f_0(x^\star)) \xrightarrow{d} \left(\frac{4\sigma^2(x^\star) f_0'(x^\star)}{h(x^\star)}\right)^{1/3} \mathbb{C},$$

where $\mathbb{C}$ has the Chernoff distribution obtained as the minimizer of a drifted two-sided Brownian motion starting at 0.

Wright (1981) The asymptotic behavior of monotone regression estimates. The Annals of Statistics.

Pointwise asymptotics

In general, if f_0 satisfies

$$f_0'(x^\star) = \cdots = f_0^{(\beta-1)}(x^\star) = 0,$$
$$f_0^{(\beta)}(x^\star) = \beta! A.$$

$$\frac{|f_0(x) - f_0(x^\star)|}{|x - x^\star|^\beta} = A(1 + o(1)) \quad \text{as} \quad x \to x^\star,$$

for some $\beta \geq 1/2$, then

$$n^{\beta/(2\beta+1)}(\widehat{f}(x^\star) - f_0(x^\star)) \xrightarrow{d} \left(\frac{\sigma^2(x^\star)A^{1/\beta}}{h(x^\star)(\beta+1)^{1/\beta}}\right)^{\beta/(2\beta+1)} \mathbb{C}_\beta,$$

where $\mathbb{C}_\beta$ is the slope at zero of the greatest convex minorant of $B(t) + |t|^{\beta+1}$.

Wright (1981) The asymptotic behavior of monotone regression estimates. The Annals of Statistics.

Some notes

- The rate of convergence and the limiting distribution depends on the local smoothness of f_0.
- The limiting distribution depends on additional (unknown) nuisance components such as the design density and heteroscedasticity.
- Estimation of those nuisance components involves tuning parameters in general but estimating $f_0(x^\star)$ does not.
- Traditional methods of inference such as Wald's, bootstrap, and subsampling fall short in various ways.

Some notes (cont.)

- Similar comments hold true for other shape constrained problems such as convex regression and current status models.
- These comments also hold for multivariate covariates.
- Recently, Deng, Han, and Zhang (2020) derived a pivotal statistic using the kinks of the estimator in the model where errors are independent of covariates and uniformly distributed covariates.

Deng, Han, and Zhang (2020+). Confidence intervals for multiple isotonic regression and other monotone models. The Annals of Statistics.

Introducing The HulC

Motivation

- Suppose $\hat{\theta}$ is a consistent estimator of $\theta_0 \in \mathbb{R}$.
- With $\hat{\theta}^{(1)}, \hat{\theta}^{(2)}, \ldots,$ representing independent copies of $\hat{\theta}$, we *expect* that the minimum and maximum of the estimators contain θ_0.

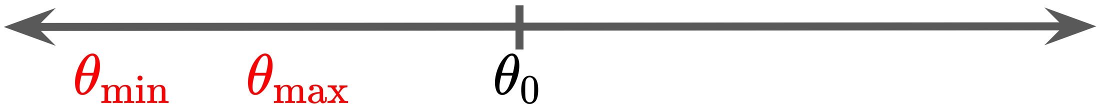
Problem: When does this hold true? And how many estimators are needed for valid coverage?

Answer: The estimator is not pathologically “asymmetric.”

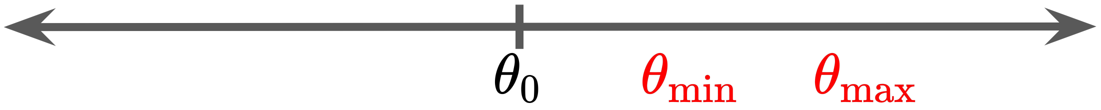
Simple Calculations

Suppose we have two estimators $\hat{\theta}^{(1)}, \hat{\theta}^{(2)},$ symmetric about θ_0

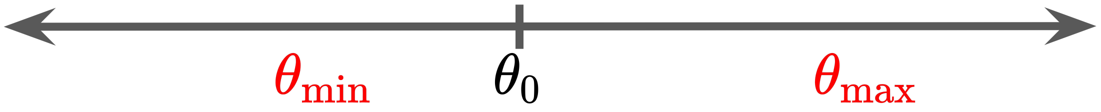
$$\theta_{\min} = \min\{\hat{\theta}^{(1)}, \hat{\theta}^{(2)}\}, \quad \text{and} \quad \theta_{\max} = \max\{\hat{\theta}^{(1)}, \hat{\theta}^{(2)}\}.$$



$$\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$



$$\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

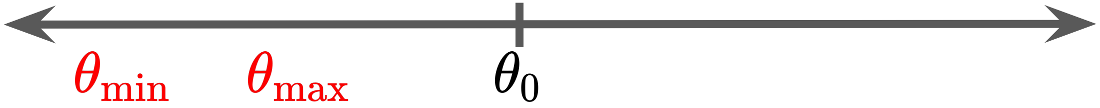


$$\frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2}$$

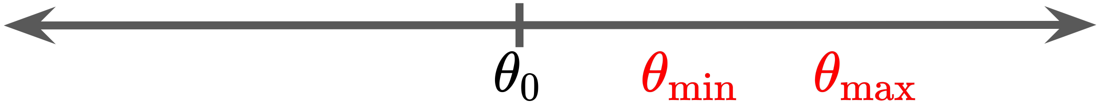
Simple Calculations

Suppose we have two estimators $\hat{\theta}^{(j)}, 1 \leq j \leq B,$ symmetric about θ_0

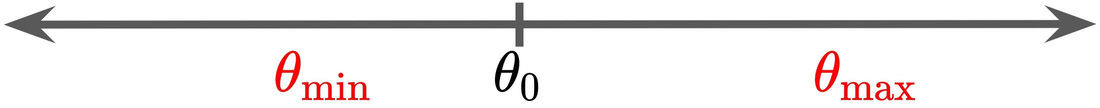
$$\theta_{\min} = \min\{\hat{\theta}^{(j)} : j \leq B\}, \quad \text{and} \quad \theta_{\max} = \max\{\hat{\theta}^{(j)} : j \leq B\}.$$



$$\frac{1}{2^B}$$



$$\frac{1}{2^B}$$



$$1 - \frac{2}{2^B}$$

General result

If

Median bias of the estimators.

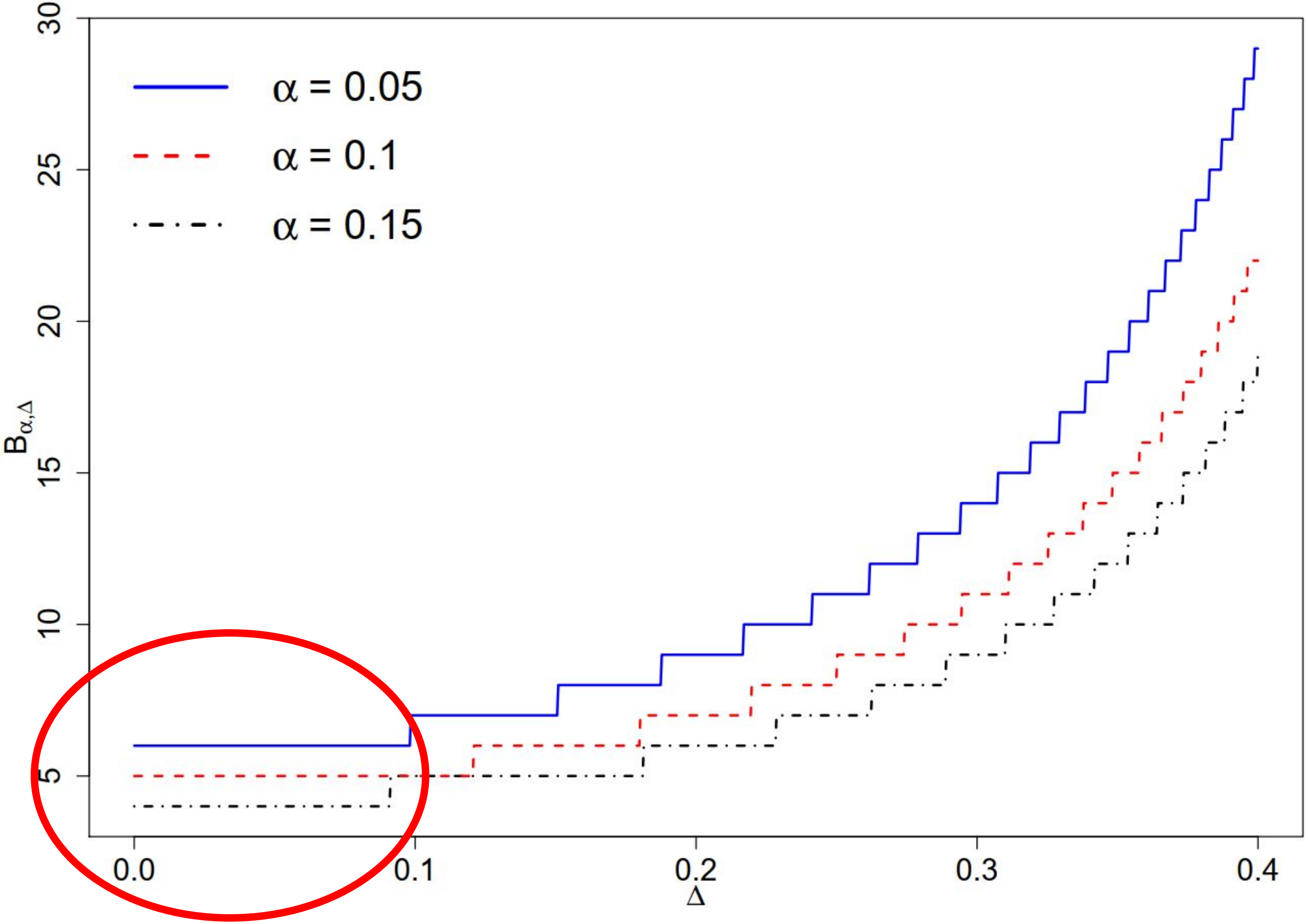
$$\Delta := \max_{j \geq 1} \left(\frac{1}{2} - \min \left\{ \mathbb{P}(\hat{\theta}_j \leq \theta_0), \mathbb{P}(\hat{\theta}_j \geq \theta_0) \right\} \right)_+ ,$$

then

$$\mathbb{P}\left(\theta_0 \notin [\theta_{\min}, \theta_{\max}]\right) \leq P(B; \Delta) = \left(\frac{1}{2} - \Delta\right)^B + \left(\frac{1}{2} + \Delta\right)^B .$$

Hence, $B = B_{\alpha,\Delta}$ with $P(B;\Delta) \leq \alpha$ many estimators to get a coverage of at least $1 - \alpha$.

Independent copies of the estimator can be obtained by splitting the data.



α = 0.05
α = 0.1
α = 0.15
30
25
20
15
10
5
$B_{\alpha,\Delta}$
0.0
0.1
0.2
0.3
0.4
Δ

Recall: pointwise asymptotics

If f_0 satisfies

$$\frac{|f_0(x) - f_0(x^\star)|}{|x - x^\star|^\beta} = A(1 + o(1)) \quad \text{as} \quad x \to x^\star,$$

for some $\beta \geq 1/2$, then

$$n^{\beta/(2\beta+1)}(\widehat{f}(x^\star) - f_0(x^\star)) \xrightarrow{d} \left(\frac{\sigma^2(x^\star)A^{1/\beta}}{h(x^\star)(\beta+1)^{1/\beta}}\right)^{\beta/(2\beta+1)} \mathbb{C}_\beta,$$

where $\mathbb{C}_\beta$ is the slope at zero of the greatest convex minorant of $B(t) + |t|^{\beta+1}$.

Simple Case

If β is known, then the median bias

$$\Delta := \max_{j\geq 1}\left(\frac{1}{2} - \min\left\{\mathbb{P}(\hat{\theta}_j \leq \theta_0), \mathbb{P}(\hat{\theta}_j \geq \theta_0)\right\}\right)_+,$$

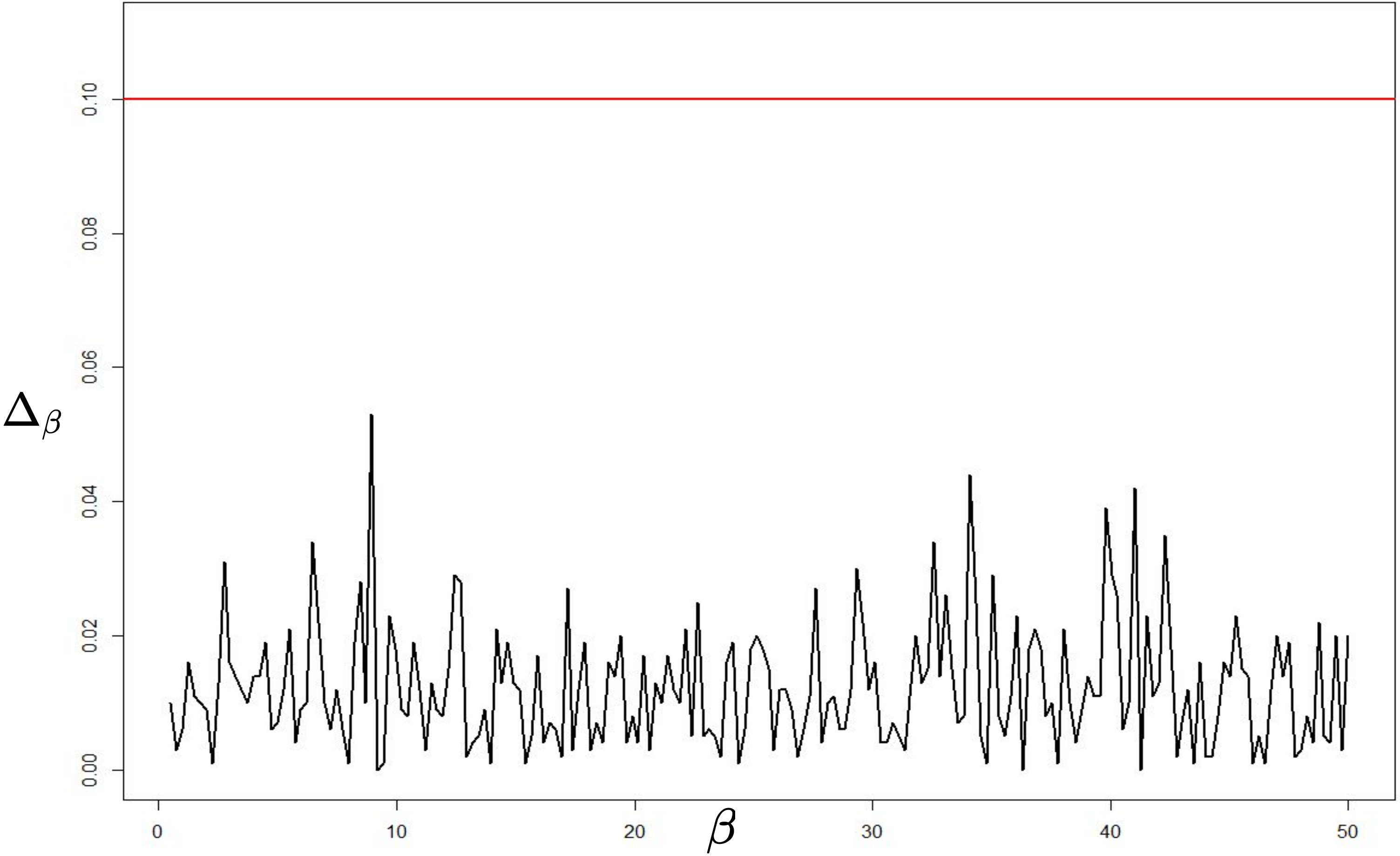
converges to

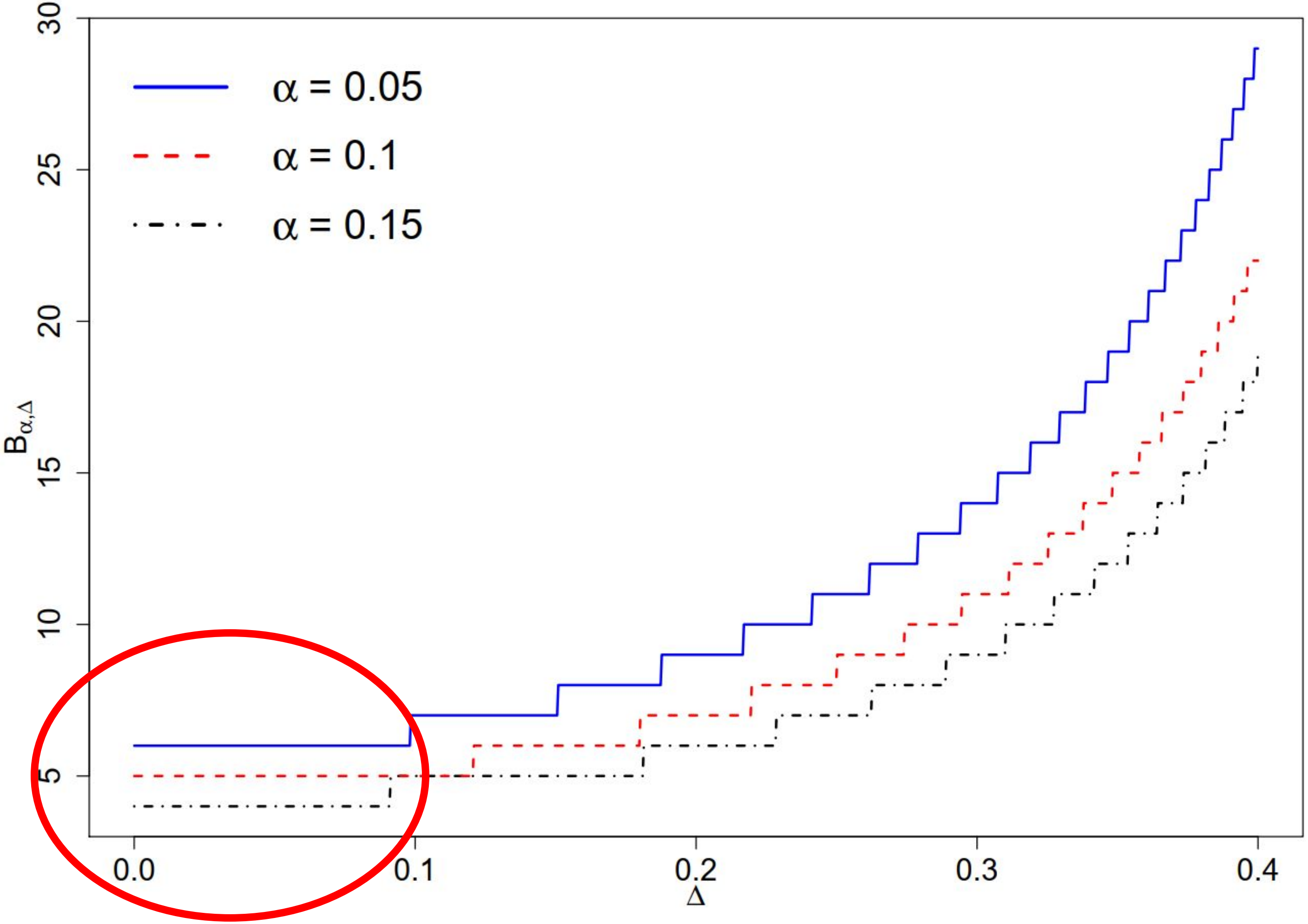
$$\Delta_\beta := \left|\frac{1}{2} - \mathbb{P}(\mathbb{C}_\beta \leq 0)\right|.$$

Hence, without the need to estimate the nuisance components, one can perform inference for isotonic regression at a point.

What if β is unknown?

Take Δ to be $\max\{\Delta_\beta : \beta \geq 1/2\}$





α = 0.05
α = 0.1
α = 0.15
30
25
20
15
10
5
0.0
0.1
0.2
0.3
0.4
Δ
$B_{\alpha,\Delta}$

Adaptive HulC

Adaptive HulC

$$\Delta := \max_{j\geq 1}\left(\tfrac{1}{2} - \min\left\{\mathbb{P}(\hat{\theta}_j \leq \theta_0),\, \mathbb{P}(\hat{\theta}_j \geq \theta_0)\right\}\right)_+,$$

can be estimated using subsampling:

b is the subsample size.

$$\widehat{\Delta} := \left|\frac{1}{2} - \frac{1}{K_n}\sum_{j=1}^{K_n}\mathbf{1}\{\hat{\theta}_j^{(b)} \leq \hat{\theta}\}\right|.$$

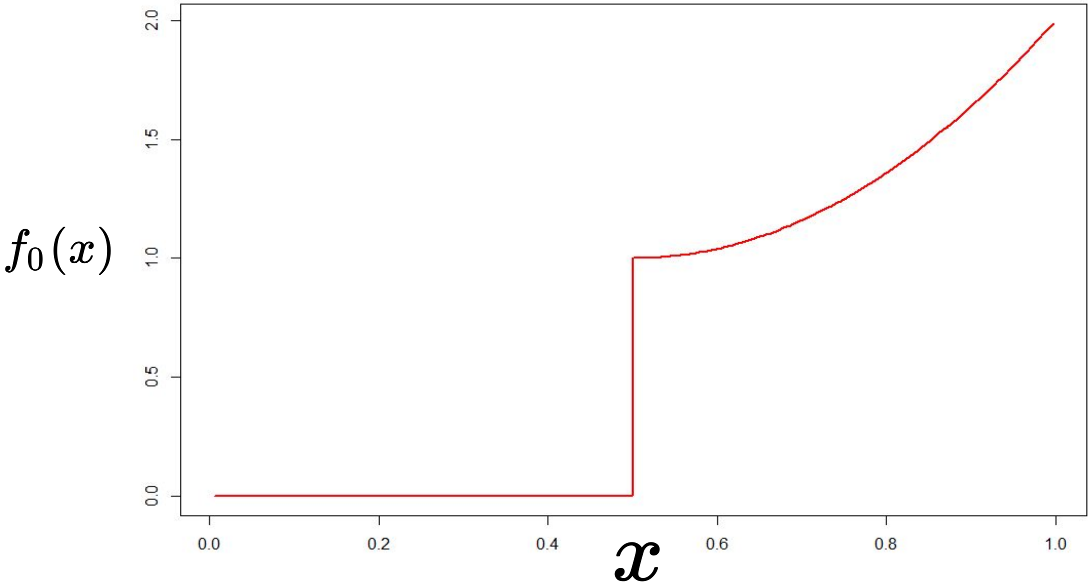
This does not require knowing the rate of convergence of the estimator, while the traditional application of subsampling requires such knowledge.

Numerical Examples

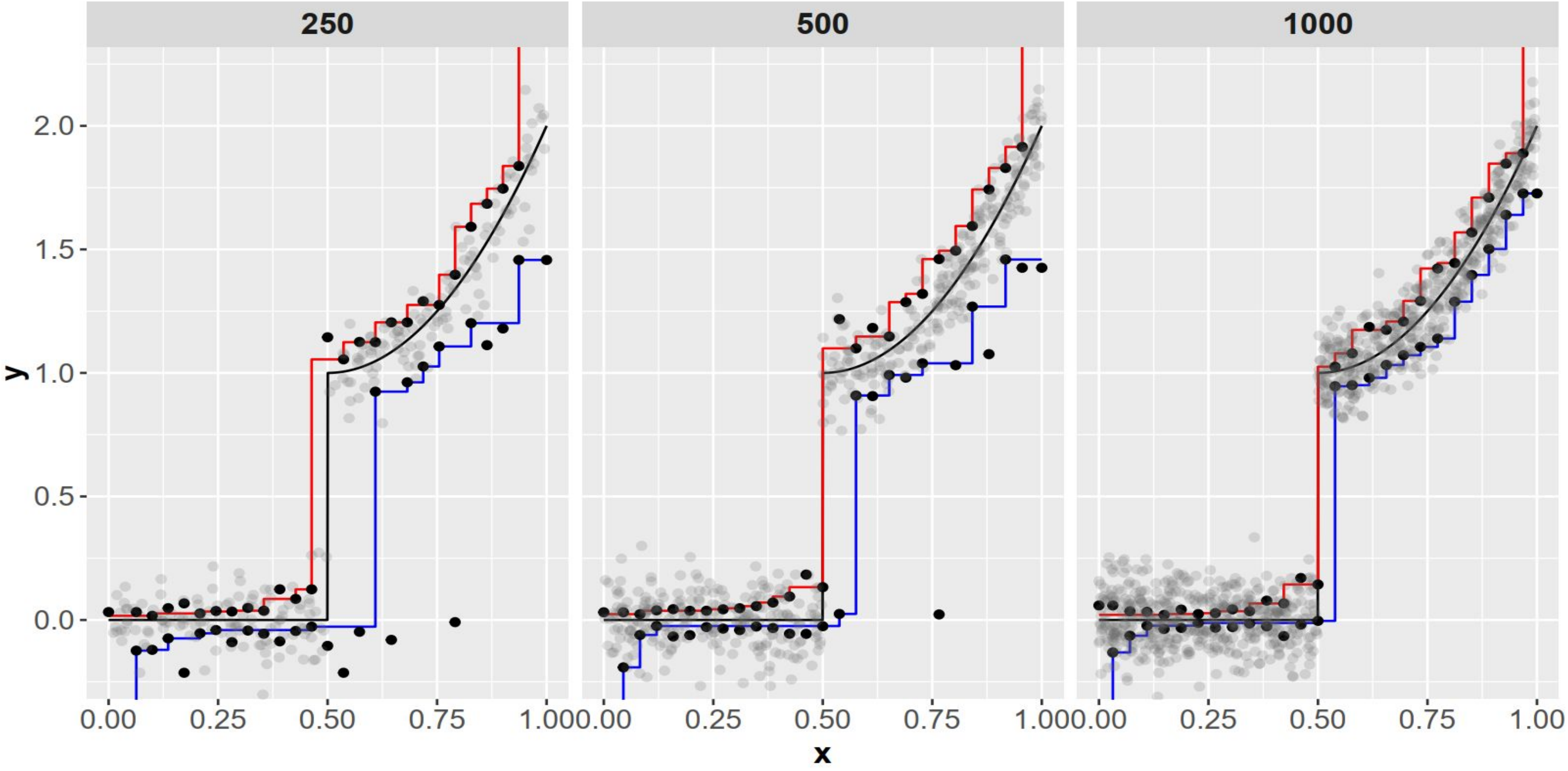
Simulation model

$$Y_i = f_0(X_i) + \xi_i$$

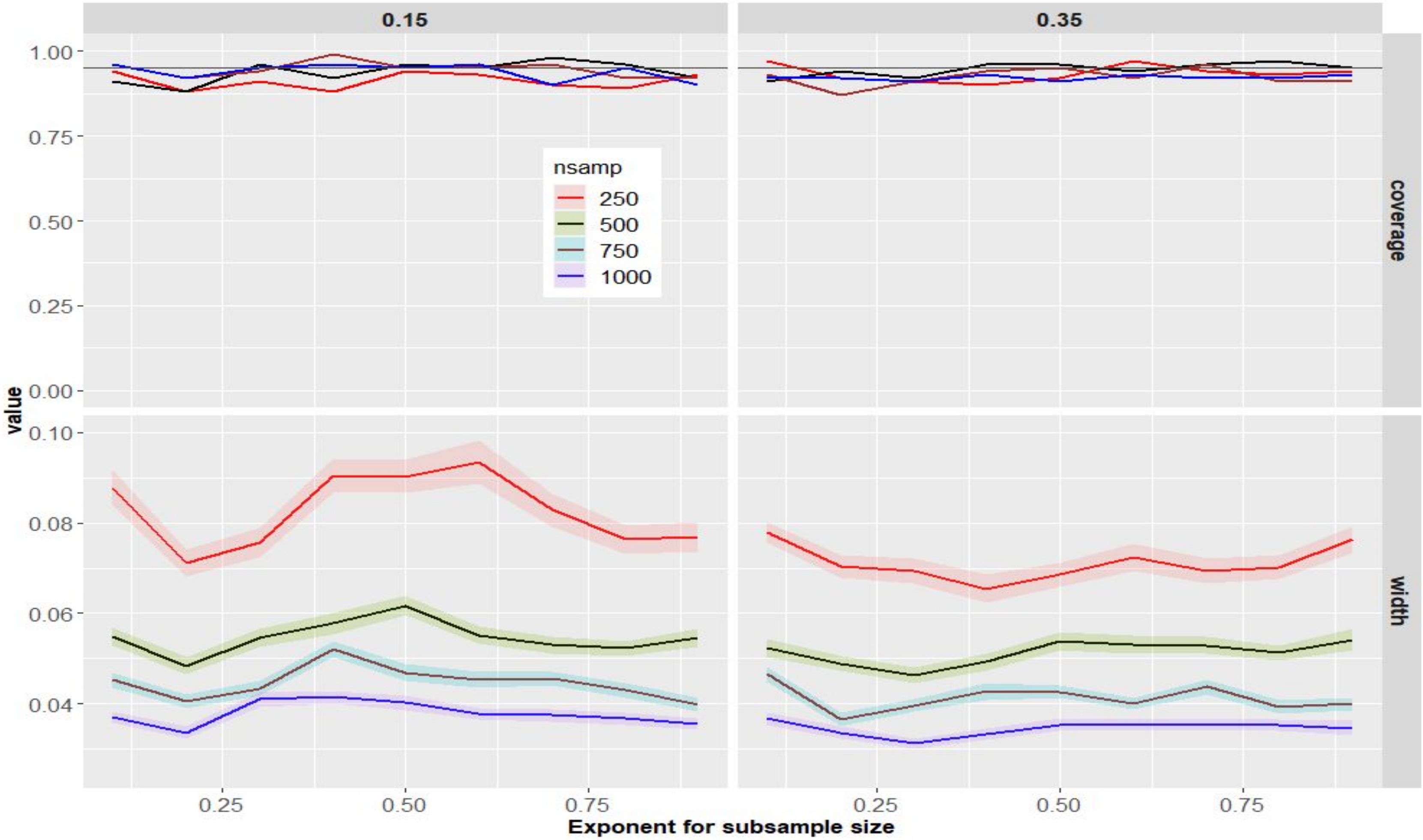
$$X_i \sim U[0,1] \quad \text{and} \quad \xi_i \sim N(0, 0.1^2).$$



Adaptive HulC



Effect of subsample size



Summary

- Inference in shape-constrained regression raises some interesting problems because traditional methods fall short.
- HulC can help solving these problems and much more.
- HulC only requires the estimation procedure and the *rough* knowledge of asymptotic median bias.
- Adaptive HulC estimates the asymptotic median bias using subsampling and the resulting method appears robust to the subsample size.

Summary (cont.)

The HulC applies to many more nonparametric examples without the requirement of undersmoothing. It provides pointwise inference for

- conditional mean, (Li and Racine, 2004, Stat. Sin. & Racine and Li, 2004, J. Econom.)
- conditional probability density, (Hall, Racine, and Li, 2004, JASA)
- conditional mean with sparsity, (Hall, Li, and Racine, 2007, Rev Econ Stat)
- conditional quantiles, (Li, Lin, and Racine, 2013, J. Bus. Econ. Stat.)
- conditional distribution functions, (Li, Lin, and Racine, 2013, J. Bus. Econ. Stat.).

The HulC also applies to all the semiparametric and parametric problems where the limiting distribution is Gaussian and often involves several nuisance components e.g., convex single index model (K., Patra, and Sen, 2021, JASA).

Thank you!!