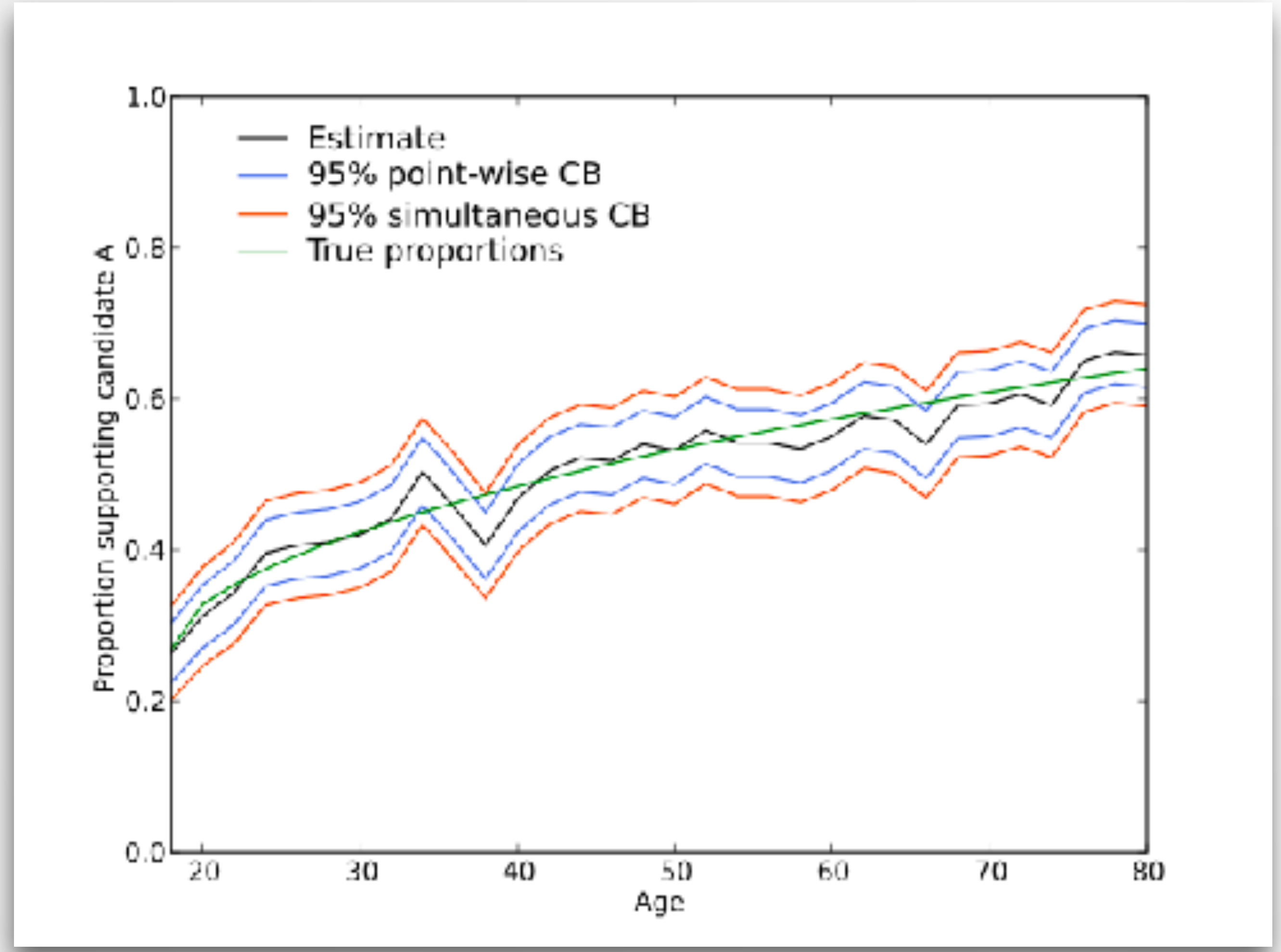
Uniform confidence bands for GRF estimates

Kainat Khowaja
Wolfgang Karl Härdle
Chen Huang

Humboldt-Universität zu Berlin
IRTG 1792
Ladislaus von Bortkiewicz Professor of Statistics
BRC Blockchain Research Center
Charles University, WISE XMU, NCTU 玉山学者

Simultaneous confidence intervals / UCB

- Summarise statistical uncertainty in both parametric and non parametric models
- Easy assess to statistical accuracy and perform various hypothesis tests about the function without access to the data



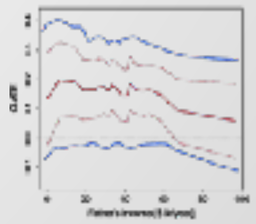
Source: Wikipedia

$\theta(x)$

$\hat{\theta}(x)$

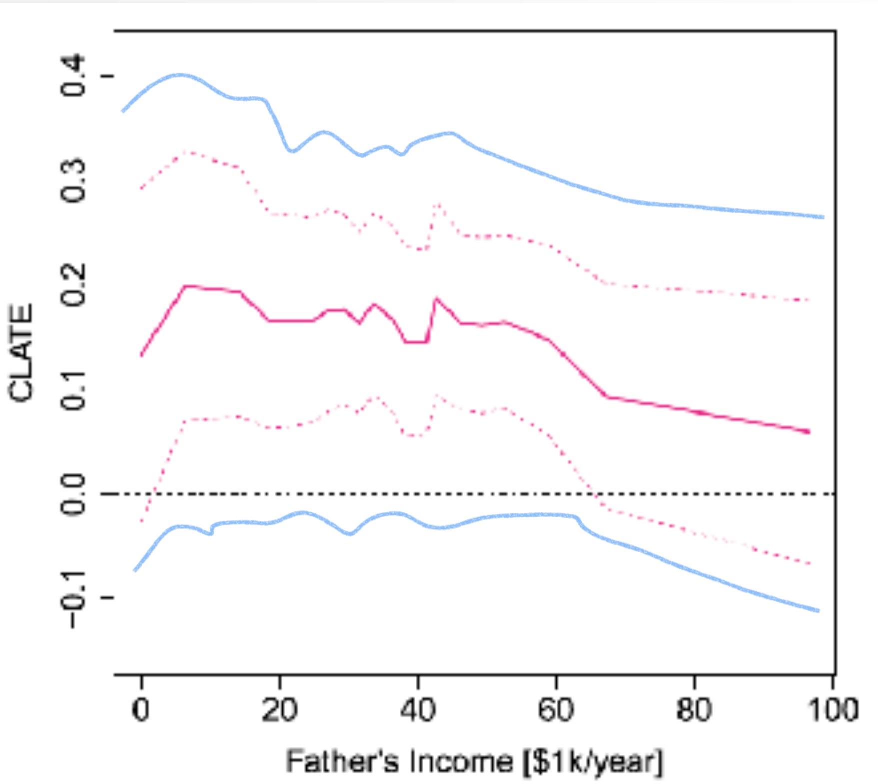
$\Pr(\hat{\theta}(x) - w(x) \leq \theta(x) \leq \hat{\theta}(x) + w(x)) = 1 - \alpha$

$\Pr(\hat{\theta}(x) - w(x) \leq \theta(x) \leq \hat{\theta}(x) + w(x) \forall x) = 1 - \alpha$



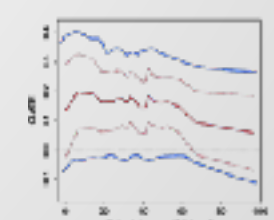
UCB for non parametric estimates

- Treatment effect: effect of having third child on labor force participation of mothers in the US in Athey et al. (2019)
- Covariate: father's income
- Conclusion: Observed treatment effect is driven by mothers whose husbands have a lower income



Mother 18 years old at first birth

Fig.: GRF estimates with pointwise 95% conf. int. that mother works for pay. A positive conditional local average treatment effect means that the treatment reduces the probability that the mother works.
Source: Fig. 3 in Athey et. al 2019

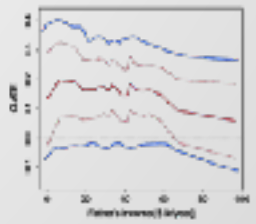


Relevant literature

- Horowitz et al. (2012) *Uniform confidence bands for functions estimated nonparametrically with instrumental variables*
  - UCB with bootstrap using max modulus method

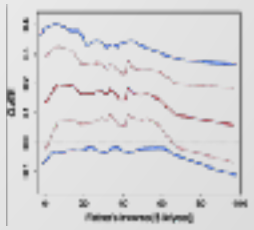
- Härdle et. al. (2013) *Tie the straps: uniform bootstrap confidence bands for bounded influence curve estimators*
  - UCB for a class of smoothers using asymptotic theory

- Härdle et. al. (2010) *Uniform confidence bands for pricing kernels*
  - *UCB using bootstrap for empirical pricing kernels*



Outline

1. Motivation ✓
2. Review of GRF
3. Uniform Confidence Bands
4. Numerical simulations
5. References



Problem setup

- Data: $\{(X_i, Y_i)\}_{i=1}^n \in \mathscr{X} \times \mathbb{R}$
- $Y_i = \theta(X_i) + \varepsilon_i$ where $\theta(x)$-unknown
- $\theta(x)$ obtained via

$$\arg\min_{\theta} \mathsf{E}\{\rho(Y_i - \theta) \,|\, X_i = x\}$$

- ρ is loss function. Eg. Quantile loss: $\rho(u) = |u(\tau - \mathbf{1}_{\{u \le 0\}})|$

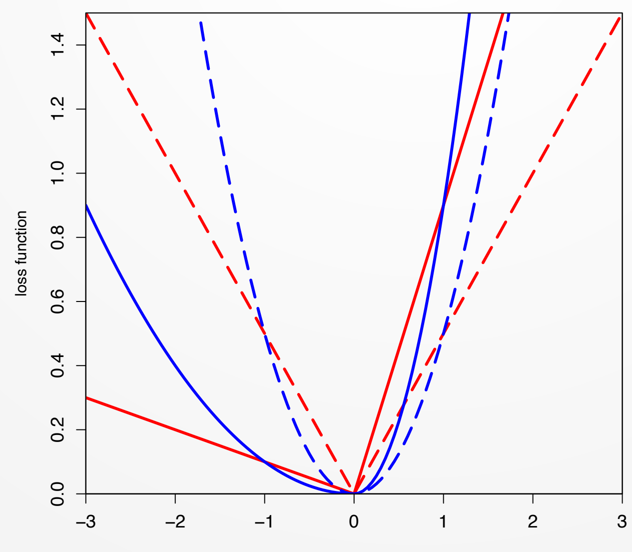
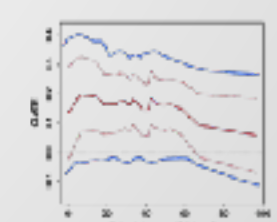


Figure: Loss function of expectiles and quantiles for $\tau = 0.5$ (dashed) and $\tau = 0.9$ (solid)

LQRcheck



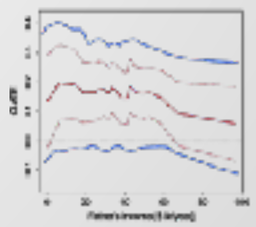
Estimation via GRF

GRF vs RF - 1st diff

- Not an average of predictions across ensemble of different trees
- $\alpha_i \succ$ similarity weights
- Estimation relies on local solutions
- Measure the relevance of i-th training example in fitting $\theta(\,.\,)$ at x
- Estimation of parameter (any parameter identified via local moment condition) via

$$\hat{\theta}(x) = \arg\min_{\theta} \sum_{i=1}^{n} \alpha_i(x)\rho(Y_i - \theta)$$

GRF vs RF - 2nd diff



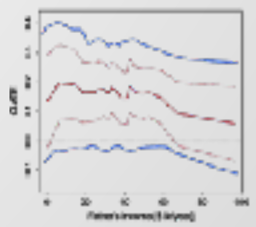
Effective weights

- Given a set of trees, indexed by $b = 1,\ldots,B$ define
  - the set of training samples in the same leaf as x in tree b by $L_b(x)$
  - the frequency that the i-th training sample falls into the same leaf as x by

$$\alpha_{bi}(x) = \frac{\mathbf{I}(\{X_i \in L_b(x)\})}{|L_b(x)|}$$

  - the forest-based adaptive neighbourhood of x for the i-th training sample by

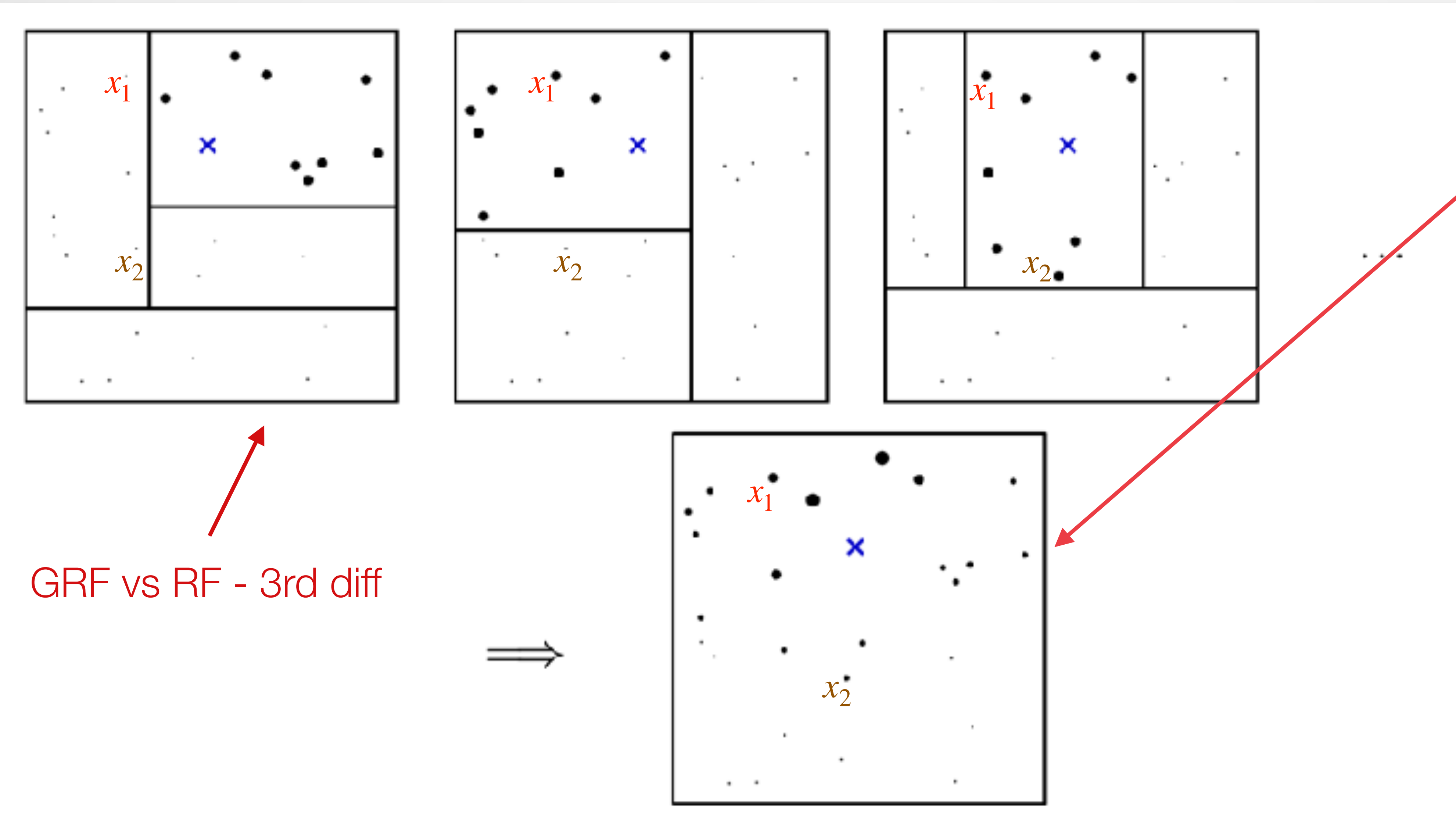
$$\alpha_i(x) = B^{-1}\sum_{b=1}^{B} \alpha_{bi}(x)$$



Effective weights

$\alpha_{11}(x) = 0/8$ $\alpha_{21}(x) = 1/8$ $\alpha_{31}(x) = 1/8$

$\alpha_{12}(x) = 0/8$ $\alpha_{22}(x) = 0/8$ $\alpha_{32}(x) = 1/8$

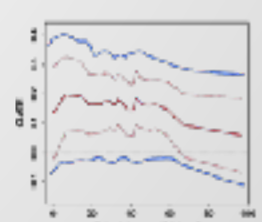


$$\alpha_1(x) = \frac{1}{3}\left(\frac{0}{8} + \frac{1}{8} + \frac{1}{8}\right)$$

$$\alpha_2(x) = \frac{1}{3}\left(\frac{0}{8} + \frac{0}{8} + \frac{1}{8}\right)$$

$$\vdots$$

Fig.: Illustration of the random forest weighting function. Each tree starts by giving equal (positive) weight to the training examples in the same leaf as our test point x of interest, and zero weight to all the other training examples. Then the forest averages all these tree-based weightings, and effectively measures how often each training example falls into the same leaf as x.
Source: Fig. 1 in Athey et. al 2019



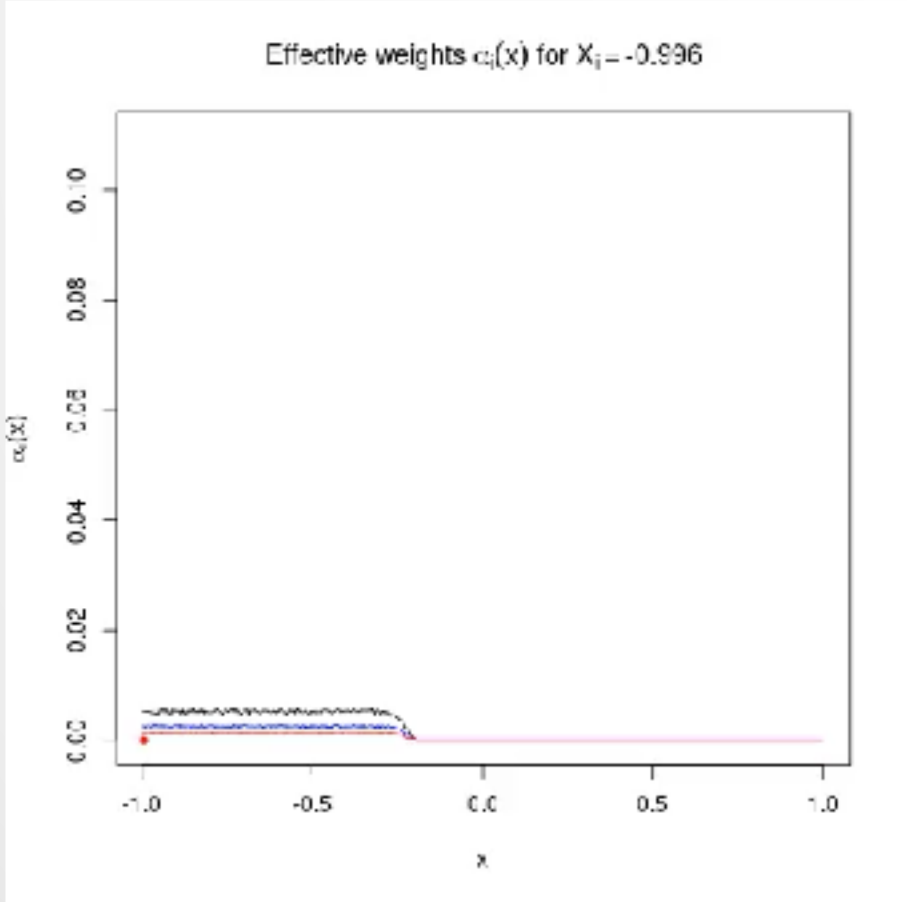
Example

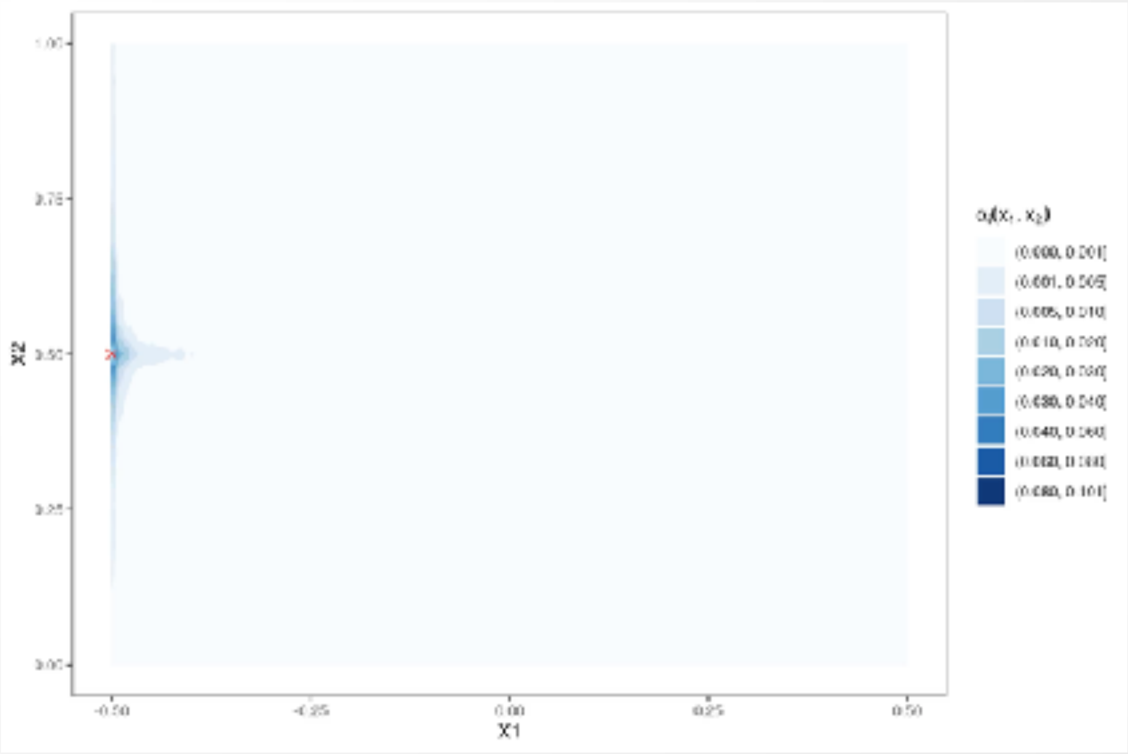
$n = 500, 1000, 2000$
$X_i = -1 + 2i/n,\ i = 1,\ldots,n$
$\theta(x_1, x_2) = \max(0, 1 - |x_1|/\eta),\ \eta = 0.2$
$Y_i = \theta(X_i) + \varepsilon_i,\ \varepsilon_i \sim N(0, \sigma_\varepsilon^2)$

$\alpha_i(x)$ based on RF trained on $(X_i, Y_i)_i$

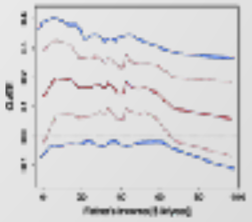
$\sigma_\varepsilon = 0$

eff. weights at $x_i = (0, 0.5)$ $\sigma_\varepsilon = 0.1$





GRF_effective_weights2D

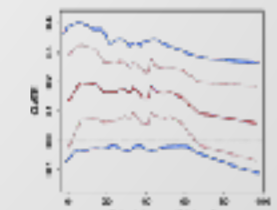


Asymptotic Behaviour /CLT for GRF estimates

- GRF estimator $\hat{\theta}(x)$ according to specifications and conditions
  - Then $\sqrt{(n)}\left(\dfrac{\hat{\theta}_n(x) - \theta(x)}{\sigma_n(x)}\right) \xrightarrow[n\to\infty]{\mathscr{L}} \mathcal{N}(0,1)$

- CIs for $\hat{\theta}(x)$:
  - Given consistent estimator of $\hat{\sigma}(x)$
  - Thm 5 in Athey et al + Slutzky's Lemma

$$P\left(\theta(x) \in \left[\hat{\theta}_n(x) \pm \Phi^{-1}(1-\alpha/2)\hat{\sigma}_n(x)\right]\right) \xrightarrow[n\to\infty]{} 1-\alpha$$



Variance of a forest

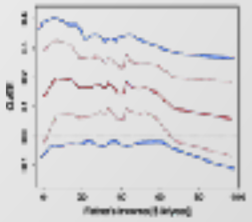
- Estimator for $\hat{\sigma}_n(x)$

$$\hat{\sigma}_n(x)^2 = \xi^\top \hat{V}(x)^{-1} \hat{H}(x) \{ \hat{V}(x)^{-1} \}^\top \xi$$

  - ξ is a vector that picks out the θ-coordinate.
- $V(x)$ is problem specific curvature parameter

$$V = \partial_\theta \, \mathrm{E}\{\psi_\theta(Y_i)\} \,|_{\theta=\theta_0}$$

- $H(x)$ is inner variance / sandwiched estimator
  - Bootstrap of little bags
  - $\widehat{H}$ can be expressed as between group and within group variance terms
  - Use ANOVA decomposition to consistently estimate the sampling variance

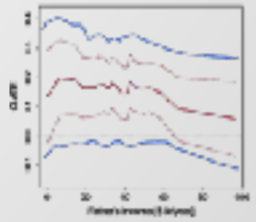


Uniform inference

- Check estimate $\hat{\theta}(x)$, as function of x, is uniformly converging to $\theta(x)$
- Given a data set at hand one needs to find a critical value C_α such that

$$P\left(\sup_x \frac{|\hat{\theta}(x) - \theta(x)|}{\sigma_n(x)} \leq C_\alpha\right) \xrightarrow[n\to\infty]{} 1 - \alpha$$

- $\widehat{\theta}(x) - \theta(x)/\sigma(x)$ is a gaussian process and EVT says the limit distribution of sup of it is Gumbel type
- Problem: Very slow asymptotic $\dfrac{1}{\log n_\theta}$
- Idea: Multiplier bootstrap for the critical value

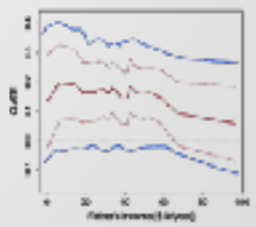


Multiplier Bootstrap

- Test statistic

$$T_n^*(x) = \sum_{i=1}^{n} \alpha_i(x) e_i \hat{\varepsilon}_i(x) \hat{\sigma}_n(x)^{-1},$$

- $e_i \sim_{i.i.d.} \mathrm{N}(0,1)$ -wild bootstrap multipliers, independent of O_i .
- $\hat{\varepsilon}_i(x) = -\xi^\top \hat{V}(x)^{-1} \psi_{\hat{\theta}(x)}(O_i),$
- $\hat{\sigma}_n(x)^2 = \xi^\top \hat{V}(x)^{-1} \hat{H}(x) \{\hat{V}(x)^{-1}\}^\top \xi,$
- $\hat{H}(x) = \mathrm{Var} \sum_{i=1}^{n} \alpha_i(x) \psi_{\hat{\theta}(x)}(O_i).$

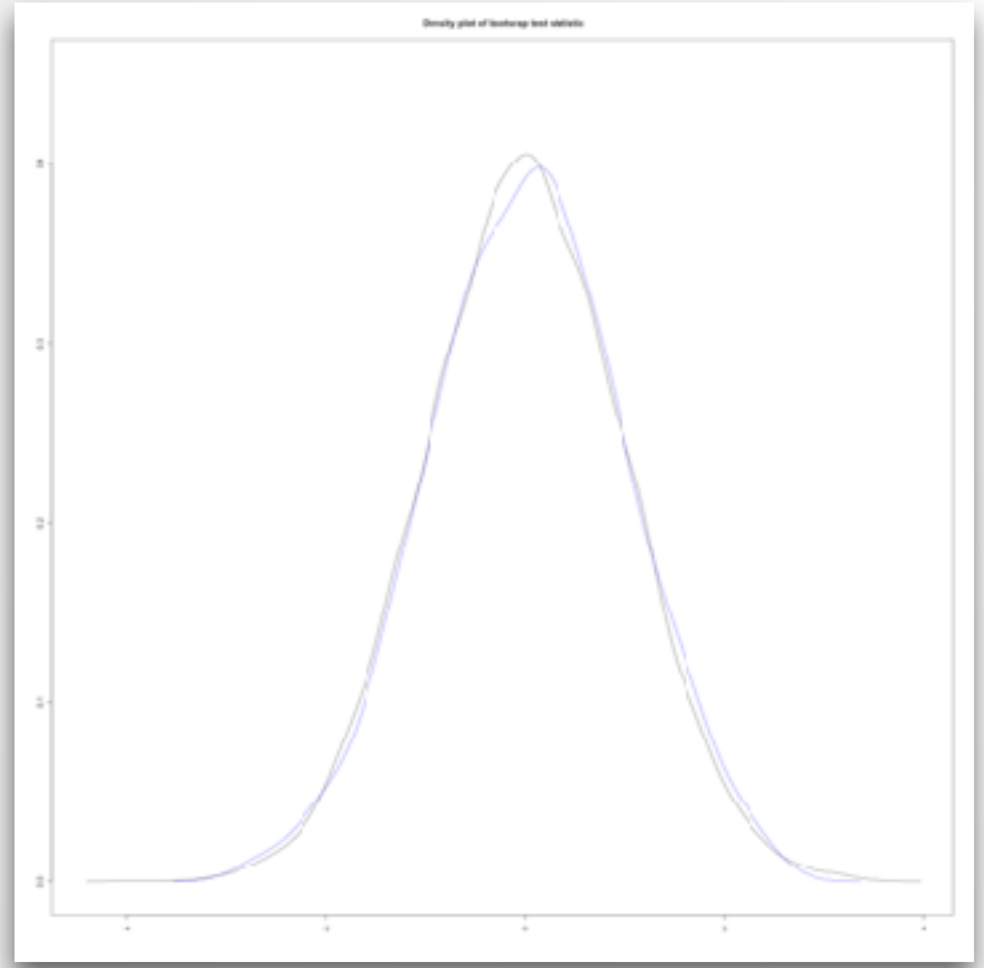


Distribution of test statistic

$x \in [-0.5, 0.5]$

$y = 1 + x + 2x^2 + 3x^3 + \varepsilon, \varepsilon \sim N(0,\sigma)$, regression forest

n=500, σ=1, h = 3

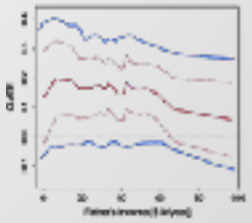


Bootstrap test stat KDE and asymptotic normal KDE for a single x (x = -0.5) in

Squared error for the single point (x= -0.5) averaged over replications (MSE) = 0.06799

θ = 0.625

E($\hat{\theta}$) = 0.612



Simultaneous confidence intervals: Maximum Modulus Method (Don Andrews)

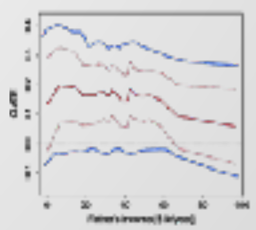
- $T_1, \ldots, T_k \succ$ set of pairwise uncorrelated rvs with the distribution t_ν
- $U \stackrel{\text{def}}{=} \max \left\{ \left| T_j \right|, j = 1, \ldots, k \right\}$ with Studentized Maximum Modulus dist $u_{k,\nu}$
- Let the $100(1-\alpha)$-percentile of the distribution be denoted $u^{\alpha}_{k,\nu}$.

$$P\left(\left| T_j \right| \le u^{\alpha}_{k,n-r}, \forall j = 1, \ldots, k \right) = P\left(\max \left\{ \left| T_j \right|, j = 1, \ldots, k \right\} \le u^{\alpha}_{k,n-r} \right) = 1 - \alpha$$

and a set of simultaneous confidence intervals is

$$I_j = \left[\hat{\theta}_j - u^{\alpha}_{k,n-r} \hat{\sigma}\left(\hat{\theta}_j\right), \hat{\theta}_j + u^{\alpha}_{k,n-r} \hat{\sigma}\left(\hat{\theta}_j\right) \right]$$

Hence, $P\left(\theta \in I_j, \forall j = 1, \ldots, k \right) = 1 - \alpha$



DGP

$n = 500, \quad \sigma = 1,$ reps = 100

$x \in [0,1.5],\ y = sin(x) + \varepsilon\ ,\ \varepsilon \sim N(0,\sigma)$

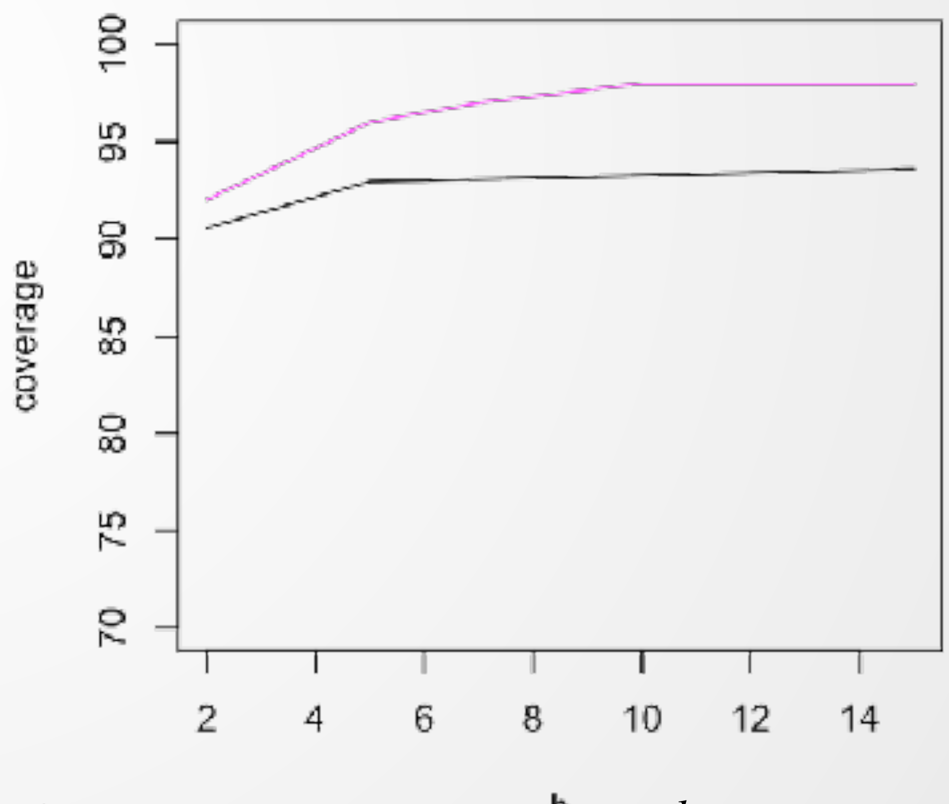
$$\hat{\theta}(x) = \frac{1}{2} \arg\min_{\theta} \sum_{i=1}^{n} \alpha_i(x)\left(Y_i - \theta\right)^2,\ V(x) = -1$$

$$\tilde{\varepsilon}_i(x) = \psi_{\hat{\theta}(x)}\left(Y_i\right) = Y_i - \hat{\theta}(x), \hat{H}(x) = \mathrm{Var}\sum_{i=1}^{n} \alpha_i(x)\left(Y_i - \hat{\theta}(x)\right), \hat{\sigma}_n(x)^2 = \hat{H}(x)$$

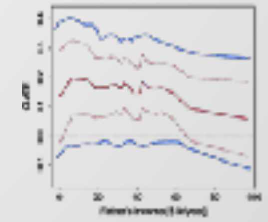
$T_n^*(x) = \sum_{i=1}^{n} \alpha_i(x) e_i \hat{\varepsilon}_i(x) \hat{\sigma}_n(x)^{-1}$, where $e_i \sim_{i.i.d.} \mathrm{N}(0,1)$

$\alpha_i(x)$ based on RF trained on $(X_i, Y_i)_i$

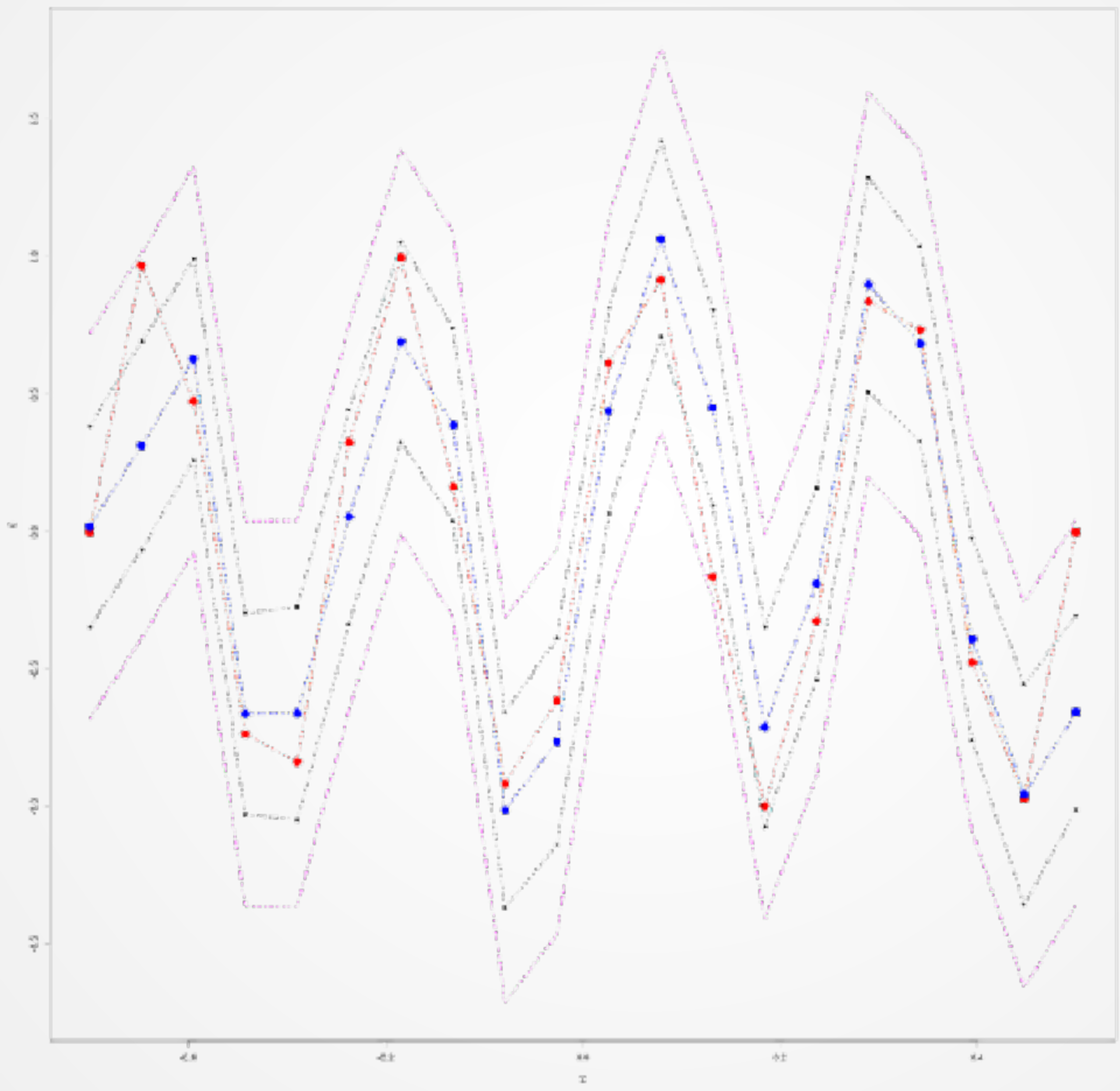
Coverage of $\theta(x)$: $\frac{1}{20 * J} \sum_{r=1}^{20} \sum_{j=1}^{J} \mathbf{1}\{\theta(x_j) \in CI_r(x_j)\}$



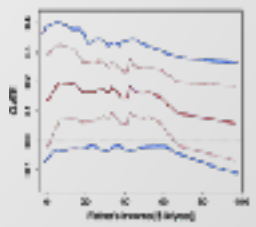
Coverage with changing bandwidth h for point wise confidence intervals and uniform confidence bands for bootstrap test statistic



$\sin(8\pi x)$

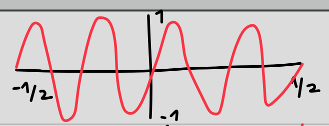
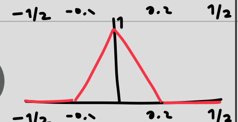
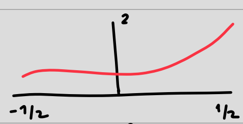
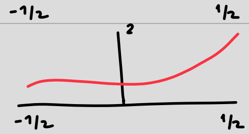
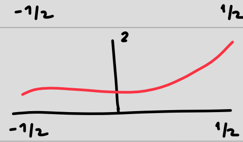


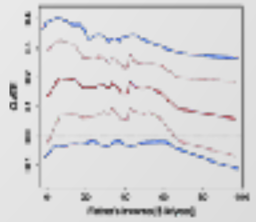
True theta, estimated theta, pointwise confidence intervals for bootstrap test statistic and uniform confidence bands GRF_effective_weights2D_bootstrap



Coverages

Monte Carlo experiments illustrate the finite-sample performance of the uniform confidence band.

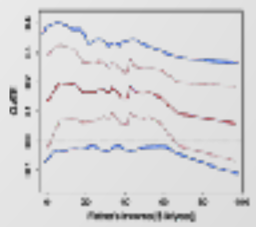
Function	h	Pointwise		Uniform	
		boostrap	asymptotic	bootstrap	asymptotic
$\sin(8\pi x)$ 	5	89	89.25	80	80
$\sin(5\pi x)$ 	5	92.75	92.75	95	100
$\sin(3\pi x)$ 	5	92	92	95	95
$\max(0,1-\lvert x_1\rvert/0.2)$ 	5	92.75	92	95	95
$\max(0,1-\lvert x_1\rvert/0.2)$ 	3	92.5	92.75	100	90
$\max(0,1-\lvert x_1\rvert/0.2)$ 	1	91.75	91.75	90	85
$1+x+2x^2+3x^3$ 	5	92	92.5	95	95
$1+x+2x^2+3x^3$ 	3	92.25	92.25	90	95
$1+x+2x^2+3x^3$ 	1	92	92	85	80



Size performance

Case	n	Grids	h	Sigma	Pointwise		Uniform	
					bootstrap	asymptotic	bootstrap	Asymptotic
1	500	20	5	1	0.075	0.0775	0	0.05
2	500	20	3	1	0.0775	0.0725	0	0.1
3	500	20	2	1	0.075	0.075	0.1	0.05
4	500	20	1	1	0.0825	0.0825	0.1	0.15
5	500	20	5	0.5	0.075	0.0775	0	0.05
6	500	20	3	0.5	0.0775	0.725	0	0.1
7	500	20	1	0.5	0.0825	0.0825	0.1	0.15
8	500	20	1	0.1	0.0825	0.0825	0.1	0.15
9	500	50	5	0.1	0.085	0.083	0.1	0.1
10	500	50	3	0.1	0.103	0.098	0.1	0.15
11	500	50	1	0.1	0.123	0.113	0.2	0.35
12	500	10	5	0.1	0.09	0.09	0.05	0.05
13	500	10	3	0.1	0.1	0.09	0.1	0.05
14	1000	20	1	0.1	0.1025	0.1	0.15	0.15
15	1000	20	1	10	0.1025	0.1	0.15	0.15
16	1000	20	10	1	0.07	0.07	0	0
17	1000	20	15	1	0.065	0.065	0	0
18	1000	50	15	1	0.084	0.083	0.1	0
19	1000	10	1	1	0.08	0.075	0.05	0.05
20	1000	5	1	1	0.18	0.18	0.15	0.1
21	500	5	1	1	0.17	0.17	0.15	0.15

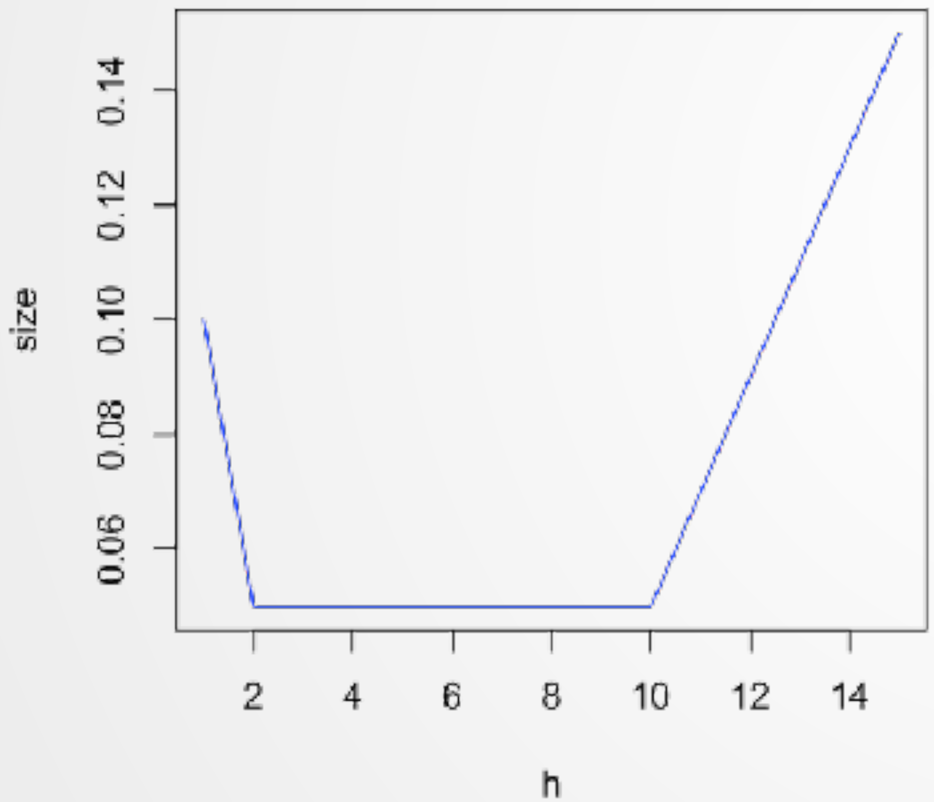
green = bootstrap test statistic size is lower or same as asymptotic normal



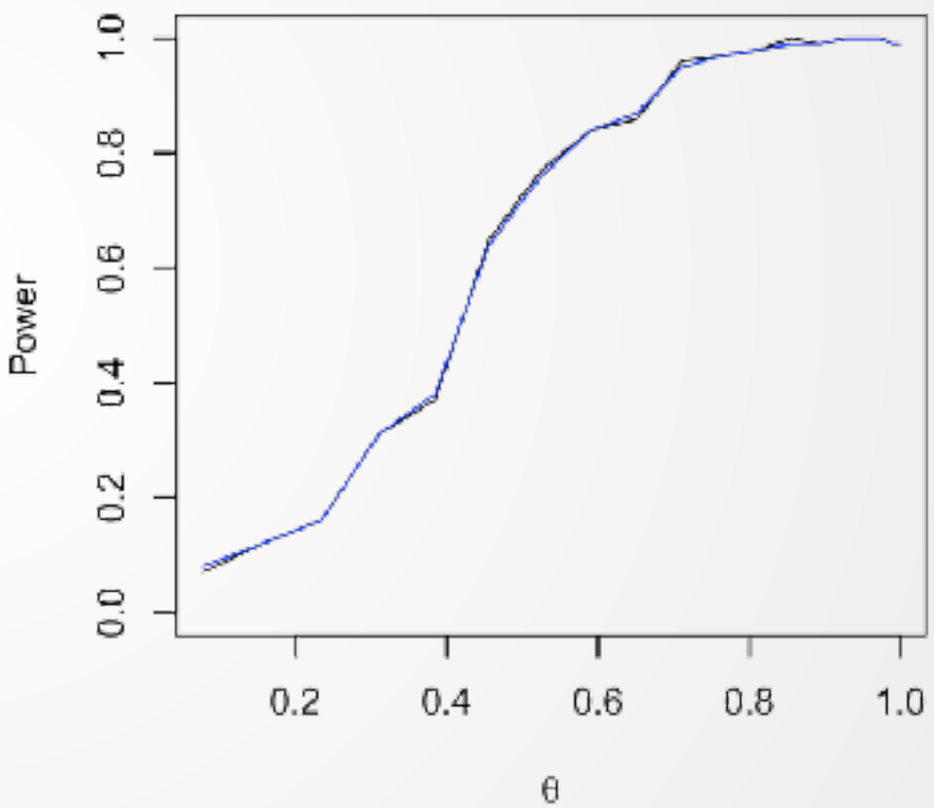
Size performance and power

$$\text{Size Performance } = \Pr\left(\text{type I error} \mid \theta_0 = 0\right)$$

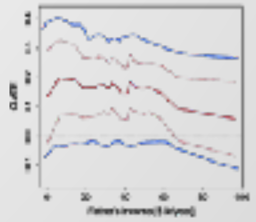
$$\text{Power curve } = 1 - \Pr\left(\text{type II error} \mid \theta_j \neq 0\right), j = 1,\ldots,J$$



Power of bootstrap test statistic and asymptotic distribution



Pointwise Power of bootstrap test statistic and asymptotic distribution



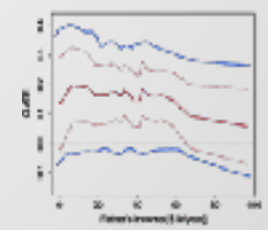
Important Literature

Härdle WK, Marron JS (1991)
Bootstrap Simultaneous Error Bars for Nonparametric Regression
Ann. Statist. 19(2): 778-796 (June, 1991). DOI: 10.1214/aos/1176348120

Athey S, Tibshirani J, Wager S (2019)
Generalized Random Forests
Annals of Statistics, Vol. 47(2), 1148-1178 , DOI: 10.1214/18-AOS1709

Härdle WK, Song S (2010)
Confidence Bands in Quantile Regression
Econometric Theory, vol. 26, no. 4, 2010, pp. 1180–1200. JSTOR, www.jstor.org/stable/40800877. Accessed 11 Jan. 2021.

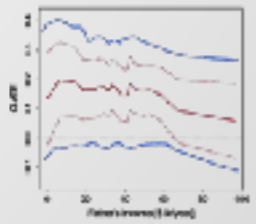
Meinshausen N (2006)
Quantile Regression Forests
J. Mach. Learn. Res. 7, 2006, pp. 983-999



Härdle WK, Ritov Y, Wang W (2015)
Tie the straps: Uniform bootstrap confidence bands for bounded influence curve estimators
J. Multivariate Analysis, 134, 129-145, doi: https://doi.org/10.1016/j.jmva.2014.11.003

Härdle WK, Okhrin Y, Wang W (2015) *Uniform confidence bands for pricing kernels*
J. Financial Econometrics, 13 (2): 376-413, DOI: https://doi.org/10.1093/jjfinec/nbu002

Horowitz JL, Lee S (2012) *Uniform confidence bands for functions estimated nonparametrically with instrumental variables.*
J Statistical Planning and Inference, 115, 521-542.



Uniform confidence bands for GRF estimates

Kainat Khowaja

Wolfgang Karl Härdle

Chen Huang

IRTG 1792
High Dimensional Nonstationary Time Series
Humboldt-Universität zu Berlin
IRTG1792.HU-Berlin.de