Inference on quantile processes with a finite number of clusters

Andreas Hagemann

Stephen M. Ross School of Business
University of Michigan

UNIVERSITY OF MICHIGAN

(Personal) motivation

- Uniform inference on quantile treatment effect in Project STAR

(Personal) motivation

- Uniform inference on quantile treatment effect in Project STAR
- Control units only, no effect, 5% nominal level, 16 clusters

	bootstrap	analytical methods		
	H ('17, JASA)	(1)	(2)	(3)
% reject	9.1	21.2	31.8	37.9

(Personal) motivation

- Uniform inference on quantile treatment effect in Project STAR

- Control units only, no effect, 5% nominal level, 16 clusters

	CRK H ('23, JoE)	bootstrap H ('17, JASA)	analytical methods (1)	(2)	(3)
% reject	**4.3**	9.1	21.2	31.8	37.9

- This presentation: **CRK** (cluster-randomized Kolmogorov) test

Contribution of this paper

- I establish validity of randomization inference with sign changes for quantile-like objects with finitely many large and arbitrarily heterogenous clusters

Contribution of this paper

- I establish validity of randomization inference with sign changes for quantile-like objects with finitely many large and arbitrarily heterogenous clusters

- Includes method to avoid matching clusters as in Canay, Romano, and Shaikh (2017, Ecma)

Contribution of this paper

- I establish validity of randomization inference with sign changes for quantile-like objects with finitely many large and arbitrarily heterogenous clusters

- Includes method to avoid matching clusters as in Canay, Romano, and Shaikh (2017, Ecma)

- Technical contribution: new results on randomization inference when limiting experiment is vector of heterogeneous Gaussian processes

Setup

- Parameter of interest is quantile or quantile-like scalar function $u \mapsto \delta(u)$

Setup

- Parameter of interest is quantile or quantile-like scalar function $u \mapsto \delta(u)$
- q clusters (villages, firms, states)

Setup

- Parameter of interest is quantile or quantile-like scalar function $u \mapsto \delta(u)$
- q clusters (villages, firms, states)
- δ is identified in each cluster separately (for now)

Setup

- Parameter of interest is quantile or quantile-like scalar function $u \mapsto \delta(u)$
- q clusters (villages, firms, states)
- δ is identified in each cluster separately (for now)
- q independent estimates $\hat{\delta} = (\hat{\delta}_1, \ldots, \hat{\delta}_q)$ of δ

Setup

- Parameter of interest is quantile or quantile-like scalar function $u \mapsto \delta(u)$
- q clusters (villages, firms, states)
- δ is identified in each cluster separately (for now)
- q independent estimates $\hat{\delta} = (\hat{\delta}_1, \ldots, \hat{\delta}_q)$ of δ
- Want to test $H_0 : \delta(u) = \delta_0(u)$ for all $u \in \mathcal{U} \subset (0, 1)$, e.g., $\delta_0 \equiv 0$

Example

- Notation: experiment with $2q$ villages, first q of $2q$ get new technology, predetermined pairs $(j, j+q)$. View pair as cluster.

Example

- Notation: experiment with $2q$ villages, first q of $2q$ get new technology, predetermined pairs $(j, j+q)$. View pair as cluster.

- Each pair identifies QTE $u \mapsto \delta(u) = F^{-1}_{Y(1)}(u) - F^{-1}_{Y(0)}(u)$

Example

- Notation: experiment with $2q$ villages, first q of $2q$ get new technology, predetermined pairs $(j, j+q)$. View pair as cluster.

- Each pair identifies QTE $u \mapsto \delta(u) = F^{-1}_{Y(1)}(u) - F^{-1}_{Y(0)}(u)$

- Each pair provides separate estimate $u \mapsto \hat{\delta}_j(u) = \hat{F}^{-1}_{Y_j}(u) - \hat{F}^{-1}_{Y_{j+q}}(u)$ of QTE

Example

- Notation: experiment with $2q$ villages, first q of $2q$ get new technology, predetermined pairs $(j, j+q)$. View pair as cluster.
- Each pair identifies QTE $u \mapsto \delta(u) = F^{-1}_{Y(1)}(u) - F^{-1}_{Y(0)}(u)$
- Each pair provides separate estimate $u \mapsto \hat{\delta}_j(u) = \hat{F}^{-1}_{Y_j}(u) - \hat{F}^{-1}_{Y_{j+q}}(u)$ of QTE
- Same idea works if we run quantile regressions

Example

- Notation: experiment with $2q$ villages, first q of $2q$ get new technology, predetermined pairs $(j, j+q)$. View pair as cluster.
- Each pair identifies QTE $u \mapsto \delta(u) = F^{-1}_{Y(1)}(u) - F^{-1}_{Y(0)}(u)$
- Each pair provides separate estimate $u \mapsto \hat{\delta}_j(u) = \hat{F}^{-1}_{Y_j}(u) - \hat{F}^{-1}_{Y_{j+q}}(u)$ of QTE
- Same idea works if we run quantile regressions
- If q stays fixed but sample grows large, standard asymptotics don't apply

Example

- Notation: experiment with $2q$ villages, first q of $2q$ get new technology, predetermined pairs $(j, j+q)$. View pair as cluster.
- Each pair identifies QTE $u \mapsto \delta(u) = F^{-1}_{Y(1)}(u) - F^{-1}_{Y(0)}(u)$
- Each pair provides separate estimate $u \mapsto \hat{\delta}_j(u) = \hat{F}^{-1}_{Y_j}(u) - \hat{F}^{-1}_{Y_{j+q}}(u)$ of QTE
- Same idea works if we run quantile regressions
- If q stays fixed but sample grows large, standard asymptotics don't apply
- $\sqrt{n}(\hat{\delta} - \delta 1_q)$ converges to nice Gaussian process, covariances unknowable

Randomization inference with sign changes

- $g = (g_1, \ldots, g_q)$ is operation of attaching $+$ or $-$, and $\mathcal{G}$ set of all g

Randomization inference with sign changes

- $g = (g_1, \ldots, g_q)$ is operation of attaching $+$ or $-$, and $\mathcal{G}$ set of all g
- Vector of data $X = (X_1, \ldots, X_q)$, some statistic $T(X)$ like mean or t-stat

Randomization inference with sign changes

- $g = (g_1, \ldots, g_q)$ is operation of attaching $+$ or $-$, and $\mathcal{G}$ set of all g
- Vector of data $X = (X_1, \ldots, X_q)$, some statistic $T(X)$ like mean or t-stat
- Ordered values of $T(gX)$ over $\mathcal{G}$: $T^{(1)}(X, \mathcal{G}) \leq \cdots \leq T^{(|\mathcal{G}|)}(X, \mathcal{G})$.

Randomization inference with sign changes

- $g = (g_1, \ldots, g_q)$ is operation of attaching $+$ or $-$, and $\mathcal{G}$ set of all g
- Vector of data $X = (X_1, \ldots, X_q)$, some statistic $T(X)$ like mean or t-stat
- Ordered values of $T(gX)$ over $\mathcal{G}$: $T^{(1)}(X, \mathcal{G}) \leq \cdots \leq T^{(|\mathcal{G}|)}(X, \mathcal{G})$.
- Classical randomization hypothesis: if $X \sim gX$ for every $g \in \mathcal{G}$, then

$$\mathrm{P}\big(T(X) > T^{1-\alpha}(X, \mathcal{G})\big) \leq \alpha, \qquad \text{where } T^{1-\alpha}(X, \mathcal{G}) := T^{(|\mathcal{G}|(1-\alpha))}(X, \mathcal{G})$$

Randomization inference with sign changes

- $g = (g_1, \ldots, g_q)$ is operation of attaching $+$ or $-$, and $\mathcal{G}$ set of all g
- Vector of data $X = (X_1, \ldots, X_q)$, some statistic $T(X)$ like mean or t-stat
- Ordered values of $T(gX)$ over $\mathcal{G}$: $T^{(1)}(X, \mathcal{G}) \leq \cdots \leq T^{(|\mathcal{G}|)}(X, \mathcal{G})$.
- Classical randomization hypothesis: if $X \sim gX$ for every $g \in \mathcal{G}$, then

$$\mathrm{P}\big(T(X) > T^{1-\alpha}(X, \mathcal{G})\big) \leq \alpha, \qquad \text{where } T^{1-\alpha}(X, \mathcal{G}) := T^{(|\mathcal{G}|(1-\alpha))}(X, \mathcal{G})$$

- If X is a stochastic process, then $X \sim gX$ does not make sense. What now?

(AN) $\sqrt{n}(\hat{\delta} - \delta 1_q) \rightsquigarrow X = (X_1, \ldots, X_q)$ independent Gaussian processes on $\mathcal{U}$ with continuous sample paths

(AN) $\sqrt{n}(\hat{\delta} - \delta 1_q) \rightsquigarrow X = (X_1, \ldots, X_q)$ independent Gaussian processes on $\mathcal{U}$ with continuous sample paths

- Covariances unknown and not assumed to be estimable

(AN) $\sqrt{n}(\hat{\delta} - \delta 1_q) \rightsquigarrow X = (X_1, \ldots, X_q)$ independent Gaussian processes on $\mathcal{U}$ with continuous sample paths

- Covariances unknown and not assumed to be estimable
- $T(X) = \sup_{u \in \mathcal{U}} \frac{1}{q} \sum_{j=1}^{q} X_j(u)$. Have $X(u) \sim gX(u)$ at every $u \in \mathcal{U}$. Is that enough?

(AN) $\sqrt{n}(\hat{\delta} - \delta 1_q) \rightsquigarrow X = (X_1, \ldots, X_q)$ independent Gaussian processes on $\mathcal{U}$ with continuous sample paths

- Covariances unknown and not assumed to be estimable

- $T(X) = \sup_{u \in \mathcal{U}} \frac{1}{q} \sum_{j=1}^{q} X_j(u)$. Have $X(u) \sim gX(u)$ at every $u \in \mathcal{U}$. Is that enough?

- Yes! If X is smooth enough to rule out ties and T is a sup-statistic.

(AN) $\sqrt{n}(\hat{\delta} - \delta 1_q) \rightsquigarrow X = (X_1, \ldots, X_q)$ independent Gaussian processes on $\mathcal{U}$ with continuous sample paths

- Covariances unknown and not assumed to be estimable
- $T(X) = \sup_{u \in \mathcal{U}} \frac{1}{q} \sum_{j=1}^{q} X_j(u)$. Have $X(u) \sim gX(u)$ at every $u \in \mathcal{U}$. Is that enough?
- Yes! If X is smooth enough to rule out ties and T is a sup-statistic.

THEOREM (SIZE)

If **(AN)** holds and $\mathrm{P}(X_j(u) = -X_j(u')) = 0$ for all $u, u' \in \mathcal{U}$ and $1 \leq j \leq q$, then

$$\mathrm{P}\big(T(\hat{\delta} - \delta 1_q) > T^{1-\alpha}(\hat{\delta} - \delta 1_q, \mathcal{G})\big) \to \mathrm{P}\big(T(X) > T^{1-\alpha}(X, \mathcal{G})\big) \leq \alpha$$

What if effect is identified across clusters?

- Draw treated and control cluster matches so that each cluster is used only once. Randomize and repeat many times.

What if effect is identified across clusters?

- Draw treated and control cluster matches so that each cluster is used only once. Randomize and repeat many times.
- $T(X) > T^{1-\alpha}(X, \mathcal{G})$ if and only if p-value $p(X, \mathcal{G}) \leq \alpha$

What if effect is identified across clusters?

- Draw treated and control cluster matches so that each cluster is used only once. Randomize and repeat many times.
- $T(X) > T^{1-\alpha}(X, \mathcal{G})$ if and only if p-value $p(X, \mathcal{G}) \leq \alpha$
- Take average p-value $\bar{p}$, satisfies Rüschendorf's inequality asymptotically

What if effect is identified across clusters?

- Draw treated and control cluster matches so that each cluster is used only once. Randomize and repeat many times.
- $T(X) > T^{1-\alpha}(X, \mathcal{G})$ if and only if p-value $p(X, \mathcal{G}) \leq \alpha$
- Take average p-value $\bar{p}$, satisfies Rüschendorf's inequality asymptotically

THEOREM (SIZE)

Under slightly strenghted conditions,

$$\limsup_{n \to \infty} \mathrm{P}\big(\bar{p} \leq \alpha/2\big) \leq \alpha$$

What I didn't show you

- Test have local and global power, explicit lower bounds on local power

What I didn't show you

- Test have local and global power, explicit lower bounds on local power
- More complex quasi-experimental setups

What I didn't show you

- Test have local and global power, explicit lower bounds on local power
- More complex quasi-experimental setups
- What if # of treated and untreated clusters not equal?

What I didn't show you

- Test have local and global power, explicit lower bounds on local power
- More complex quasi-experimental setups
- What if # of treated and untreated clusters not equal?
- **Thank you!**