# EC708 Discussion 11
# Bootstrap

Yan Liu[1]

Department of Economics
Boston University

April 23, 2021

[1]Parts of the materials are from Horowitz (2001, 2019), lecture notes by Xiaoxia Shi, and textbook by Bruce Hansen (2021).

# Outline

1. Nonparametric Bootstrap
   - Asymptotic Refinement
   - Hypothesis Testing
   - Bootstrap Failure

2. Extensions
   - Wild Bootstrap
   - Block Bootstrap
   - Nonsmooth Statistics

# Outline

# What is Bootstrap?

- "The bootstrap is a method for estimating the distribution of an estimator or test statistic by resampling one's data or a model estimated from the data." (Horowitz, 2019)
- Bootstrap is an alternative to asymptotic-based inference, but not a substitute of asymptotic theory.

# Nonparametric Bootstrap

- A random sample $\{X_1, \ldots, X_T\}$ from distribution $P_0$.
- Statistic $T_T(X_1, \ldots, X_T)$ (estimator or test statistic)
- $J_T(\tau, P)$: exact finite-sample distirbution of $T_T$ when data are sampled from $P$.
- Nonparametric bootstrap approximates $J_T(\tau, P_0)$ by replacing $P_0$ with empirical CDF $\hat{P}_T$.

**Algorithm:**

1. Generate a bootstrap sample $\{X_t^*\}_{t=1}^T$ from original data randomly with replacement.
2. Compute $T_T^* = T_T(X_1^*, \ldots, X_T^*)$.
3. Use results of many repetitions of steps 1 and 2 to compute $J_T(\tau, \hat{P}_T)$.

# How Many Bootstrap Replications?

- Computation cost is essentially linear in $B$ while accuracy (standard errors or $p$-values) is proportional to $B^{-1/2}$.
- For daily quick and investigatory calculations, $B = 100$ may be sufficient for rough estimates.
- For final calculations, $B = 10,000$ is desired, with $B = 1000$ a minimal choice.
- Stata by default sets $B = 50$.

# Bootstrap Standard Errors

Denote $\bar{T}_{T,B} = \frac{1}{B}\sum_{b=1}^{B} T^*_{T,b}$. Simulated bootstrap standard error is

$$\sqrt{\frac{1}{B-1}\sum_{b=1}^{B}(T^*_{T,b} - \bar{T}_{T,B})(T^*_{T,b} - \bar{T}_{T,B})'}.$$

**Remarks:**

- Bootstrap variance is consistent for smooth functions with a bounded $p^{\text{th}}$ order derivative. Counterexample: $\theta = \mu_1/\mu_2$ where $\mu_i = E(y_i)$. Need to use a trimmed estimator by excluding tails.
- Bootstrap standard errors are desired if asymptotic variance involves nonparametric functions (e.g. sample quantile), whose estimators have slow convergence rate and are sensitive to bandwidth choices.
- Bootstrap standard errors are robust to misspecification or heteroskedasticity, and less susceptible to algebraic or coding errors.

# Asymptotic Refinement

Bootstrap can provide more accurate approximations to the distributions of statistics than does the conventional asymptotic distribution theory.

- First-order bias reduction of nonlinear estimators
- Higher-order refinements to rejection probabilities of tests and coverage probabilities of confidence intervals

# Bias Correction

- Nonlinear estimators are prone to finite-sample bias. Bootstrap offers a way to estimate the bias up to some asymptotic order.
- For any estimator $\hat{\theta}_T$, bootstrap bias corrected estimator is

$$\hat{\theta}_T - (E^*(\hat{\theta}_T^*) - \hat{\theta}_T) = 2\hat{\theta}_T - E^*(\hat{\theta}_T^*).$$

- Let's look at a concrete example.

# Bias Correction

- For a random vector $X$, consider $\mu = E[X]$ and $\theta = g(\mu)$ where $g$ is a twice continuously differentiable nonlinear function.
- For a random sample $\{X_t\}_{t=1}^T$, define $\bar{X}_T = \frac{1}{T}\sum_{t=1}^T X_t$. A consistent estimator is $\hat{\theta}_T = g(\bar{X}_T)$.
- $\hat{\theta}_T$ is biased: $E[\hat{\theta}_T] = E[g(\bar{X}_T)] \neq g(E[\bar{X}_T]) = g(\mu) = \theta$.
- Characterize the bias by a Taylor expansion:

$$E[\hat{\theta}_T - \theta] = \underbrace{\frac{1}{2}E\left[(\bar{X}_T - \mu)' G_2(\mu)(\bar{X}_T - \mu)\right]}_{\text{first-order bias } B_T} + O(T^{-2})$$

where $G_2$ is the matrix of second derivative of $g$.

# Bias Correction

- In the bootstrap world, the true parameter is $\hat{\theta}_T$.
- Denote the bootstrap estimator as $\hat{\theta}_T^* = g(\bar{X}_T^*)$ where $\bar{X}_T^* = \frac{1}{T}\sum_{t=1}^T X_t^*$ is the bootstrap sample mean.
- Taylor expansion is now

$$E^*[\hat{\theta}_T^* - \hat{\theta}_T] = \underbrace{\frac{1}{2}E^*\left[(\bar{X}_T^* - \bar{X}_T)' G_2(\bar{X}_T)(\bar{X}_T^* - \bar{X}_T)\right]}_{\text{first-order bootstrap bias } B_T^*} + O(T^{-2})$$

  $B_T^*$ can be computed with arbitrary accuracy by Monte Carlo simulation since we know the empirical distribution.
- Can show that $E(B_T^*) = B_T + O(T^{-2})$. Hence,

$$\hat{\theta}_{bc} = \hat{\theta}_T - B_T^* = 2\hat{\theta}_T - E^*(\hat{\theta}_T^*)$$

  satisfies $E(\hat{\theta}_{bc} - \theta) = O(T^{-2})$.

# Higher-order Refinement

Work with the statistic $T_T = \sqrt{T}[H(\bar{X}) - H(\mu)]$ where $H$ has sufficiently many derivatives and $X$ satisfies the Cramér condition. Cramér condition Recall

1. $J_T(\tau, P_0)$: exact finite-sample distribution of $T_T$;
2. $J_\infty(\tau, P_0)$: exact asymptotic distribution of $T_T$;
3. $J_T(\tau, \hat{P}_T)$: bootstrap distribution of $T_T$.

**Edgeworth expansion:**

$$J_T(\tau, P_0) = J_\infty(\tau, P_0) + \frac{1}{\sqrt{T}} g_1(\tau, \kappa_1) + \frac{1}{T} g_2(\tau, \kappa_2) + \frac{1}{\sqrt{T^3}} g_3(\tau, \kappa_3) + O(T^{-2})$$

$$J_T(\tau, \hat{P}_T) = J_\infty(\tau, \hat{P}_T) + \frac{1}{\sqrt{T}} g_1(\tau, \kappa_{T1}) + \frac{1}{T} g_2(\tau, \kappa_{T2}) + \frac{1}{\sqrt{T^3}} g_3(\tau, \kappa_{T3}) + O(T^{-2})$$

- Leading term in the difference is $J_\infty(\tau, P_0) - J_\infty(\tau, \hat{P}_T) = O(T^{-1/2})$
  $\Rightarrow$ same accuracy as conventional approximations
- If $T_T$ is pivotal, leading term is $\frac{1}{\sqrt{T}} g_1(\tau, \kappa_1) - \frac{1}{\sqrt{T}} g_1(\tau, \kappa_{T1}) = O(T^{-1})$
  $\Rightarrow$ more accurate than conventional approximations

# Bootstrap Critical Values for Hypothesis Testing

- Suppose $T_T$ is an asymptotically pivotal test statistic
- Let $q_{1-\alpha}$ denote the $1-\alpha$ quantile of distribution of $|T_T|$, then $P(|T_T| \leq q_{1-\alpha}) = 1-\alpha$.
- $q_{1-\alpha}$ is unknown in most settings. Let $q^*_{1-\alpha}$ be the $1-\alpha$ quantile of bootstrap distribution of $|T^*_T|$.
- Edgeworth expansion of distribution of $|T_T| - q^*_{1-\alpha}$ shows that under the null,

$$P(|T_T| > q^*_{1-\alpha}) - \alpha = O(T^{-2}).$$

- In contrast, for asymptotic critical value $z_{1-\alpha/2}$,

$$P(|T_T| > z_{1-\alpha/2}) - \alpha = O(T^{-1}).$$

- **Takeaway:** Size distortion converges to zero more rapidly using bootstrap critical values.

# General Hypothesis Testing with Bootstrap

Consider a general hypothesis testing problem:

$$H_0 : P \in \mathcal{P}_0, \quad H_1 : P \in \mathcal{P}_1, \quad \mathcal{P}_0 \cap \mathcal{P}_1 = \emptyset.$$

Having picked a suitable $T_T$, our goal is to construct a (data-dependent) critical value $c_T(1-\alpha)$ s.t.

- when $P \in \mathcal{P}_0$, $P\{T_T > c_T(1-\alpha)\} \to \alpha$ as $T \to \infty$;
- when $P \in \mathcal{P}_1$, $P\{T_T > c_T(1-\alpha)\} \to 1$ as $T \to \infty$.

Let $G_T(x, P) = \Pr_P(T_T < x)$. A bootstrap critical value can be defined by

$$g_T(1-\alpha, \hat{Q}_T) = \inf\{x : G_T(x, \hat{Q}_T) \geq 1-\alpha)\}$$

where $\hat{Q}_T$ is an estimate of $P \in \mathcal{P}_0$ so that $\hat{Q}_T$ satisfies the null hypothesis constraints.

# General Hypothesis Testing with Bootstrap

Choice of resampling distribution $\hat{Q}_T$ should satisfy:

- if $P \in \mathcal{P}_0$, $\hat{Q}_T$ is near $P$ so that $g_T(1-\alpha, P) \approx g_T(1-\alpha, \hat{Q}_T)$;
- if $P \in \mathcal{P}_1$, $\hat{Q}_T$ should not approach $P$ but some $P_0 \in \mathcal{P}_0$.

Notice that we would not want to replace $\hat{Q}_T$ by empirical distribution $\hat{P}_T$.

# General Hypothesis Testing with Bootstrap

**Example** (Testing the mean)

- Let $X_1, \ldots, X_T$ be real-valued with finite mean and variance.
- Test whether the mean is zero. $T_T = T\bar{X}_T^2$.
- Let $\hat{Q}_T$ be the distribution in $\mathcal{P}_0$ closest to $\hat{P}_T$.
  - Closeness can be described by Kullback-Leibler divergence

$$\delta_{KL}(P, Q) = \int \ln\left(\frac{dP}{dQ}\right) dP.$$

  - Let $\hat{Q}_T = \min_{Q \in \mathcal{P}_0} \delta_{KL}(\hat{P}_T, Q)$. Then $\hat{Q}_T$ assigns $w_t$ to $X_t$ where

$$w_t \propto \frac{(1 + lX_t)^{-1}}{\sum_{s=1}^{T}(1 + lX_s)^{-1}}$$

    where $l$ is chosen s.t. $\sum_{t=1}^{T} w_t X_t = 0$.
- Alternatively, can directly use $T_T = T\delta_{KL}(\hat{P}_T, \hat{Q}_T)^2$.

[2] Any testing problem can be formulated by $T_T = \tau_T \delta(\hat{P}_T, \tau(\hat{P}_T))$ with $\tau(P)$ defined by $\rho(P, \tau(P)) = \inf_{Q \in \mathcal{P}_0} \rho(P, Q)$ for some metrics $\rho$ and $\delta$ and scaling $\tau_T$.

# A Partial Guide for Practitioners

- To use bootstrap for standard error, the estimator must be asymptotically normal.[3]
- To achieve higher-order refinement, the test statistic needs to be asymptotically pivotal.
- Do not use bootstrap for weak IV regressions.
  - Discontinuity in limiting distribution (Andrews and Guggenberger, 2010)
  - Systematic errors in estimating the strength of instruments (Andrews, Stock, and Sun, 2019)

[3] In some cases the bootstrap standard error is not consistent even though the estimator is asymptotically normal. One notable example is nearest neighbor matching estimator (Abadie and Imbens, 2008).

# Bootstrap Failure: Parameter on the Boundary

- Consider a random sample $\{X_t\}_{t=1}^T$ from $N(\mu, 1)$ where $\mu \geq 0$.
- MLE estimate of $\mu$ is $\hat{\mu}_{MLE} = \max\{\bar{X}_T, 0\}$ where $\bar{X}_T = \frac{1}{T}\sum_{t=1}^T X_t$.
- Assume $\text{Var}(X) = 1$ is known. $t$-statistic has asymptotic distribution:

$$\sqrt{T}(\hat{\mu}_{MLE} - \mu) \xrightarrow{d} \begin{cases} Z & \text{if } \mu > 0 \\ \max\{Z, 0\} & \text{if } \mu = 0 \end{cases} \quad \text{where } Z \sim N(0,1).$$

- When $\mu = 0$, bootstrap $t$-statistic $\sqrt{T}(\hat{\mu}^*_{MLE} - \hat{\mu}_{MLE}) \overset{d}{\not\to} \max\{Z, 0\}$ conditional on almost all paths $X_1, X_2, \ldots$. For any $c > -x > 0$,

$$\begin{aligned} &\Pr(\sqrt{T}(\hat{\mu}^*_{MLE} - \hat{\mu}_{MLE}) \leq x | \sqrt{T}\bar{X}_T > c) \\ \geq &\Pr(\max\{\sqrt{T}(\bar{X}^*_T - \bar{X}_T), -c\} \leq x | \sqrt{T}\bar{X}_T > c) \\ \to &\Pr(\max\{Z, -c\} \leq x) > \Pr(\max\{Z, 0\} \leq x). \end{aligned}$$

# Bootstrap Failure: Maximum of a Sample

- Consider a random sample $\{X_t\}_{t=1}^T$ from $U(0, \theta_0)$. $\theta_0 = 1$ is unknown.
- MLE estimate of $\theta_0$ is $\hat{\theta}_{MLE} = \max\{X_1, \ldots, X_T\}$.
- Define $T_T = T(\hat{\theta}_{MLE} - 1)$. As $T \to \infty$, $P(T_T \leq -z) = e^{-z}$ for any $z \geq 0$. Moreover, $P(T_T = 0) = 0$ for all $T$.
- The bootstrap analog $T_T^* = T(\hat{\theta}_{MLE}^* - \hat{\theta}_{MLE})$ satisfies

$$P_T^*(T_T^* = 0) = 1 - (1 - T^{-1})^T \to 1 - e^{-1} \text{ as } T \to \infty.$$

Hence, nonparametric bootstrap does not estimate the distribution of $T_T$ consistently.

# Outline

1. Nonparametric Bootstrap
   - Asymptotic Refinement
   - Hypothesis Testing
   - Bootstrap Failure
2. Extensions
   - Wild Bootstrap
   - Block Bootstrap
   - Nonsmooth Statistics

# Wild Bootstrap

Take the linear regression model

$$y_t = X_t'\beta + \varepsilon_t, \quad E(\varepsilon_t|X_t) = 0.$$

- **Nonparametric bootstrap** samples $(y_t^*, X_t^*)$ i.i.d. from original observations: inaccurate because it does not impose $E(\varepsilon_t|X_t) = 0$.
- **Residual bootstrap** draws errors independent from $X_t$: stronger than conditional mean restriction
- **Wild bootstrap** (Liu, 1988; Mammen, 1993) generates bootstrap samples from

$$y_t^* = X_t'\hat{\beta} + \varepsilon_t^*.$$

where $\hat{\beta}$ is the OLS estimator.

# Wild Bootstrap

$\varepsilon_t^*$'s can be generated by two methods:

1. Let $\varepsilon_t^* = \xi_t^* \hat{\varepsilon}_t$, where $\hat{\varepsilon}_t$ are OLS residuals and $\xi_t^*$ are i.i.d. auxiliary random variables with distributions
   - Rademacher random variables: $P(\xi_t^* = 1) = P(\xi_t^* = -1) = \frac{1}{2}$
   - Mammen (1993) two-point distribution: $P(\xi_t^* = \frac{1+\sqrt{5}}{2}) = \frac{\sqrt{5}-1}{2\sqrt{5}}, P(\xi_t^* = \frac{1-\sqrt{5}}{2}) = \frac{\sqrt{5}+1}{2\sqrt{5}}$
2. **Multiplier bootstrap** (Davidson and Flachaire, 2008): $\varepsilon_t^* = U_t f(\hat{\varepsilon}_t)$, where $U_t$ are random variables that are independent of each other and $\hat{\varepsilon}_t$ with $E(U_t) = 0$ and $E(U_t^2) = 1$, and $f(\cdot)$ is a transformation.

Bootstrap sampling must capture dependence of DGP.

- $\{X_t\}$ is generated by a stationary, invertible, finite-order ARMA model

$$A(L,\alpha)X_t = B(L,\beta)U_t$$

- $\{X_t\}$ is generated by a stationary, linear process

$$X_t - \mu = \sum_{j=1}^{\infty} \alpha_j (X_{t-j} - \mu) U_t$$

  **Sieve bootstrap:** generate bootstrap samples according to $AR(p)$

- $\{X_t\}$ is a stationary Markov process

  **Markov bootstrap:** generate bootstrap samples by a nonparametric estimate of Markov transition density

# Block Bootstrap

Divide data into blocks and sample blocks randomly with replacement

- Fixed block length $l$
  - Non-overlapping blocks: $\{X_1, \ldots, X_l\}, \{X_{l+1}, \ldots, X_{2l}\}, \ldots$
  - Overlapping blocks: $\{X_1, \ldots, X_l\}, \{X_2, \ldots, X_{l+1}\}, \ldots$
- Random block length: **Stationary bootstrap**

**Remarks:**

- Block length must increase with sample size for consistency
- In terms of asymptotic RMSE, stationary bootstrap is unattractive relative to block bootstrap with fixed-length blocks.
- Block bootstrap does not exactly replicate dependence structure
  $\Rightarrow$ Need to develop special versions of test statistics to obtain asymptotic refinements (e.g. Hall and Horowitz, 1996)

# Nonsmooth Statistics: Maximum Score Estimator

Consider the binary response model $Y_t = 1\{X_t'\beta + U \geq 0\}$ with $P(U \geq 0|X) = 0.5$. Manski (1975, 1985)'s maximum score estimator is

$$\hat{b}_{MS} = \arg\max_{b \in B} \sum_{t=1}^{T} (2Y_t - 1) 1\{X_t' b \geq 0\}, \quad \text{where } B = \{b : |b_1| = 1\}.$$

$\hat{b}_{MS}$ converges slowly and has a complicated limiting distribution. Horowitz (1992) proposes the smoothed maximum score (SMS) estimator

$$\hat{b}_{SMS} = \arg\max_{b \in B} \sum_{t=1}^{T} (2Y_t - 1) H\left(\frac{X_t' b}{h_T}\right)$$

where $H(\cdot)$ is sufficiently smooth s.t. $H(v) = 1$ if $v \geq 1$ and $H(v) = 0$ if $v \leq -1$, $\{h_T\}$ are decreasing positive constants s.t. $h_T \to 0$ as $T \to \infty$.

- $(Th_T)^{1/2}(\hat{b}_{SMS} - \beta)$ is asymptotically normal.
- Asymptotic refinements: $P(|t_T| > z^*_{1-\alpha}) = \alpha + o[(Th_T)^{-1}]$.

# Cramér Condition

Let $\tau$ be a vector of constants with the same dimension as $X$. Let $i = \sqrt{-1}$. $X$ satisfies the Cramér condition if

$$\lim_{\|\tau\| \to \infty} \sup |E \exp(i\tau' X)| < 1.$$

- Cramér condition is satisfied if $X$ is continuously distributed but not if $X$ is discrete.
- Let $X = X(Z)$. Cramér condition is satisfied even if some components of $Z$ are discrete but $X(Z)$ is continuously distributed.

Back

# Bibliography


Abadie, A. and Imbens, G. W. (2008), "On the failure of the bootstrap for matching estimators," *Econometrica*, 76, 1537–1557.

Andrews, D. W. and Guggenberger, P. (2010), "Asymptotic size and a problem with subsampling and with the m out of n bootstrap," *Econometric Theory*, 426–468.

Andrews, I., Stock, J. H., and Sun, L. (2019), "Weak instruments in instrumental variables regression: Theory and practice," *Annual Review of Economics*, 11, 727–753.

Davidson, R. and Flachaire, E. (2008), "The wild bootstrap, tamed at last," *Journal of Econometrics*, 146, 162–169.

Hall, P. and Horowitz, J. L. (1996), "Bootstrap critical values for tests based on generalized-method-of-moments estimators," *Econometrica: Journal of the Econometric Society*, 891–916.

Horowitz, J. L. (1992), "A smoothed maximum score estimator for the binary response model," *Econometrica: journal of the Econometric Society*, 505–531.

— (2001), "The bootstrap," in *Handbook of econometrics*, Elsevier, vol. 5, pp. 3159–3228.

— (2019), "Bootstrap methods in econometrics," *Annual Review of Economics*, 11, 193–224.

Liu, R. Y. (1988), "Bootstrap procedures under some non-iid models," *Annals of Statistics*, 16, 1696–1708.

Mammen, E. (1993), "Bootstrap and wild bootstrap for high dimensional linear models," *The annals of statistics*, 255–285.

Manski, C. F. (1975), "Maximum score estimation of the stochastic utility model of choice," *Journal of econometrics*, 3, 205–228.

— (1985), "Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator," *Journal of econometrics*, 27, 313–333.