

11. Bootstrap Methods

© A. Colin Cameron & Pravin K. Trivedi 2006

These transparencies were prepared in 20043.

They can be used as an adjunct to

Chapter 11 of our subsequent book

Microeconometrics: Methods and Applications

Cambridge University Press, 2005.

Original version of slides: May 2004

Bootstrap Methods

Outline

1. Introduction
2. Bootstrap Overview
3. Bootstrap Example
4. Bootstrap Theory
5. Bootstrap Extensions
6. Bootstrap Applications

1 Introduction

- Exact finite sample results usually unavailable. Instead use asymptotic theory - limit normal and χ^2 .
- **Bootstrap** due to Efron (1979) is an alternative method.
Approximate distribution of statistic by Monte Carlo simulation, with sampling from the empirical distribution or the fitted distribution of the observed data.
- Like conventional methods relies on asymptotic theory so only exact in infinitely large samples.
- **Simplest bootstraps** no better than usual asymptotics but may be easier to implement.
- **More complicated bootstraps provide asymptotic refinement** so perhaps better finite sample approximation.

2 Bootstrap Summary

- Sample: $\{\mathbf{w}_1, \dots, \mathbf{w}_N\}$ iid over i
Usually: $\mathbf{w}_i = (y_i, \mathbf{x}_i)$
- Estimator: $\hat{\theta}$ is smooth root- N consistent and is asymptotically normal.
Simplicity: state results for scalar θ .
- Statistics of interest:
 - estimate $\hat{\theta}$
 - standard errors $s_{\hat{\theta}}$
 - t –statistic $t = (\hat{\theta} - \theta_0) / s_{\hat{\theta}}$ where θ_0 is H_0 value
 - associated critical value or p –value for the test
 - confidence interval.

2.1 Bootstrap Without Refinement

- Estimate variance $V[\hat{\mu}]$ of the sample mean $\hat{\mu} = \bar{y} = N^{-1} \sum_{i=1}^N y_i$ where y_i is iid $[\mu, \sigma^2]$.
- Obtain S samples of size N from the population.
Gives S estimates $\hat{\mu}_s = \bar{y}_s$, $s = 1, \dots, S$.
Then $\hat{V}[\hat{\mu}] = (S - 1)^{-1} \sum_{s=1}^S (\hat{\mu}_s - \bar{\hat{\mu}})^2$,
where $\bar{\hat{\mu}} = S^{-1} \sum_{s=1}^S \hat{\mu}_s$.
- This approach not possible as only have one sample.
Bootstrap: treat the sample as the population.
So finite population is the actual data $y_1, \dots, y_N$.
Draw B bootstrap samples from this population of size N , sampling from $y_1, \dots, y_N$ with replacement.
[In each bootstrap sample some original data points appear more than once while others appear at all.]
Gives B estimates $\hat{\mu}_b = \bar{y}_b$, $b = 1, \dots, B$.
Then $V[\hat{\mu}] = (B - 1)^{-1} \sum_{b=1}^B (\hat{\mu}_b - \bar{\hat{\mu}})^2$,
where $\bar{\hat{\mu}} = B^{-1} \sum_{b=1}^B \hat{\mu}_b$.

2.1.1 Bootstrap Example

- Data generated from an exponential distribution with an exponential mean with two regressors

$$\begin{aligned}y_i|x_{2i}, x_{3i} &\sim \text{exponential}(\lambda_i), \quad i = 1, \dots, 50 \\ \lambda_i &= \exp(\beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i}) \\ (x_{2i}, x_{3i}) &\sim \mathcal{N}[0.1, 0.1; 0.1^2, 0.1^2, 0.005] \\ (\beta_1, \beta_2, \beta_3) &= (-2, 2, 2).\end{aligned}$$

- MLE for 50 observations gives $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$
Focus on statistical inference for β_3 :
 $\hat{\beta}_3 = 4.664 \quad s_3 = 1.741 \quad t_3 = 2.679.$
- Use paired bootstrap with (y_i, x_{2i}, x_{3i}) jointly re-sampled with replacement $B = 999$ times.
- **Bootstrap standard error estimate = 1.939**
since $\hat{\beta}_{3,1}^*, \dots, \hat{\beta}_{3,999}^*$ had mean 4.716 and standard deviation of 1.939.
[compared to usual asymptotic estimate of 1.741).

2.2 Asymptotic Refinements

- Example: Nonlinear estimators asymptotically unbiased but biased in finite samples

- Often can show that

$$E[\hat{\theta} - \theta_0] = \frac{a_N}{N} + \frac{b_N}{N^2} + \frac{c_N}{N^3} + \dots$$

where a_N , b_N and c_N are bounded constants that vary with the data and estimator.

- Consider alternative estimator $\tilde{\theta}$ with

$$E[\tilde{\theta} - \theta_0] = \frac{B_N}{N^2} + \frac{C_N}{N^3} + \dots$$

where B_N and C_N are bounded constants.

- Both estimators are unbiased as $N \rightarrow \infty$.

The latter is an **asymptotic refinement**:

asymptotic bias of size $O(N^{-2}) < O(N^{-1})$.

- Bootstrap one of several asymptotic refinement methods.
- Smaller asymptotic bias not guaranteed to give smaller finite sample bias:

$$B_N/N^2 > (a_N/N + b_N/N^2)$$

is possible for small N .

- Need simulation studies to confirm finite sample gains.

2.3 Asymptotically Pivotal Statistic

- Asymptotic refinement for tests and conf. intervals.
 α = nominal size for a test, e.g. $\alpha = 0.05$.
Actual size = $\alpha + O(N^{-1/2})$ for usual one-sided tests.
- Asymptotic refinement requires statistic to be an **asymptotically pivotal statistic**, meaning limit distribution does not depend on unknown parameters.
- e.g. Sampling from $y_i \sim [\mu, \sigma^2]$.
 - $\hat{\mu} = \bar{y} \stackrel{a}{\sim} \mathcal{N}[\mu, \sigma^2/N]$ is not asymptotically pivotal since distribution depends on unknown σ^2 .
 - $t = (\hat{\mu} - \mu_0)/s_{\hat{\mu}} \stackrel{a}{\sim} \mathcal{N}[0, 1]$, the **studentized statistic**, is asymptotically pivotal.
- Estimators are usually not asymptotically pivotal.
Conventional tests and CI's are asymp. pivotal.

2.4 The Bootstrap

A general **bootstrap algorithm** is as follows:

1. Given data $\mathbf{w}_1, \dots, \mathbf{w}_N$
draw a bootstrap sample of size N (see below)
denote this new sample $\mathbf{w}_1^*, \dots, \mathbf{w}_N^*$.
2. Calculate an appropriate statistic using the bootstrap sample. Examples include:
 - (a) estimate $\hat{\theta}^*$ of θ ;
 - (b) standard error $s_{\hat{\theta}}^*$ of estimate $\hat{\theta}^*$
 - (c) t -statistic $t^* = (\hat{\theta}^* - \hat{\theta})/s_{\hat{\theta}}^*$ centered at $\hat{\theta}$.
3. Repeat steps 1-2 B independent times.
Gives B bootstrap replications of $\hat{\theta}_1^*, \dots, \hat{\theta}_B^*$ or $t_1^*, \dots, t_B^*$
or
4. Use these B bootstrap replications to obtain a bootstrapped version of the statistic (see below).

2.4.1 Bootstrap Sampling Methods (step 1)

- **Empirical distribution function (EDF) bootstrap** or **nonparametric bootstrap**.

$\mathbf{w}_1^*, \dots, \mathbf{w}_N^*$ obtained by sampling with replacement from $\mathbf{w}_1, \dots, \mathbf{w}_N$ (used earlier).

Also called **paired bootstrap** as for single equation models the pair $(y_i, \mathbf{x}_i)$ is being resampled.

- **Parametric bootstrap** for fully parametric models.
Suppose $y|\mathbf{x} \sim F(\mathbf{x}, \boldsymbol{\theta}_0)$ and have estimate $\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0$.
Bootstrap generate y_i^* by random draws from $F(\mathbf{x}_i, \hat{\boldsymbol{\theta}})$ where $\mathbf{x}_i$ may be the original sample
[or may first resample $\mathbf{x}_i^*$ from $\mathbf{x}_1, \dots, \mathbf{x}_N$]
- **Residual bootstrap** for additive error regression
Suppose $y_i = g(\mathbf{x}_i, \boldsymbol{\beta}) + u_i$.
Form fitted residuals $\hat{u}_1, \dots, \hat{u}_N$.
Bootstrap residuals $(\hat{u}_1^*, \dots, \hat{u}_N^*)$ yield bootstrap sample $(y_1^*, \mathbf{x}_1), \dots, (y_N^*, \mathbf{x}_N)$, for $y_i^* = g(\mathbf{x}_i, \hat{\boldsymbol{\beta}}) + \hat{u}_i^*$.

2.4.2 The Number of Bootstraps (step 3)

- Asymptotic refinement can occur even for low B but higher B is even better.
- For standard error computation Efron and Tibsharani (1993, p.52) say $B = 50$ is often enough and $B = 200$ is almost always enough.
- Hypothesis tests and confidence intervals at standard levels of statistical significance involve the tails of the distribution, so more replications are needed.
See Davidson and MacKinnon (1999a,b and 2000).
For hypothesis testing at level α choose B so that $\alpha(B + 1)$ is an integer.
e.g. at $\alpha = .05$ let $B = 399$ rather than 400.
 $B \geq 399$ if $\alpha = 0.05$ and $B \geq 1499$ if $\alpha = 0.01$.
- Andrews and Buchinsky (2000) present an application-specific three-step numerical method to determine B for a given desired percentage deviation of the bootstrap quantity from that when $B = \infty$.

2.5 Standard Error Estimation

- The **bootstrap estimate of variance** of an estimator is the usual formula for estimating a variance, applied to the B bootstrap replications $\hat{\theta}_1^*, \dots, \hat{\theta}_B^*$:

$$s_{\hat{\theta}, \text{Boot}}^2 = \frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}_b^* - \bar{\hat{\theta}}^*)^2,$$

where $\bar{\hat{\theta}}^* = B^{-1} \sum_{b=1}^B \hat{\theta}_b^*$.

- The **bootstrap estimate of the standard error**, $s_{\hat{\theta}, \text{Boot}}$, is obtained by taking the square root.
- This bootstrap provides no asymptotic refinement. But extraordinarily useful when it is difficult to obtain standard errors using conventional methods:
 - Sequential two-step m-estimator
 - 2SLS estimator with heteroskedastic errors (if no White option).
 - Functions of other estimates e.g. $\hat{\theta} = \hat{\alpha}/\hat{\beta}$
 - Clustered data with many small clusters, such as short panels. Then resample the clusters.

2.6 Tests with Asymptotic Refinement

- $T_N = (\hat{\theta} - \theta_0)/s_{\hat{\theta}} \stackrel{a}{\sim} \mathcal{N}[0, 1]$ is asymptotically pivotal, so bootstrap this.

- Bootstrap gives B test statistics $t_1^*, \dots, t_B^*$, where

$$t_b^* = (\hat{\theta}_b^* - \hat{\theta})/s_{\hat{\theta}_b^*},$$

is **centered around the original estimate** $\hat{\theta}$ since resampling is from a distribution centered around $\hat{\theta}$.

- Let $t_1^*, \dots, t_B^*$ denote statistics ordered from smallest to largest.
- Upper alternative test: $H_0 : \theta \leq \theta_0$ vs $H_a : \theta > \theta_0$
the **bootstrap critical value** (at level α) is the upper α quantile of the B ordered test statistics.
e.g. if $B = 999$ and $\alpha = 0.05$ then
the critical value is the 950th highest value of t^* ,
since then $(B + 1)(1 - \alpha) = 950$.

- Can also compute a **bootstrap p -value**.
e.g. if original statistic t lies between the 914th and 915th largest values and $B = 999$ then $p_{\text{boot}} = 1 - 914/(B + 1) = 0.086$.
- For a **non-symmetrical test** the bootstrap **critical values** (at level α) are the lower $\alpha/2$ and upper $\alpha/2$ quantiles of the ordered test statistics t^* .
Reject H_0 if the original t lies outside this range.
- For a **symmetrical test** we instead order $|t^*|$ and the bootstrap **critical value** (at level α) is the upper α quantile of the ordered $|t^*|$.
Reject H_0 if $|t|$ exceeds this critical value.
- These tests, using the **percentile-t method**, provide asymptotic refinements.
One-sided t test and a non-symmetrical two-sided t test the true size $\alpha + O(N^{-1/2}) \rightarrow \alpha + O(N^{-1})$.
Two-sided symmetrical t test or an asymptotic chi-square test true size $\alpha + O(N^{-1}) \rightarrow \alpha + O(N^{-2})$.

2.6.1 Tests without Asymptotic Refinement

- Alternative bootstrap methods can be used. While asymptotically valid there is no asymptotic refinement.

1. Compute $t = (\hat{\theta} - \theta_0) / s_{\hat{\theta}, \text{boot}}$ where $s_{\hat{\theta}, \text{boot}}$ replaces the usual estimate $s_{\hat{\theta}}$.
Compare this test statistic to standard normal critical values.

2. For two-sided test of $H_0 : \theta = \theta_0$ against $H_a : \theta \neq \theta_0$ find the lower $\alpha/2$ and upper $\alpha/2$ quantiles of the bootstrap estimates $\hat{\theta}_1^*, \dots, \hat{\theta}_B^*$.
Reject H_0 if the original estimate $\hat{\theta}$ follows outside this region.

This is called the **percentile method**.

- These two bootstraps do not require computation of $s_{\hat{\theta}}$, the usual standard error estimate based on asymptotic theory.

2.6.2 Bootstrap Example (continued)

- Summary statistics and percentiles based on 999 paired bootstrap resamples for:
 - (1) estimate $\hat{\beta}_3^*$;
 - (2) the associated statistics $t_3^* = (\hat{\beta}_3^* - 4.664)/s_{\hat{\beta}_3^*}$;
 - (3) $t(47)$ quantiles;
 - (4) standard normal quantiles.

	$\hat{\beta}_3^*$	t_3^*	$z=t(\infty)$	$t(47)$
Mean	4.716	0.026	1.021	1.000
St.Dev.	1.939	1.047	1.000	1.021
1%	-.336	-2.664	-2.326	-2.408
2.5%	0.501	-2.183	-1.960	-2.012
5%	1.545	-1.728	-1.645	-1.678
25%	3.570	-0.621	-0.675	-0.680
50%	4.772	0.062	0.000	0.000
75%	5.971	0.703	0.675	0.680
95%	7.811	1.706	1.645	1.678
97.5%	8.484	2.066	1.960	2.012
99.0%	9.427	2.529	2.326	2.408

- **Hypothesis Testing with Asymptotic Refinement:**

Test $H_0 : \beta_3 = 0$ vs $H_a : \beta_3 \neq 0$ at level .05.

Compute $t_3^* = (\hat{\beta}_3^* - 4.664) / s_{\hat{\beta}_3^*}$.

Bootstrap critical values are -2.183 and 2.066 .

Reject H_0 since original $t_3 = 2.679 > 2.066$,

- Figure plots bootstrap estimate of the density of t_3 and compares it to the standard normal.

- **Hypothesis Testing without Asymptotic Refinement 1:**

Use the bootstrap standard error estimate of 1.939, rather than the asymptotic standard error estimate of 1.741.

Then $t_3 = (4.664 - 0)/1.939 = 2.405$.

Reject H_0 at level .05 as $2.405 > 1.960$ using standard normal critical values.

- **Hypothesis Testing without Asymptotic Refinement 2:**

From Table 11.1, 95 percent of the bootstrap estimates $\hat{\beta}_3^*$ lie in the range (0.501, 8.484).

Reject $H_0 : \beta_3 = 0$ as this range does not include the hypothesized value of 0.

2.7 Confidence Intervals

- Statistics literature focuses on confidence intervals.
- **Asymptotic refinement using the percentile-t method:**
The $100(1 - \alpha)$ percent confidence interval is

$$(\hat{\theta} - t_{[1-\alpha/2]}^* \times s_{\hat{\theta}}, \hat{\theta} + t_{[\alpha/2]}^* \times s_{\hat{\theta}}),$$

where $\hat{\theta}$ and $s_{\hat{\theta}}$ are the estimate and standard error from the original sample and

$t_{[1-\alpha/2]}^*$ and $t_{[\alpha/2]}^*$ denote the lower and upper $\alpha/2$ quantiles $\alpha/2$ of the t -statistics $t_1^*, \dots, t_B^*$.

- The **bias-corrected and accelerated (BC_a) method** detailed in Efron (1987) gives asymptotic refinement for a wider class of problems.
- **No asymptotic refinement** (though valid).
 1. $(\hat{\theta} - z_{[1-\alpha/2]} \times s_{\hat{\theta}, \text{boot}}, \hat{\theta} + z_{[\alpha/2]} \times s_{\hat{\theta}, \text{boot}})$
 2. $(\hat{\theta}_{[1-\alpha/2]}^*, \hat{\theta}_{[\alpha/2]}^*)$ the percentile method.

2.8 Bias Reduction

- Most estimators are biased in finite samples.
We wish to estimate the bias $E[\hat{\theta}] - \theta$.
- Let $\bar{\hat{\theta}}^*$ denote the average of $\hat{\theta}^*$ from the bootstraps.
The **bootstrap estimate of the bias** is then

$$\text{Bias}_{\hat{\theta}} = (\bar{\hat{\theta}}^* - \hat{\theta}).$$

- If $\hat{\theta} = 4$ and $\bar{\hat{\theta}}^* = 5$ then $\hat{\theta}$ **overestimates** by 1
(truth was 4 and the many estimates averaged 5).
So the bias-corrected estimate is $4 - 1 = 3$.
- The **bootstrap bias-corrected estimator** of θ is

$$\hat{\theta}_{\text{Boot}} = \hat{\theta} - (\bar{\hat{\theta}}^* - \hat{\theta}) = 2\hat{\theta} - \bar{\hat{\theta}}^*.$$

Note that $\bar{\hat{\theta}}^*$ itself is not the bias-corrected estimate.
The asymptotic bias of $\hat{\theta}_{\text{Boot}}$ is instead $O(N^{-2})$.

- Bias correction is seldom used for $\sqrt{N}$ consistent estimators, as $\hat{\theta}_{\text{Boot}}$ can be more variable than $\hat{\theta}$ and bias is often small relative to standard error.

2.8.1 Bootstrap Example (continued)

- Bootstrap: $\overline{\hat{\beta}_3}^* = 4.716$
 Original estimate $\hat{\beta}_3 = 4.664$.
 Estimated bias $= (4.716 - 4.664) = 0.052$
 Bias-corrected estimate $= 4.664 - 0.052 = 4.612$
- The bias of 0.052 is small compared to the standard error $s_3 = 1.741$.

3 Bootstrap in Stata

- * For exponential model MLE command is
* `ereg y x z`

- * To bootstrap
`bs "ereg y x z" " _b[z]", reps(999) level(95)`

Var	Reps	Observe	Bias	St.err.	95% CI
_bs_1	999	4.663	.052	1.939	(.86,8.47) (N)
					(.50,8.48) (P)
					(.23,8.17) (BC)

- N = normal is no refinement method 1
P = percentile is no refinement method 2
BC = bias-corrected is Efron (1986) refinement
- Stata does not do the percentile-t method.
But can save the 999 estimated coefficients.
Then use these to form 999 t-statistics, etc.

4 Bootstrap Theory

- See Horowitz (2001) - detailed though challenging.
- For root-N asymptotically normal statistic
 - The bootstrap should be from distribution F_N consistent for the dgp F_0 .
 - The bootstrap requires smoothness and continuity in F_0 (the dgp) and G_N (the finite sample distribution of the statistic).
 - The bootstrap assumes existence of low order moments, as low order cumulants appear in the Edgeworth expansions that the bootstrap implements.
 - Asymptotic refinement requires use of an asymptotically pivotal statistic.
 - Bootstrap refinement presented assumes iid data. Care needed even for heteroskedastic errors.

5 Bootstrap Extensions

- 1-2 permit consistent bootstrap for a wider range of applications. 3-5 give asymptotic refinement.
 - 1. **Subsample Bootstrap:** bootstrap resample of size m substantially smaller than sample size N . For nonsmooth estimators.
 - 2. **Moving blocks bootstrap:** split sample into non-overlapping blocks and sample with replacement from these blocks. For dependent data.
 - 3. **Nested bootstrap:** bootstrap within a bootstrap.
Provides asymptotic refinement for t-test even if don't have formula for standard error.
 - 4. **Recentering and Rescaling:** Ensure bootstrap based on estimate that imposes all the conditions of the model under consideration. e.g. over-identified GMM.
 - 5. **Jackknife:** Precursor of the bootstrap.

6 Bootstrap Applications

1. **Heteroskedastic Errors:** For asymptotic refinement use the wild bootstrap.
2. **Panel Data and Clustered Data:** Resample all observation in the cluster.
e.g. For panel data (i, t) resample over i only.
No refinement but gives heteroscedasticity and panel robust standard errors. on the cluster unit.
3. **Hypothesis and Specification Tests:** consider more complicated tests.
e.g. Can implement Hausman test without asymptotic refinement by bootstrap.
e.g. can do bootstrap refinement on LM tests that have “poor finite sample properties”
4. **GMM and EL in Over-identified Models:** For overidentified models need to recenter or use empirical likelihood.

5. **Nonparametric Regression:** Nonparametric density and regression estimators converge at rate less than $\text{root-}N$ and are asymptotically biased. This complicates inference such as confidence intervals. Bootstrap works.
6. **Non-Smooth Estimators:** e.g. LAD.
7. **Time Series:** bootstrap residuals if possible or do moving blocks.

7 Some References

Andrews, D.W.K. and M. Buchinsky (2000), "A Three-step Method for choosing the Number of Bootstrap Replications," *Econometrica*, 68, 23-51.

Brownstone, D. and C. Kazimi (1998), "Applying the The Bootstrap," Unpublished manuscript, Department of Economics, University of California - Irvine.

Efron, B. (1979), "Bootstrap Methods: Another Look at the Jackknife," *Annals of Statistics*, 7, 1-26.

Efron, B. (1982), "Better Bootstrap Confidence Intervals," *Journal of the American Statistical Association*, 82, 171-185.

Efron, B. and J. Tibsharani (1993), *An Introduction to the Bootstrap*, London: Chapman and Hall.

Hall, P. and J.L Horowitz (1996), "Bootstrap critical values for tests based on Generalized Method of Moments Estimators," *Econometrica*, 64, 891-916.

Horowitz, J.L. (1994), "Bootstrap-based Critical Values for the Information Matrix Test," *Journal of Econometrics*, 61, 395-411.

Horowitz, J.L. (2001), "The Bootstrap," *Handbook of Econometrics: Volume 5*, Amsterdam: North-Holland.