

Econ 424/Amath 540

The Bootstrap

Eric Zivot

July 17, 2012

Bootstrapping

Motivation

- Before modern computers, doing statistical analysis involved using mathematics and probability theory to derive statistical formulas for standard errors and confidence intervals. Often these formulas are approximations that rely on large samples
- With modern computers and statistical software, resampling methods (e.g. bootstrapping) can be used to produce standard errors and confidence intervals without the use of formulas that are often more reliable than statistical formulas

Advantages of Bootstrapping

- Fewer assumptions
 - Do not need data to be normally distributed
- Greater accuracy
 - Do not rely on very large sample sizes in contrast to Central Limit Theorem
- Generality
 - Same method applies to a wide variety of statistics

Bootstrapping in the CER Model

$$r_t = \mu + \epsilon_t \quad t = 1, \dots, T$$
$$\epsilon_t \sim \text{iid } N(0, \sigma^2)$$

Observed sample:

$$\{r_1, \dots, r_T\}$$

Goal: Compute $\widehat{SE}(\hat{\mu})$ and $\widehat{SE}(\hat{\sigma})$ and 95% confidence intervals for μ and σ using the bootstrap. The analytic formulas are

$$\widehat{SE}(\hat{\mu}) = \frac{\hat{\sigma}}{\sqrt{T}}, \quad \widehat{SE}(\hat{\sigma}) \approx \frac{\hat{\sigma}}{\sqrt{2T}}$$
$$\hat{\theta} \pm 2 \cdot \widehat{SE}(\hat{\theta}), \quad \hat{\theta} = \hat{\mu}, \hat{\sigma}$$

Procedure for Nonparametric Bootstrapping

1. Resample - create B bootstrap samples by sampling *with replacement* from the original data. Each bootstrap sample has T observations (same as original sample)

$$\{r_{11}^*, \dots, r_{1T}^*\} = \text{1st bootstrap sample}$$

\vdots

$$\{r_{B1}^*, \dots, r_{BT}^*\} = \text{Bth bootstrap sample}$$

2. Calculate bootstrap distribution for statistic of interest $\hat{\theta}$ - for each bootstrap sample compute $\hat{\theta}^*$. There will be B values of $\hat{\theta}^*$:

$$\{\hat{\theta}_1^*, \dots, \hat{\theta}_B^*\}$$

3. Use the bootstrap distribution - the bootstrap distribution gives information about the shape, center and spread of the unknown pdf $f(\hat{\theta})$

Bootstrap estimate of bias

$$\text{bias}_{boot}(\hat{\theta}, \theta) = \frac{1}{B} \sum_{j=1}^B \hat{\theta}_j^* - \hat{\theta}$$

bootstrap mean - estimate

Bootstrap estimate of $\widehat{SE}(\hat{\theta})$

$$\widehat{SE}_{boot}(\hat{\theta}) = \sqrt{\frac{1}{B-1} \sum_{j=1}^B \left(\hat{\theta}_j^* - \frac{1}{B} \sum_{j=1}^B \hat{\theta}_j^* \right)^2}$$

sample SD of bootstrap values $\{\hat{\theta}_1^*, \dots, \hat{\theta}_B^*\}$

Bootstrap 95% Confidence Intervals

1. If bootstrap distribution is symmetric and looks normal use

$$\hat{\theta} \pm 2 \cdot \widehat{SE}_{boot}(\hat{\theta})$$

2. If bootstrap distribution is not symmetric and looks non-normal use

$$[q_{.025}^*, q_{.975}^*]$$

$q_{.025}^* =$ 2.5% quantile from bootstrap distribution

$q_{.975}^* =$ 97.5% quantile from bootstrap distribution

Performing the Bootstrap in R

- Brute force
- R package boot

Brute force: perform the steps of the bootstrap by hand

1. sample with replacement from original data using `sample()` function B times
2. compute statistic from B bootstrap samples using for loop or `apply()` function
3. compute bootstrap bias and SE from bootstrap samples using `mean()` and `sd()` functions

Bootstrapping Value-at-Risk

In the CER model, 5% Value-at-Risk on an investment of $\$W$ is estimated using

$$\widehat{\text{VaR}}_{.05} = (e^{\hat{q}_{.05}} - 1)W$$

$$\hat{q}_{.05} = \hat{\mu} + \hat{\sigma} \cdot (-1.645)$$

$$-1.645 = 5\% \text{ quantile from } N(0, 1)$$

Bootstrapping can be used to compute

$$\widehat{SE}(\widehat{\text{VaR}}_{.05})$$

as well as confidence intervals.