1 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 Boostrapping

- 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

1.1 Bootstrapping in the CER Model

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

Observed sample:

$$
\{r_1, \ldots, r_T\}
$$

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

$$
\begin{aligned}
\widehat{\mathsf{SE}}(\hat{\mu}) &= \frac{\hat{\sigma}}{\sqrt{T}},\ \widehat{\mathsf{SE}}(\hat{\sigma}) \approx \frac{\hat{\sigma}}{\sqrt{2T}} \\
\hat{\theta} \pm 2 \cdot \widehat{\mathsf{SE}}(\hat{\theta}),\ \hat{\theta} &= \hat{\mu},\ \hat{\sigma}
\end{aligned}
$$

Procedure for 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)

$$
\begin{aligned}
\{r_{11}^*, \ldots, r_{1T}^*\} &= \text{1st bootstrap sample} \\
&\vdots \\
\{r_{B1}^*, \ldots, r_{BT}^*\} &= \text{Bth bootstrap sample}
\end{aligned}
$$

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

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

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

Bootstrap estimate of bias

$$\frac{1}{B}\sum_{j=1}^{B}\hat{\theta}_j^* - \hat{\theta}$$

bootstrap mean - estimate

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

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

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

Bootstrap 95% Confidence Intervals

1. If bootstrap distribution is symmetric use

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

2. If bootstrap distribution is not symmetric use

$$[q^*_{.025},\ q^*_{.975}]$$
$$q^*_{.025} = \text{ 2.5\% quantile from bootstrap distribution}$$
$$q^*_{.975} = \text{ 97.5\% quantile from bootstrap distribution}$$

1.1.1 Bootstrapping Value-at-Risk

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

$$\begin{aligned} \widehat{\mathrm{VaR}}_{.05} &= (e^{\hat{q}_{.05}} - 1)W \\ \hat{q}_{.05} &= \hat{\mu} + \hat{\sigma} \cdot (-1.645) \\ -1.645 &= 5\% \text{ quantile from } N(0,1) \end{aligned}$$

Bootstrapping can be used to compute

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

as well as confidence intervals.