Small-Sample Inference and Bootstrap

Leonid Kogan

MIT, Sloan

15.450, Fall 2010
Outline

1. Small-Sample Inference
2. Bootstrap
Overview

- So far, our inference has been based on asymptotic results: LLN and CLT.
- Asymptotic inference is sometimes difficult to apply, too complicated analytically.
- In small samples, asymptotic inference may be unreliable:
  - Estimators may be consistent but biased.
  - Standard errors may be imprecise, leading to incorrect confidence intervals and statistical test size.
- We can use simulation methods to deal with some of these issues:
  - Bootstrap can be used instead of asymptotic inference to deal with analytically challenging problems.
  - Bootstrap can be used to adjust for bias.
  - Monte Carlo simulation can be used to gain insight into the properties of statistical procedures.
Outline

1. Small-Sample Inference
2. Bootstrap
Example: Autocorrelation

- We want to estimate first-order autocorrelation of a time series $x_t$ (e.g., inflation), $\text{corr}(x_t, x_{t+1})$.
- Estimate by OLS (GMM)

$$x_t = a_0 + \rho_1 x_{t-1} + \varepsilon_t$$

- We know that this estimator is consistent:

$$\text{plim}_{T \to \infty} \hat{\rho}_1 = \rho_1$$

- We want to know if this estimator is biased, i.e., we want to estimate

$$E(\hat{\rho}_1) - \rho_1$$
Example: Autocorrelation, Monte Carlo

- Perform a Monte Carlo study to gain insight into the phenomenon.
- Simulate independently $N$ random series of length $T$.
- Each series follows an AR(1) process with persistence $\rho_1$ and Gaussian errors:

$$x_t = \rho_1 x_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, 1)$$

- Compute $\hat{\rho}_1(n)$, $n = 1, \ldots, N$ for each simulated sample.
- Estimate the bias:

$$\hat{\mathbb{E}}(\hat{\rho}_1) - \rho_1 = \frac{1}{N} \sum_{n=1}^{N} \hat{\rho}_1(n) - \rho_1$$

Standard error of our simulation-based estimate is

$$\hat{\sigma} = \sqrt{\frac{1}{N} \sum_{n=1}^{N} \left( \hat{\rho}_1(n) - \hat{\mathbb{E}}(\hat{\rho}_1) \right)^2}$$
Example: Autocorrelation, Monte Carlo

MATLAB® Code

```matlab
phi = 0.9; % AR(1) coefficient
T = 100; % Sample lengths
N = 100000; % Number of simulated samples
varss = 1/(1-phi^2); % STD of steady-state distribution
for n=1:N
    x = zeros(T,1);
    x(1) = sqrt(varss)*randn(1,1); % Draw initial value
    noise = randn(T-1,1);
    for t=2:T
        x(t) = phi*x(t-1) + noise(t-1);
    end
    X = [ones(T-1,1) x(1:T-1)];
    b = (X' * X) \ (X' * x(2:T)); rho(n) = b(2); % Run OLS
end
MeanBias = mean(rho) - phi
StdErrorBias = std(rho)/sqrt(N)
```
Example: Autocorrelation, Monte Carlo

- We use 100,000 simulations to estimate the average bias

<table>
<thead>
<tr>
<th>$\rho_1$</th>
<th>$T$</th>
<th>Average Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
<td>50</td>
<td>$-0.0826 \pm 0.0006$</td>
</tr>
<tr>
<td>0.0</td>
<td>50</td>
<td>$-0.0203 \pm 0.0009$</td>
</tr>
<tr>
<td>0.9</td>
<td>100</td>
<td>$-0.0402 \pm 0.0004$</td>
</tr>
<tr>
<td>0.0</td>
<td>100</td>
<td>$-0.0100 \pm 0.0006$</td>
</tr>
</tbody>
</table>

Bias seems increasing in $\rho_1$, and decreasing with sample size.

- There is an analytical formula for the average bias due to Kendall:

$$ E(\hat{\rho}_1) - \rho_1 \approx -\frac{1 + 3\rho_1}{T} $$

- When explicit formulas are not known, can use bootstrap to estimate the bias.
Example: Predictive Regression

- Consider a predictive regression (e.g., forecasting stock returns using dividend yield)

\[
\begin{align*}
  r_{t+1} &= \alpha + \beta x_t + u_{t+1} \\
  x_{t+1} &= \theta + \rho x_t + \varepsilon_{t+1} \\
  (u_t, \varepsilon_t)' &\sim \mathcal{N}(0, \Sigma)
\end{align*}
\]

- Stambaugh bias:

\[
E(\hat{\beta} - \beta) = \frac{\text{Cov}(u_t, \varepsilon_t)}{\text{Var}(\varepsilon_t)} E(\hat{\rho} - \rho) \approx -\frac{1 + 3\rho}{T} \frac{\text{Cov}(u_t, \varepsilon_t)}{\text{Var}(\varepsilon_t)}
\]

- In case of dividend yield forecasting stock returns, the bias is positive, and can be substantial compared to the standard error of \(\hat{\beta}\).
Predictive Regression: Monte Carlo

- Predictive regression of monthly S&P 500 excess returns on log dividend yield:
  \[
  r_{t+1} = \alpha + \beta x_t + u_{t+1} \\
  x_{t+1} = \theta + \rho x_t + \epsilon_{t+1}
  \]

- Data: CRSP, From 1/31/1934 to 12/31/2008.

- Parameter estimates:
  \[
  \hat{\beta} = 0.0089, \quad \hat{\rho} = 0.9936,
  \]

- S.E.(\hat{\beta}) = 0.005.
Predictive Regression: Monte Carlo

- Generate 1,000 samples with parameters equal to empirical estimates. Use 200 periods as burn-in, retain samples of the same length as historical.
- Tabulate $\hat{\beta}$ and standard errors for each sample. Use Newey-West with 6 lags to compute standard errors.

- Average of $\hat{\beta}$ is 0.013.
- Average bias in $\hat{\beta}$ is 0.004.
- Average standard error is 0.005.
- Average $t$-stat on $\beta$ is 0.75.
Testing the Mean: Non-Gaussian Errors

- We estimate the mean $\mu$ of a distribution by the sample mean. Tests are based on the asymptotic distribution

$$\frac{\hat{\mu} - \mu}{\hat{\sigma}/\sqrt{T}} \sim \mathcal{N}(0, 1)$$

- How good is the normal approximation in finite samples if the sample comes from a Non-Gaussian distribution?

- Assume that the sample is generated by a lognormal distribution:

$$x_t = e^{-\frac{1}{2} + \varepsilon_t}, \quad \varepsilon_t \sim \mathcal{N}(0, 1)$$
Monte Carlo experiment: $N = 100,000$, $T = 50$. Document the distribution of the $t$-statistic

$$\hat{t} = \frac{\hat{\mu} - 1}{\hat{\sigma}/\sqrt{T}}$$

Asymptotic theory dictates that $\text{Var}(\hat{t}) = 1$. We estimate

$$\text{Var}(\hat{t}) = 1.2542^2$$

Tails of the distribution of $\hat{t}$ are far from the asymptotic values:

$$\text{Prob}(\hat{t} > 1.96) \approx 0.0042, \quad \text{Prob}(\hat{t} < -1.96) \approx 0.1053$$
Outline

1 Small-Sample Inference

2 Bootstrap
Bootstrap: General Principle

- Bootstrap is a re-sampling method which can be used to evaluate properties of statistical estimators.
- Bootstrap is effectively a Monte Carlo study which uses the empirical distribution as if it were the true distribution.
- Key applications of bootstrap methodology:
  - Evaluate distributional properties of complicated estimators, perform bias adjustment;
  - Improve the precision of asymptotic approximations in small samples (confidence intervals, test rejection regions, etc.)
Bootstrap for IID Observations

- Suppose we are given a sample of IID observations $x_t, t = 1, \ldots, T$.
- We estimate the sample mean as $\hat{\mu} = \hat{E}(x_t)$. What is the 95% confidence interval for this estimator?
- Asymptotic theory suggests computing the confidence interval based on the Normal approximation

$$\sqrt{T} \frac{\hat{E}(x_t) - \mu}{\hat{\sigma}} \sim \mathcal{N}(0, 1), \quad \hat{\sigma}^2 = \frac{\sum_{t=1}^{T} [x_t - \hat{E}(x_t)]^2}{T}$$

- Under the empirical distribution, $x$ is equally likely to take one of the values $x_1, x_2, \ldots, x_T$. 
Key Idea of Bootstrap

REAL WORLD

Unknown probability model

Parameter of interest
θ = θ(P)

Estimated probability model

Estimated parameter
θ = θ(\(\hat{\theta}\))

Biasp(\(\hat{\theta}, \theta\))

BOOTSTRAP WORLD

Observed data

x = (x₁, x₂, ..., xₙ)

Bootstrap sample

x* = (x₁*, x₂*, ..., xₙ*)

Bootstrap replicate of \(\hat{\theta}\)

\(\hat{\theta}^*\) = s(x*)

Biasp(\(\hat{\theta}, \hat{\theta}^*\))

Source: Efron and Tibshirani, 1994, Figure 10.4.
Bootstrap Confidence Intervals

- Bootstrap confidence interval starts by drawing $R$ samples from the empirical distribution.

- For each bootstrapped sample, compute $\hat{\mu}^*$. "*" denotes statistics computed using bootstrapped samples.

- Compute 2.5% and 97.5% percentiles of the resulting distribution of $\hat{\mu}^*$:

$$\hat{\mu}^*_{2.5\%}, \hat{\mu}^*_{97.5\%}$$

- Approximate the distribution of $\hat{\mu} - \mu$ with the simulated distribution of $\hat{\mu}^* - \hat{\mu}$. Estimate the confidence interval as

$$\left(\hat{\mu} - (\hat{\mu}^*_{97.5\%} - \hat{\mu}), \hat{\mu} - (\hat{\mu}^*_{2.5\%} - \hat{\mu})\right)$$
Example: Lognormal Distribution

- Fix a sample of 50 observations from a lognormal distribution $\ln x_t \sim \mathcal{N}(-1/2, 1)$ and compute the estimates

  $\hat{\mu} = 1.1784$, $\hat{\sigma} = 1.5340$

- Population mean

  $\mu = E(x_t) = E(e^{-\frac{1}{2} \epsilon_t}) = 1$, $\epsilon_t \sim \mathcal{N}(0, 1)$

- Asymptotic approximation produces a confidence interval

  $(\hat{\mu} - 1.96 \frac{\hat{\sigma}}{\sqrt{T}}, \hat{\mu} + 1.96 \frac{\hat{\sigma}}{\sqrt{T}}) = (0.7532, 1.6036)$

- Compare this to the bootstrapped distribution.
Lognormal Distribution

Use bootstrap instead of asymptotic inference.

MATLAB® Code

```matlab
R = 10000;
muvec = zeros(R,1);
for r=1:R
    y = x(ceil(T*rand(T,1)));  % Sample with replacement
    muvec(r) = mean(y);
end
muvec = sort(muvec);
% 5 percent confidence interval
LeftEnd = muhat - (muvec(ceil(0.975*R)) - muhat)
RightEnd = muhat - (muvec(floor(0.025*R)) - muhat)
```

Bootstrap estimate of the confidence interval

\((0.7280, 1.5615)\)
Testing the Mean: Bootstrap
Lognormal Example

- Consistent with Monte Carlo results: small-sample distribution of t-statistics exhibits left-skewness.
- Variance of the bootstrapped t-statistic is $1.1852^2$. Normal approximation: $\text{Var}(\hat{t}) = 1$. Monte Carlo estimate: $\text{Var}(\hat{t}) = 1.2542^2$.

A histogram of $\hat{t}$ statistic

Bootstrap (10,000 samples)  Monte Carlo (100,000 samples)
Bootstrap Confidence Intervals

- The basic bootstrap confidence interval is valid, and can be used in situations when asymptotic inference is too difficult to perform.
- Bootstrap confidence interval is as accurate asymptotically as the interval based on the normal approximation.
- For $t$-statistic, bootstrapped distribution is more accurate than the large-sample normal approximation.
- Many generalizations of basic bootstrap have been developed for wider applicability and better inference quality.
Parametric Bootstrap

- Parametric bootstrap can handle non-IID samples.
- Example: a sample from an AR(1) process: \( x_t, t = 1, \ldots, T \):

\[
x_t = a_0 + a_1 x_{t-1} + \varepsilon_t
\]

- Want to estimate a confidence interval for \( \hat{a}_1 \).

  - Estimate the parameters \( \hat{a}_0, \hat{a}_1 \) and the residuals \( \hat{\varepsilon}_t \).
  - Generate \( R \) bootstrap samples for \( x_t \).
    - For each sample: generate a long series according to the AR(1) dynamics with \( \hat{a}_0, \hat{a}_1 \), drawing shocks with replacement from the sample \( \hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_T \);
    - Retain only the last \( T \) observations (drop the burn-in sample).
  - Compute the confidence interval as we would with basic nonparametric bootstrap using \( R \) samples.
Bootstrap Bias Adjustment

- Want to estimate small-sample bias of a statistic $\hat{\theta}$:

$$E \left[ \hat{\theta} - \theta_0 \right]$$

**REAL WORLD**
- Unknown probability model
  - $P$
  - Parameter of interest $\theta = \theta(P)$
  - Estimated data $x = (x_1, x_2, \ldots, x_n)$
  - Estimate of $\theta$ $\hat{\theta} = s(x)$
  - Bias $\text{Bias}_P(\hat{\theta}, \theta)$

**BOOTSTRAP WORLD**
- Estimated probability model
  - $P$
  - Bootstrap sample $x^* = (x^*_1, x^*_2, \ldots, x^*_n)$
  - Bootstrap replicate of $\hat{\theta}$ $\hat{\theta}^* = s(x^*)$
  - Estimated parameter $\hat{\theta} = \theta(\hat{P})$
  - Bias $\text{Bias}_P(\hat{\theta}^*, \hat{\theta})$

Source: Efron and Tibshirani, 1994, Figure 10.4.

Image by MIT OpenCourseWare.
Bootstrap Bias Adjustment

- Bootstrap provides an intuitive approach:
  \[ E \left[ \hat{\theta} - \theta_0 \right] \approx E_R \left[ \hat{\theta}^* - \hat{\theta} \right] \]

  where \( E_R \) denotes the average across the \( R \) bootstrapped samples.

- Intuition: treat the empirical distribution as exact, compute the average bias across bootstrapped samples.

- Caution: by estimating the bias, we may be adding sampling error. Correct for the bias if it is large compared to the standard error of \( \hat{\theta} \).
Example: Predictive Regression

- Use parametric bootstrap: 1,000 samples, 200 periods as burn-in, retain samples of same length as historical.
- Tabulate $\hat{\beta}$ and standard errors for each sample. Use Newey-West with 6 lags to compute standard errors.

- Average of $\hat{\beta}$ is 0.0125.
- Average bias in $\hat{\beta}$ is 0.0036.
- Average standard error is 0.005.
- Average $t$-stat on $\beta$ is 0.67.
Discussion

- Asymptotic theory is very convenient when available, but in small samples results may be inaccurate.
- Use Monte Carlo simulations to gain intuition.
- Bootstrap is a powerful tool. Use it when asymptotic theory is unavailable or suspect.

Bootstrap is not a silver bullet:
- Does not work well if rare events are missing from the empirical sample;
- Does not account for more subtle biases, e.g., survivorship, or sample selection.
- Does not cure model misspecification.

No substitute for common sense!
Readings

15.450 Analytics of Finance
Fall 2010

For information about citing these materials or our Terms of Use, visit: http://ocw.mit.edu/terms.