Lecture 2-c
Hypothesis Testing and Confidence Intervals

 1

Review – Returns

- Net or simple (total) return, R_t:

$$R_t = \frac{(P_t - P_{t-1}) + D_t}{P_{t-1}} = \text{capital gain + dividend yield}$$

where P_t = Stock price or Value of investment at time t
D_t = Dividend or payout of investment at time t

- Another commonly used definition of return, the *log (total) return*, r_t, defined as the log of the gross return:

$$r_t = \log(1 + R_t) = \log(P_t + D_t) - \log(P_{t-1})$$

<u>Note</u>: When the values are small (-0.1 to +0.1), both returns are *approximately* the same:

$$r_t = \log(1 + R_t) \approx R_t.$$

2

Review – Multi-period Returns & Real Returns

• Multi-period holding return

k-period gross holding return $(1 + R_{t,t+k}) = \prod_{j=0}^{k-1}(1 + R_{t+j,t+j+1})$

With the log approximation: $r_{t,t+k} = \sum_{j=0}^{k-1} r_{t+j,t+j+1}$

• If $r_{t+j,t+j+1} = r$ $\Rightarrow$ $r_{t,t+k} = \sum_{j=0}^{k-1} r_{t+j,t+j+1} = k * r$

• If $r_{t+j,t+j+1}$ are independent with a constant variance equal to σ_r^2

$\Rightarrow$ Var$(r_{t,t+k}) = \sum_{j=0}^{k-1} var(r_{t+j,t+j+1}) = k * \sigma_r^2$

$\Rightarrow$ SD$(r_{t,t+k}) = \sqrt{k * \sigma_r^2} = \sqrt{k} * \sigma_r$

3

Review – Returns: Sampling Distribution of $\bar{X}$

• Recall that the sampling distribution of the sample mean is:

$\bar{X} \sim N(\mu, \sigma^2/N)$

Example: Using Shiller's S&P 500 monthly log returns (N = **1805**).

Estimated Monthly mean return = $\bar{X}$ = **0.007378**

Estimated Var($\bar{X}$) = s^2/N = **0.040455**^2 /**1805** = 9.067075e-07

The SD of the monthly mean (also called the Standard Error, SE):

S.E.($\bar{X}$) = sqrt(9.067075e-07) = **0.000952** (or 0.095%).

$\Rightarrow \bar{X} \sim$ N(**0.007378**, **.00095**^{2})

Note: Compared to returns, expected returns, estimated by the sample mean, are more precisely estimated (**0.10%** vs **4.05%**). Not surprised, the sampling distribution of the mean shrinks towards the population mean as N increases.

4

Review – Returns: Sampling Distribution of s^2

- Under normal variates, the sampling distribution of s^2 is:

$$(N-1)\, s^2/\sigma^2 \sim \chi^2_{N-1}.$$

Example: Using Shiller's S&P 500 monthly log returns (N = **1805**).

Estimated Monthly sample variance = s^2 = **0.040455**^2 = 0.001637

Estimated Var[s^2] = $2 * \sigma^4/(N-1)$ = 2 * **0.040455**^4/1804 =

= 2.969493e-09

The SD of the monthly mean (also called the Standard Error, SE):

SE[s^2] = sqrt(2.969493e-09) = **0.0000545** (or 0.0055%).

Note: Compared to $\bar{X}$, s^2 is more precisely estimated (**0.00005%** vs **0.0001**).

5

Review – Yields

- Consider an n −period discount bond. Time is measured in years. Today is t. Bond (asset) pays F_{t+n} dollars n years from now, at $t+n$.

F_{t+n} = Face value (value at time $t+n$).

P_t = Market price of the bond (Today's initial capital).

$r_{n,t}$ = Yield to maturity (YTM) at time t for a maturity of n years.

n = Maturity of bond.

$$P_t = \frac{F_{t+n}}{(1+r_{n,t})^n}$$

As the number of compounding periods grow, $m \to \infty$, the value of the payoff at time $t+n$ becomes, in the limit:

$$P_t = e^{-r_t n}\, F_{t+n}$$

6

Review – Continuous compounding

• Suppose the continuously compounded bond at maturity pays $1 (=$F_{t+n}$) and the remaining duration is D units of time. Then,

$$P_t = e^{-r_t D}\ \$1$$

Then, the log return per unit of time is:

$$\log(P_{t+1}) - \log(P_t) = D * (r_t - r_{t+1})$$

where we ignore that D has one unit of time less at time t+1. Then, the daily return of a bond is the change of yields multiplied by duration.

• Now, suppose we invest P_0 in a bond with continuous compounding at an annual rate r. Then, the value of the investment at year t is:

$$V_t = P_0\, e^{rt} \qquad (\Rightarrow \log(V_t) = \log(P_0) + t\, r\,)$$

• The *log return* (log of gross return) *per year (unit of time)* is r:

$$\log(V_{t+1}) - \log(V_t) = [\log(P_0) + (t+1)\, r] - [\log(P_0) + t\, r] = r$$ 7

Bond Yields & Stock Returns

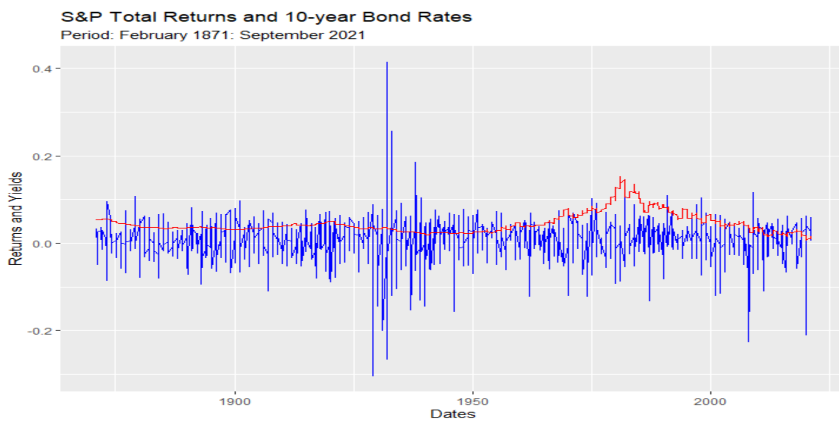
Example (continuation): We report the sample moments and descriptive statistics for **S&P total returns, 10-year interest rates** (annualized), and **total excess returns**, taken from Shiller's website.

	Total Return	10-year Interest Rates	Total Excess Returns
Mean	**0.007378**	**0.04511**	**0.003619**
Median	0.009928	0.0382	0.006395
Maximum	0.414151	0.1532	0.411264
Minimum	-0.30365	0.0062	-0.306435
Std. Dev.	**0.040455**	**0.02304**	**0.040519**
Skewness	-0.47050	1.79664	-0.450043
Kurtosis	14.61046	6.75173	14.51218
Jarque-Bera	**10204.9**	**2029.7**	**10028.0**
P-value (JB)	2.2e-16	2.2e-16	2.2e-16

8

Bond Yields & Stock Returns

Example (continuation): We plot long-run S&P returns (blue) and 10-year Bond interest rates (red) data.

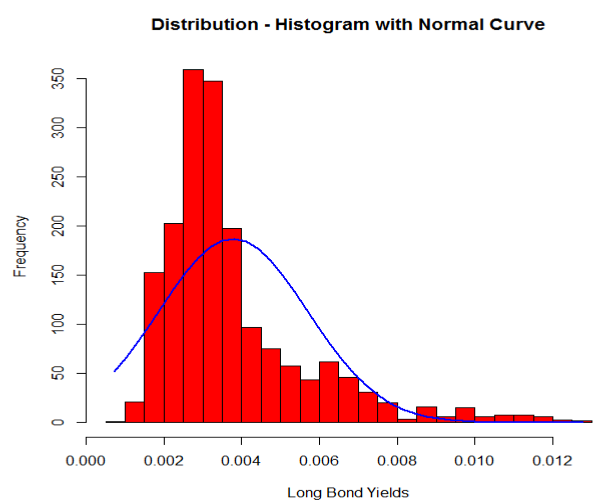


Note: Interest rates have a lower monthly mean (**0.003759** = **0.04511**/12), & lower volatility and kurtosis than stock returns.

9

Bond Yields: Distribution

Example (continuation): Bond interest rates data, from Robert Shiller's website:



10

Returns: Expected Returns & The ERP

• As mentioned above, returns are not very precisely estimated. They have a large variance. Things get better for expected returns, since the S.E. for the sample mean gets smaller with *N*.

• The expected return on any investment can be written as the sum of a *risk-free rate* and a *risk premium* to compensate for the risk in the investment.

• A key element in equity valuation models is the risk premium that investors demand for *market (equity) risk*. The difference between the expected market return and a risk-free rate is called the *equity risk premium* or **ERP**:

$$\text{ERP} = \text{E}[(r_{M,t} - r_{f,t})]$$

where $\text{E}[r_{M,t}]$ is the expected return on a *well-diversified equity market portfolio* and $r_{f,t}$ is the risk-free rate.

11

Returns: Expected Returns & The ERP

• Components of $\text{E}[(r_{M,t} - r_{f,t})]$, the ERP:

(1) The risk-free rate, $r_{f,t}$. It is approximated by the mean yield of government securities, typical examples, 3-month U.S. Treasury bill or 10-year U.S. Treasury bond rate. In general, given the upward sloping terms structure, using T-bill yields gives higher ERP estimates.

(2) Expected Market Return. To determine $\text{E}[r_{M,t}]$, we need to determine the market portfolio. In theory, the market portfolio represents **all** risky securities, not just domestic, but in the world. Returns on this market portfolio should be measured free of survivor bias. In general, market-weighted liquid portfolios are preferred.

In general, we approximate ("proxy") the Market Portfolio, with a well-diversified index, for example, the S&P 500, MSCI World Index, or Weighted Average of CRSP returns.

12

Returns: Expected Returns & The ERP

• In finance, the ERP is central to many financial theories, for example, the CAPM. The CAPM states that the expected excess return on asset i is proportional to the ERP:

$$E[r_{i,t} - r_{f,t}] = \beta_i \, E[(r_{M,t} - r_{f,t})]$$

where

$r_{i,t}$ = return on asset *i* at time t.

$r_{f,t}$ = return of riskless asset at time t.

$r_{M,t}$ = return on the market portfolio at time t. (Here, the Market portfolio represents all wealth, it includes not only equity, but bonds, real estate, collectibles, etc; usually proxied by a market equity index.)

β_i = proportionality factor, it measures the sensitivity to market (systematic) risk.

Firms routinely use the CAPM to compute the cost of equity (=*required return on equity*) as an input in the WACC (cost of capital). 13

Returns: Expected Returns & The ERP

• Q: How do we calculate $E[r_{M,t}]$?

There are three different ways to compute $E[r_{M,t}]$:

1) Surveys. Usually an average of ERPs provided by individual investors, institutional investors, managers, and, even, academics.

2) Historical data. Expectations are computed using past data. This is the most popular approach. For example, compute $E[r_{M,t}]$ with $\bar{X}$. If we use this approach, it pays to use as much data as possible –more data, lower S.E. We think of $E[r_{M,t}]$ as a *long-run* average of market returns.

3) Forward-looking data. An (implied) ERP is derived from market prices, for example, market indexes, options & futures on market indexes, etc. Of course, we also need a model (a formula) that extracts the ERP from market prices. 14

Returns: Expected Returns & The ERP

• Once we compute E[$r_{M,t}$] and chose a corresponding $r_{f,t}$, we are ready to determine the ERP. But, we make decisions along the way.

Example: We use Shiller's monthly data, with 150 years of data, to produce an estimate of the ERP = E[($r_{M,t}$ - $r_{f,t}$)]:

Annualized Market return = **0.007378** * **12** = **0.088536**

Annualized risk-free rate = **0.04511**

ERP = **0.088536** - **0.04511** = **0.043426** (**4.34%**)

Decisions made:

- Computation of returns (log returns)
- Method of computing ERP (Historical data)
- Sample period (1871-2021)
- Market portfolio (S&P Composite Index)
- Risk-free rate (10-year U.S. bond rate).

15

Returns: Expected Returns & The ERP

• We compute a **4.34%** annualized ERP = E[($r_{M,t}$ - $r_{f,t}$)]:

ERP = **0.088536** – **0.04511** = **0.043426** (**4.34%**)

• Damodaran (2024) computes an annual update of the ERP, using data from 1928. He estimates ERP = **5.23%** (geometric returns).

• Many economists would consider these estimated ERPs as "*too high*." Why? In the standard "neoclassical" theoretical model, the degree of risk aversion to justify it is unreasonable high.

• The standard "neoclassical" model gets an ERP close to **1%.**

16

Returns: The Equity Risk Premium Puzzle

• A **too high** (for economic theory) ERP was first reported by Mehra & Prescott (1985), which they estimated around **6%.** They labeled the incompatibility of theory & observed data the *equity risk premium puzzle*.

• There have been many attempts to explain the puzzle:

- Statistical artifact (estimation, survivor bias)
- Disaster insurance (peso problem/sample period)
- Taxation
- Model's preferences (standard model's preferences too simple)
- Behavioral issues (mainly, myopic loss aversion & overreactions).

17

Returns: The Equity Risk Premium Puzzle

• Damodaran (2024), who produces an annual update of the literature and the ERP estimates, said in an overview of all possible explanations:

"It is true that historical risk premiums are higher than could be justified using conventional utility models for wealth. However, that may tell us more about the dangers of using historical data and the failures of classic utility models than they do about equity risk premiums."

18

Returns: Equity Risk Premium – Wide Range

• Is it **4%**, **5%**, or **6%**? Not clear. Even with 100+ years of data for developed markets there is no consensus on an ERP. For example, for the U.S. market, one of the best in terms of data quality, Duarte and Rosa (2015) list over 20 approaches ("models") to estimate ERP.

• With 1**960-2013** data, D&R (2015) report estimates from **-0.4%** to **13.1%,** with a **5.7%** average for all models. A wide range!

• The next slide, with **50 years of monthly data** of log returns, reports annual ERP estimates for developed markets (in USD), from **0.88%** (Italy) to **11.56%** (HK), using the **sample mean return** for the MSCI country index & the average U.S. T-Bill rate for the period (≈ **4.50%**).

For emerging markets, there also big dispersion of ERP estimates (also big SDs!)

19

Returns: Equity Risk Premium – Wide Range

MSCI Index USD Equity Returns and ERP: (1970-2021)

Market (T=620)	**Equity Return**	**Standard Deviation**	**ERP**
U.S.	8.31	**15.01**	**0.0382**
Canada	7.95	19.21	0.0346
France	8.80	21.95	0.0431
Germany	8.80	21.48	0.0431
Italy	**5.37**	**25.25**	**0.0088**
Switzerland	10.34	17.64	0.0585
U.K.	7.37	21.20	0.0288
Japan	9.56	20.46	0.0506
Hong Kong	**16.06**	**33.23**	**0.1156**
Singapore	11.71	27.48	0.0722
Australia	7.35	23.42	0.0273
World	**7.66**	**14.54**	**0.0317**
EAFE	7.69	16.64	0.0306

20

Returns: Equity Risk Premium – Wide Range

MSCI Index USD Equity Returns and ERP: (1970-2021)

Market (T)	Equity Return	Standard Deviation	ERP
Argentina (404)	24.21	51.49	0.1972
Brazil (404)	**22.23**	**47.67**	**0.1774**
Mexico (404)	17.67	29.26	0.1318
Poland (344)	15.88	43.24	0.1139
Russia (320)	**21.09**	**47.54**	**0.1660**
India (344)	**12.10**	**28.35**	**0.0760**
China (344)	**4.90**	**31.94**	**0.0041**
Korea (404)	11.75	34.08	0.0726
Thailand (404)	11.58	32.24	0.0606
Egypt (320)	11.61	31.69	0.0862
South Africa (344)	9.47	26.31	0.0498
World (620)	**7.66**	**14.54**	**0.0317**
EM Asia	8.85	23.13	0.0436

21

Returns: Equity Risk Premium – Precision

• We use the SE as a measure of precision of an estimate. For the sample mean, $\bar{X}$, we have:

$$\text{S.E.}(\bar{X}) = {}^{s}\!/_{\sqrt{T}}$$

• Using the previous data, we calculate the S.E.($\bar{X}$) for several markets:

U.S.: **15.01**/sqrt(620/12) = 2.0882%

Germany: 21.48/sqrt(620/12) = 2.9883%

Hong Kong: **33.23**/sqrt(620/12) = 4.6230 % ⟸ Effect of T

Brazil: **47.67**/sqrt(404/12) = 8.2157 %

Russia: **47.54**/sqrt(320/12) = 9.2061%

India: **28.35**/sqrt(344/12) = 5.2950%

China: **31.94**/sqrt(344/12) = 5.9654%

A big difference in precision between Developed and EM.

22

Returns: Risk-Return – The Sharpe Ratio

• The most commonly cited statistics that provides a measure of the risk-return trade-off for an asset is the Sharpe ratio (SR), the ratio of the excess expected return of an asset to its risk, measured by its return volatility (SD). We estimate the SR of asset i with

$$\widehat{SR}_i = \frac{\hat{\mu}_i - r_f}{s_i} \qquad (\hat{\mu}_i = \bar{X} \text{ for the return of } i)$$

Interpretation: A 1% change in risk, increases excess returns by SR%.

Using the previous data, we calculate the SR for several markets:

U.S.:	**0.0382**/**.1501** = 0.2545	
Switzerland:	0.0585/**.1764** = 0.3316	
Hong Kong:	**0.1156**/**.3323** = 0.3479	
Russia:	**0.1660**/**.4754** = 0.3492	⟸ Best trade-off!
India:	**0.0760**/**.2835** = 0.2681.	

23

Hypothesis Testing

• Testing involves the comparison between two competing hypothesis (sometimes, they represent partitions of the world).

– H_0: The maintained hypothesis.

– H_1: The hypothesis considered if H_0.

• Idea: We collect a sample, $\boldsymbol{X} = \{X_1, X_2, \ldots, X_N\}$. We construct a statistic $T(X) = f(X)$, called the *test statistic*. Now we have a decision rule:

– If $T(X)$ is contained in space R, we reject H_0 (& we learn).

– If $T(X)$ is in the complement of R (R^C), we fail to reject H_0.

We call R, the *Rejection Region*. That is,

$$\text{If } T(X) \in R \qquad \Rightarrow \text{Reject } H_0$$

Note: $T(X)$, like any other statistic, is a RV. It has a distribution. We use the distribution of $T(X)$, under H_0, to determine R.

24

Hypothesis Testing: Rejection Region

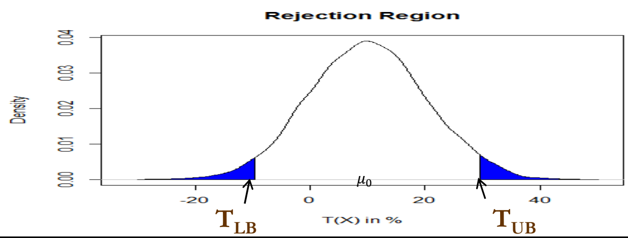
Example: Suppose we want to test H_0: $\mu = \mu_0$. Steps:

1) Collect data, $\{X_1, X_2, X_3, \dots, X_N\}$.

2) Construct statistic $T(X)$, which follows a known distribution under H_0. Suppose the distribution is normal.

3) Build a rejection region, R, in such a way that R contains α% of the area under the Normal distribution that $T(X)$ follows under H_0. Then, for T(X) = $\bar{X}$:

$$R = [\bar{X} < T_{LB}, T_{UB} > \bar{X}] \qquad \text{such that } P[R|H_0] = \alpha.$$



25

Hypothesis Testing: Rejection Region

• We call the blue area "*significance level*" (α%). If H_0 is true, the blue area represents the probability of rejecting a true H_0 or P[*Reject*|H_0].

• The significance level is arbitrary, we select it. Though, there are "standard" α's: 1%, 5%, 10%. The most popular one is **5%.**

• We impose on the distribution of $T(X)$ the values derived from H_0. For example, if $T(X) \sim N(\mu, .20)$. Then, under H_0, $T(X) \sim N(\mu_0, .20)$.

Remark: We determine T_{LB} & T_{UB} in such a way that the probability of rejecting H_0 when it is true –i.e., when $\mu = \mu_0$- is equal to α. In practice, since we try to avoid rejecting a true H_0, we usually set α (= P[R|H_0]) equal to a small number.

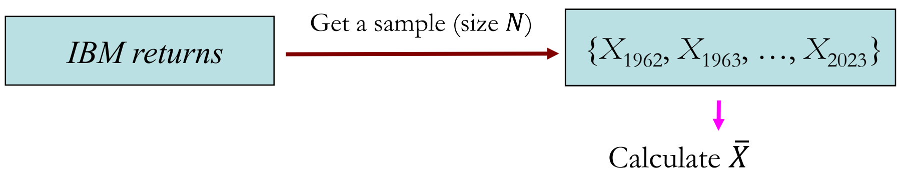
26

Hypothesis Testing: Rejection Region

Example: Suppose we want to test if the mean of IBM annual returns, μ_{IBM}, is 10%. That is, H_0: $\mu_{IBM} = \mu_0 = 10\%$.

1) Get a sample: $\{X_{1962}, X_{1963}, \ldots, X_{N=2024}\}$, with $N = 63$.

2) Compute $\bar{X}$, which is unbiased, consistent, &, assuming X is normally distributed, we know $\bar{X} \sim N(\mu, \sigma^2/N)$. (Set $\sigma = .15$.)



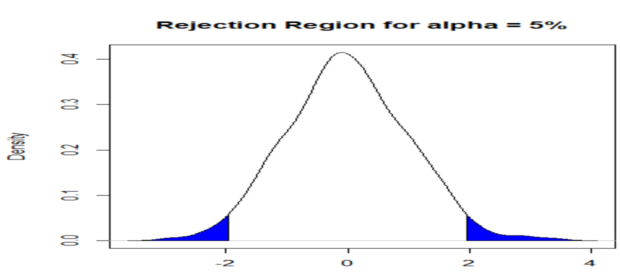
We standardize $\bar{X}$: $T(X) = (\bar{X} - \mu)/(\sigma/N) = (\bar{X} - .10)/(.15/\sqrt{63})$, which, under H_0, follows a N(0, 1).

Hypothesis Testing: Rejection Region

Example (continuation):

3) We set α & define R. Since $T(X) = (\bar{X} - .10)/(.15/\sqrt{63}) \sim N(0,1)$. Then, by setting α equal to 5% we get

$R = [T(X) < -1.96, 1.96 > T(X)]$ (that is, $P[R|H_0] = .05$.)



• Recall the above blue area represents the $P[Reject|H_0]$. In this example, the probability of rejecting H_0: $\mu_{IBM} = 10\%$. when it is true. 28

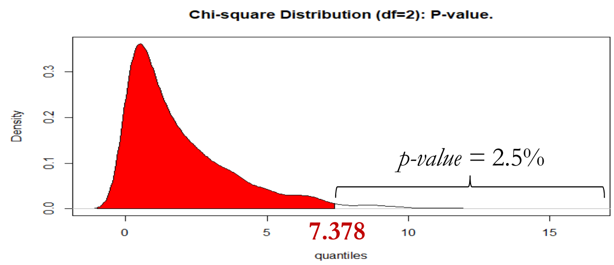
Hypothesis Testing: *p-value*

• We present the *classical approach*, which relies on the concept of *p-value*.

p-value is the probability of observing a result at least as extreme as the test statistic, under H_0.

Example: Suppose $T(X) \sim \chi^2_2$. We compute $\widehat{T(X)}$ = **7.378**. Then,

p-value($\widehat{T(X)}$ = **7.378**) = 1 – Prob[$T(X)$ < **7.378**] = 0.025



29

Hypothesis Testing: Steps

• Steps for the *classical approach*, also known as *significance testing*:

1. Identify H_0 & decide on a *significance level* (α% = P[R|H_0]) to compare your test results.

2. Determine the appropriate test statistic T(X) and its distribution under the assumption that H_0 is true.

3. Calculate T(X) from the data.

4. Rule: Reject H_0 if the *p-value* is sufficiently small, that is, we consider T(X) in R (we learn). Otherwise, we reach no conclusion (no learning).

Note: In Step 4, setting α% is equivalent to setting R.

• Q: What *p-value* is "sufficiently small" as to warrant rejection of H_0? Under this approach: smaller than α (usually, α = 5% or 1%).

30

Hypothesis Testing: Steps

Example: From the U.S. Jury System

1. Identify H_0 & set a *significance level* ($\alpha\% = P[R|H_0]$)

   H_0: The defendant is not guilty

   H_1: The defendant is guilty

Significance level α = "*beyond reasonable doubt*," presumably small level.

2. After judge instructions, each juror forms an "innocent index" $T(X)_i$.

3. Through deliberations, jury reaches a conclusion $T(X) = \sum_{i=1}^{12} T(X)_i$.

4. Rule: If *p-value* of $T(X) < \alpha \Rightarrow$ Reject H_0. That is, guilty!

   If *p-value* of $T(X) > \alpha \Rightarrow$ Fail to reject H_0. That is, non-guilty.

   Alternatively, we build a rejection region around H_0.

Note: Mistakes are made. We want to quantify these mistakes.

Hypothesis Testing: Error Types

• Type I and Type II errors

A *Type I error* is the error of rejecting H_0 when it is true.

A *Type II error* is the error of "accepting" H_0 when it is false (that is, when H_1 is true).

Example: From the U.S. Jury System

Type I error is the error of finding an innocent defendant guilty.

Type II error is the error of finding a guilty defendant not guilty.

• There is a trade-off between both errors. Traditional view: Set *Type I error* equal to a small number & find a test that minimizes *Type II error*.

The usual tests (t-tests, F-tests, Likelihood Ratio tests) incorporate this traditional view.

Hypothesis Testing: z-test & t-test

• For inferences about the population mean, the usual test statistic is the t-test. It is a modification of the z-test statistic.

• **z-test.** Assuming $\{X_1, X_2, X_3, \ldots, X_N\}$ is generated by a N(μ, σ^2), then, the sampling distribution of the sample mean is:

$$\bar{X} \sim N(\mu, \sigma^2/N).$$

Using the CLT, the distribution of the standardize sample mean, z, is:

$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{N}} \sim N(0, 1)$$

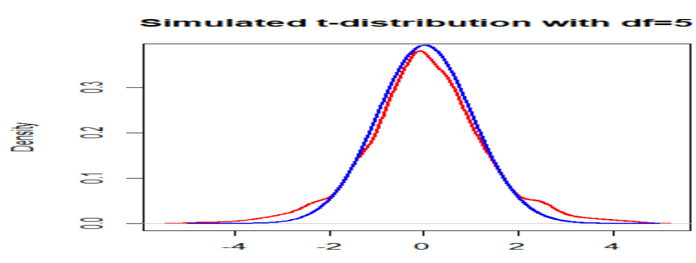
• **t-test.** In practice, σ is unknown. We need to estimate it, which we do with use s. Then, keeping the assumption $\{X_i\} \sim N(\mu, \sigma^2)$:

$$t = \frac{\bar{X} - \mu}{s/\sqrt{N}} \sim t_{N-1} \qquad \text{–when } N > 30,\ t_N \sim N(0, 1).$$

33

Hypothesis Testing: z-test & t-test

• Below, we plot a simulated t-distribution with ν =5 (in red), along a normal distribution (in blue). It has thicker tails. As ν increases, t_ν converges to a N(0, 1) distribution.



Technical Note: The distribution of t is exact if $\{X_i\} \sim N(\mu, \sigma^2)$, otherwise, the distribution is asymptotic (for large N). That is,

$$t = \frac{\bar{X} - \mu}{s/\sqrt{N}} \xrightarrow{d} N(0, 1).$$

34

Hypothesis Testing: H_0: $\mu = \mu_0$

Example: We want to test if the mean is equal to μ_0. Then,

1. H_0: $\mu = \mu_0$.
   H_1: $\mu \neq \mu_0$.

2. Appropriate T(X): *t-test* (based on σ unknown and estimated by s).
   Determine distribution of T(X) under H_0. Assuming $\{X_i\}$ ~ N(μ, σ^2), the distribution of T(X) under H_0:

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{N}} \sim t_{N-1} \qquad \text{–when } N > 30,\ t_N \sim N(0, 1).$$

3. Compute t, $\hat{t}$, using $\bar{X}$, μ_0, s, and N. Get *p-value*($\hat{t}$).

4. Rule: Set an α level. If *p-value*($\hat{t}$) < α ⇒ Reject H_0: $\mu = \mu_0$.
   Alternatively, if $|\hat{t}| > t_{N-1,(1-\alpha)/2}$ ⇒ Reject H_0: $\mu = \mu_0$.

35

Hypothesis Testing: Are Excess Returns Zero?

Example 1: We test if the **S&P 500 monthly excess return** is zero.
Data (1871-2021): $\bar{X}$ = **0.003619**, s = **0.04052**, N = 1805. Then,

1. H_0: $\mu = 0$.
   H_1: $\mu \neq 0$.

2. Appropriate test: $t = \frac{\bar{X} - \mu_0}{s/\sqrt{N}}$

3. $\hat{t} = \frac{0.003619}{0.04052/\sqrt{1805}} = \mathbf{3.7941}$ & *p-value*($\hat{t}$) = **0.00015**

4. Rule: *p-value*($\hat{t}$) = **0.00015** < α = .05 ⇒ Reject H_0: $\mu = 0$.

Alternatively, $|\hat{t} = \mathbf{3.7941}| > t_{N-1,.975} = \mathbf{1.96}$ ⇒ Reject H_0: $\mu = 0$.

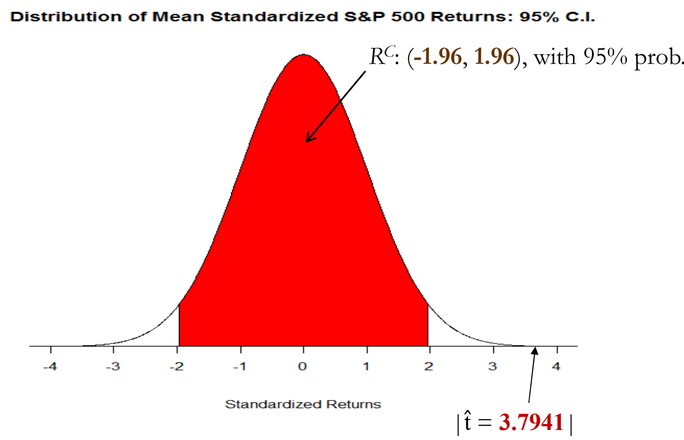
Conclusion: S&P 500 monthly mean excess returns are not equal to zero. ¶

36

Hypothesis Testing: Are Excess Returns Zero?

Example 1 (continuation): The observed $\hat{t}$ = **3.7941** is outside the non-rejection region built around H_0: R^C = (**-1.96**, **1.96**).

Note: $t_{1789, .975} = z_{.975}$ = **1.96** (since $N > 30$.)



37

Hypothesis Testing: Are Stock Returns Normal?

Example 2: We want to test if the monthly return of the S&P 500 follows a normal distribution. If the distribution is normal, skewness is zero and kurtosis is equal to 3 (or excess kurtosis equals 0). That is,

1. H_0 (Data is normal): $\gamma_1 = \frac{\mu_3^0}{\sigma^3} = 0$ and $\gamma_2 = \frac{\mu_4^0}{\sigma^4} - 3 = 0$.

   H_1 (Data is not normal): $\gamma_1 \neq 0$ and/or $\gamma_2 \neq 0$.

2. Appropriate T(X): the *Jarque-Bera test* (JB), $JB = \frac{N}{6} * (\gamma_1^2 + \frac{\gamma_2{}^2}{4})$

   Under H_0, $JB \xrightarrow{d} \chi_2^2$ (chi-square distribution with 2 *degrees of freedom*)

3. Data (1871-2020): $\hat{\gamma}_1 = -\mathbf{0.4705}$, $\hat{\gamma}_2$ = (14.6105 - 3), N = 1805,

   Then, $\widehat{JB} = \frac{1805}{6} * [(-0.4705)^2 + \frac{(11.6105)^2}{4}] =$ **10,204.89**

38

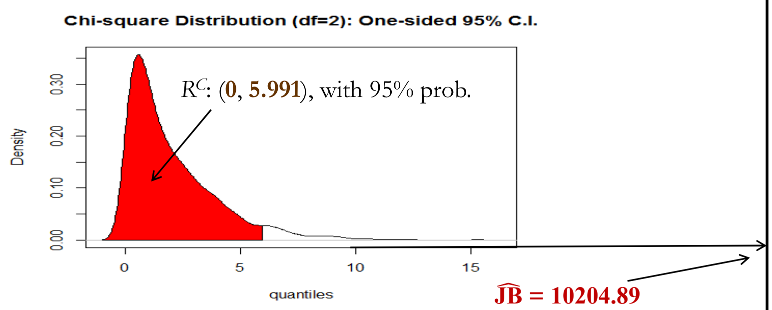
Hypothesis Testing: Are Stock Returns Normal?

Example 2 (continuation):

4. Rule: p-value($\widehat{JB}$ = **10,204.89**) ≈ 0 < α = .05 ⇒ Reject H_0.

Alternatively, compare $\widehat{JB}$ to the $\chi^2_{2,0.95}$ value ($\chi^2_{2,0.95}$ = **5.991**).

That is, $\widehat{JB} > \chi^2_{2,.95}$ ⇒ Reject H_0. (strong rejection!)



Conclusion: Monthly S&P 500 returns are not normally distributed. ¶ 39

Review – Confidence Intervals (C.I.)

• When we estimate parameters with an estimator, $\hat{\theta}$, we get a point estimate for θ, meaning that $\hat{\theta}$ is a single value in R^k. For example, in the previous example, $\bar{X}$ = **0.003619**.

• Broader concept: Estimate a set C_n, a collection of values in R^k. For example, μ ∈ {0.00155, 0.00554}.

• It is common to focus on intervals $C_n = [L_n; U_n]$, called an *interval estimate* for θ. The goal of C_n is to contain the true population value, θ. We want to see θ ∈ C_n, with high probability.

Technical detail: Since C_n is a function of the data, it is a RV and, thus, it has a pdf associated with it. The *coverage probability* of the interval $C_n = [L_n; U_n]$ is Prob[θ ∈ C_n].

40

Review – Confidence Intervals (C.I.)

• Intervals estimates C_n provide an idea of the uncertainty in the estimation of θ: The wider the interval C_n, the more uncertain we are about our estimate, $\hat{\theta}$.

• Interval estimates C_n are called *confidence intervals* (C.I.) as the goal is to set the coverage probability to equal a pre-specified target, usually 90% or 95%. C_n is called a (1 – α)% C.I.

• When we know the distribution for the point estimate, it is straightforward to construct a C.I. For example, if $\hat{\theta} \sim N(\theta, Var[\hat{\theta}])$, then a (1 – α)% C.I. is given by:

$$C_n = [\hat{\theta} + z_{\alpha/2} * \text{Estimated SE}(\hat{\theta}), \hat{\theta} + z_{(1-\alpha/2)} * \text{Estimated SE}(\hat{\theta})]$$

• This C.I. is symmetric around $\hat{\theta}$. Its length is proportional to SE($\hat{\theta}$). 41

Review – Confidence Intervals (C.I.)

• If $\hat{\theta} \sim N(\theta, Var[\hat{\theta}])$, then a (1 – α)% C.I. is given by:

$$C_n = [\hat{\theta} + z_{\alpha/2} * \text{Estimated SE}(\hat{\theta}), \hat{\theta} + z_{(1-\alpha/2)} * \text{Estimated SE}(\hat{\theta})]$$

• The z values are taken from the standard normal distribution, which is symmetric around 0. That is, $z_{(1-\alpha/2)} = -z_{\alpha/2} = |z_{\alpha/2}|$.

Thus, we can write the above (1 – α)% C.I. as:

$$C_n = [\hat{\theta} - z_{(1-\alpha/2)} * \text{Estimated SE}(\hat{\theta}), \hat{\theta} + z_{(1-\alpha/2)} * \text{Estimated SE}(\hat{\theta})]$$

• Popular values for α and z:

α = .10 ⇒ $z_{.95}$ = **1.645** ($z_{.05}$ = -**1.645**)

α = .05 ⇒ $z_{.975}$ = **1.96** ($z_{.025}$ = -**1.96**)

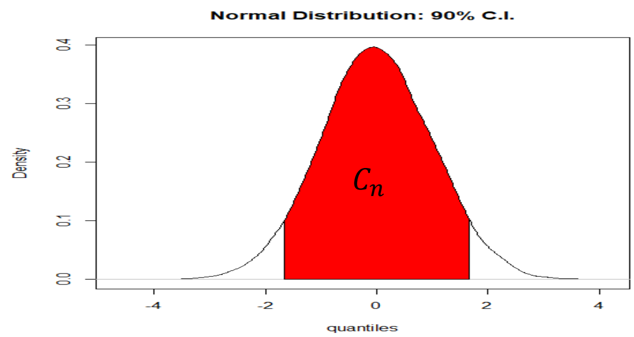
α = .02 ⇒ $z_{.99}$ = **2.33** ($z_{.01}$ = -**2.33**) 42

Review – Confidence Intervals (C.I.)

Example: Suppose $\hat{\theta}$ follows a Normal distribution. We estimate $\hat{\theta}$ to be very close to 0 ($\hat{\theta} \approx 0$) and estimated SE[$\hat{\theta}$] = 1. Then, a 90% C.I is given by

$$C_n = [0 - \mathbf{1.645} * 1, 0 + \mathbf{1.645} * 1] = [-\mathbf{1.645}, \mathbf{1.645}]$$

• C.I. is symmetric around 0, the length is proportional to 1 (=SE[$\hat{\theta}$]).

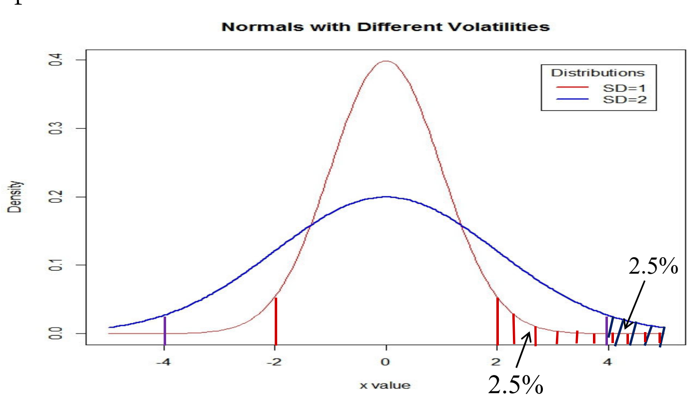


43

Review – Confidence Intervals (C.I.)

• For the mean, the size of the C.I. depends on the SD (=SE). The higher SD, the wider the C.I. ⇒ The wider, the higher uncertainty.

Example: Two 95% C.I. for the mean, with two different SD (=1, 2), are plotted below.



C.I. for Monthly Stock Excess Returns

Example: We estimate a 95% C.I. for the monthly S&P mean excess return. The sampling distribution of the sample mean (assuming normality) is $\bar{X} \sim N(\mu, \sigma^2/N)$, then, a $(1-\alpha)$% C.I. is given by:

$$C_n = [\bar{X} + z_{\alpha/2} * SD(\bar{X}),\ \bar{X} + z_{1-\alpha/2} * SD(\bar{X})]$$

Note: This C.I. is symmetric around $\bar{X}$. (Recall: $-z_{\alpha/2} = |-z_{\alpha/2}| = z_{1-\alpha/2}$)

Recall: $\bar{X}$ = **0.003619**, s = **0.04052**, N = **1805**.

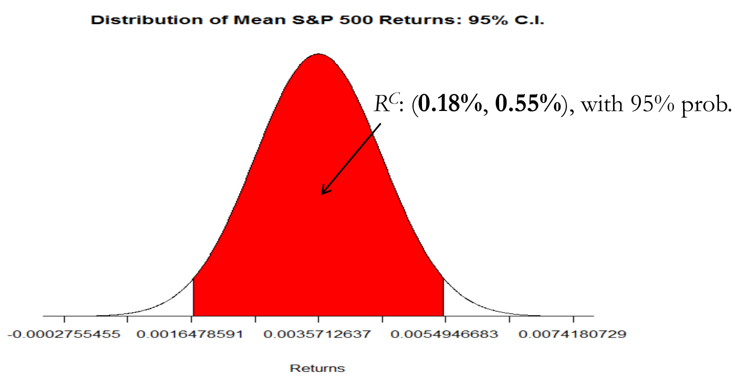
C_n = [**0.003619** – **1.96** * [**0.04052**/$\sqrt{1805}$], **0.003619** + **1.96** * [**0.04052**/$\sqrt{1805}$]
= [0.00175, 0.00549] = [**0.18%, 0.55%**]. (annual: [**2.16%, 6.60%**])

• By looking at the 95% C.I., we can reject that monthly S&P Composite excess returns are 0%, since 0% is outside the 95% C.I. But, the C.I is wide, even after 150 years of data. ¶

45

C.I. for Monthly Stock Excess Returns

Example (continuation): Reject H_0: $\mu = 0$, since 0 is outside the observed 95% C.I.



• Note: This 95% C.I. shows why Mehra & Prescott (1985) thought the ERP was *too high*. Their theoretical monthly ERP was, at most, 0.0833% (= 1%/12). ¶

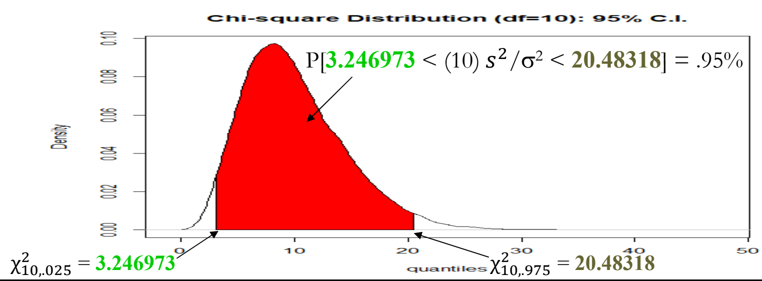
46

C.I. for the Variance of Monthly Stock Returns

Example: We want to estimate a 95% C.I. for the variance of **monthly S&P 500 mean total return.** Assuming normality, the sample variance, once scaled, is distributed: $(N-1)\, s^2/\sigma^2 \sim \chi^2_{N-1}$.

To derive a (1 - α)% C.I. for the variance, we rewrite the standard C.I. for a χ^2_υ variable:

$$P(\chi^2_{\upsilon,\alpha/2} < \chi^2_\upsilon < \chi^2_{\upsilon,1-\alpha/2}) = P(\chi^2_{\upsilon,\alpha/2} < (N\text{-}1)s^2/\sigma^2 < \chi^2_{\upsilon,1-\alpha/2}) = 1 - \alpha$$



47

C.I. for the Variance of Monthly Stock Returns

Example (continuation): the standard C.I. for a χ^2_υ variable:

$$P(\chi^2_{\upsilon,\alpha/2} < \chi^2_\upsilon < \chi^2_{\upsilon,1-\alpha/2}) = P(\chi^2_{\upsilon,\alpha/2} < (N\text{-}1)s^2/\sigma^2 < \chi^2_{\upsilon,1-\alpha/2}) = 1 - \alpha$$

After some algebra (recall inversion changes inequality signs), we derive:

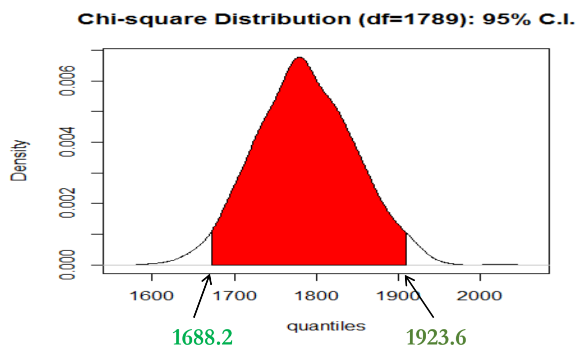
$$P[(N-1)\, s^2/\chi^2_{\upsilon,1-\alpha/2} < \sigma^2 < (N-1)\, s^2/\chi^2_{\upsilon,\alpha/2}] = 1 - \alpha.$$

Note: This C.I. is not symmetric. But, as the degrees of freedom, υ, get large, χ^2_υ starts to look like the normal distribution and, thus, CIs will look more symmetric.

48

C.I. for the Variance of Monthly Stock Returns

Example (continuation): From the χ^2_{1804} distribution, we get: $\chi^2_{1804,.025}$ = **1688.2** & $\chi^2_{1804,.975}$ = **1923.6**. (You get these values in R with qchisq(.025, df=N-1) & qchisq(.975, df=N-1), respectively.)



P[(1804) * (**0.04046**)^2/(**1923.6**) < σ^2 < (1804) * (**0.04046**)^2/(**1688.2**)] = .95

P[0.001535 < σ^2 < 0.001749] = .95

49

C.I. for the Variance of Monthly Stock Returns

Example (continuation):

P[0.001535 < σ^2 < 0.001749] = .95

Taking square root above delivers a 95% C.I. for σ:

⇒ 95% C.I. for σ is given by (**3.918%**, **4.182%**).

The C.I. is quite compact around the sample point estimate. Compared to the sample mean estimate, σ is measured with accuracy.

Note: Usually N is large. We can use the normal approximation to calculate CIs for the population σ. For the S&P data, we estimate the S.E. for the sample SD, s:

SE[s] = $s/\sqrt{2 * (N-1)}$ = **0.04046**$/\sqrt{2 * 1804}$ = 0.00067 (or .067%)

A 95% CI for σ is given by (**3.914%**, **4.178%**). (Very close!)

Derivation: (**0.04046** – **1.96** * 0.00067, **0.04046** + **1.96** * 0.00067)

50

C.I. Application: Using the ED

• In the previous examples, we assumed that we knew the distribution of the data: Stock returns follow a normal distribution. What happens when the data follows an unknown distribution, *F*?

• We still can use $\bar{X}$ or s^2 as estimates of μ and σ^2, since by LLN both are consistent estimators. If we have a "large" N, we can also use the CLT to justify a C.I. based on a normal distribution.

• But, when we have an unknown *F* and N is not large enough or we suspect the normal approximation is not a good one, we still can build a C.I. for any statistic using a new method: a *bootstrap*.

• The *bootstrap* is a method for estimating the sampling distribution of a statistic, $\theta = \theta(x_1, x_2, \ldots, x_N)$, by resampling from the ED, where

$$x_1, x_2, \ldots, x_N \sim i.i.d.\ F \text{ (unknown)}$$

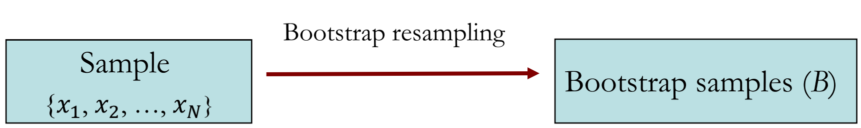
C.I. Application: Using the ED

Idea: We use the data at hand -the empirical distribution (ED)- to estimate the variation of a statistic θ that is itself computed from the same data. Recall that, for large samples drawn from *F*, the ED approximates the CDF of *F* very well.

• We need a *(bootstrap) Data Generating Process* to draw (simulated) data.

⇒ The bootstrap DGP estimates the unknown true DGP.

• The *bootstrap* is a resampling mechanism used to provide information about the sampling distribution of a statistic θ. Bootstrapping uses the ED –i.e., sample- as if it were the true CDF.



C.I. Application: Using the ED – The Bootstrap

• Suppose we have N *i.i.d.* observations drawn from $F(x)$:

$$\{x_1, x_2, \dots, x_N\}$$

From the ED, F^*, we sample with replacement N observations:

$$\{x_1^*, x_2^*, x_3^*, \dots, x_N^*\}$$

This is an *empirical bootstrap sample*, which is a resample of the same size N as the original data, drawn from F^*.

• For any statistic θ computed from the original sample data, we can define a statistic θ^* by the same formula, but computed instead using the resampled data.

• θ^* is computed by resampling the original data; we can compute many θ^* by resampling many times from F^*. Say, we resample θ^* B times.

C.I. Application: Using the ED – The Bootstrap

Example: You are interested in the relation between CEO's education (**X**) and firm's long-term performance (**y**). You have 1,500 observations on both variables. You estimate the correlation coefficient, ρ, with its sample counterpart, r. You find the correlation to be very low.

• Q: How reliable is this result? The distribution of r is complicated. You decide to use a bootstrap to study the distribution of r.

• Randomly construct a sequence of B samples (all with N=1,500). Say,

$B_1 = \{(x_1, y_1), (x_3, y_3), (x_6, y_6), (x_6, y_6), \dots, (x_{1458}, y_{1458})\} \Rightarrow \hat{\theta}_1^* = r_1$

$B_2 = \{(x_5, y_5), (x_7, y_7), (x_{11}, y_{11}), (x_{12}, y_{12}), \dots, (x_{1486}, y_{1486})\} \Rightarrow \hat{\theta}_2^* = r_2$

. . . .

$B_B = \{(x_2, y_2), (x_2, y_2), (x_2, y_2), (x_3, y_3), \dots, (x_{1499}, y_{1499})\} \Rightarrow \hat{\theta}_B^* = r_B$

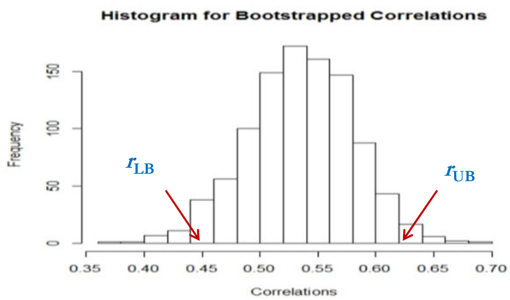
C.I. Application: Using the ED – The Bootstrap

• We have a collection of estimated θ^*:

$$\{\hat{\theta}_1^*, \hat{\theta}_2^*, \hat{\theta}_3^*, \dots, \hat{\theta}_B^*\}.$$

From this collection of $\hat{\theta}^*$'s, we can compute the mean, the variance, skewness, draw a histogram, etc., and confidence intervals.

• Below, we plot the histogram of bootstrapped correlations and defined a C.I. with [r_{LB}, r_{UB}].



C.I. Application: Using the ED – The Bootstrap

• Using the histogram or the sorted $\{\hat{\theta}_1^*, \hat{\theta}_2^*, \hat{\theta}_3^*, \dots, \hat{\theta}_B^*\}$, we can build a $(1-\alpha/2)$% C.I. Using the histogram, the lower bound leaves $\alpha/2$% of the $\hat{\theta}^*$ to the right and $(1-\alpha/2)$% of the $\hat{\theta}^*$ to the left.

• Bootstrap Steps:

1. From the original sample, draw random sample with size N.
2. Compute statistic θ from the resample in 1: $\hat{\theta}_1^*$.
3. Repeat steps 1 & 2 B times ⇒ Get B statistics: $\{\hat{\theta}_1^*, \hat{\theta}_2^*, \hat{\theta}_3^*, \dots, \hat{\theta}_B^*\}$
4. Compute moments, draw histograms, etc. for these B statistics.

C.I. Application: Using the ED – The Bootstrap

• Results (Bootstrap Principle):

1. With a large enough B, the LLN allows us to use the $\hat{\theta}^*$'s to estimate the distribution of $\hat{\theta}$, $F(\hat{\theta})$.

2. The variation in $\hat{\theta}$ is well approximated by the variation in $\hat{\theta}^*$.

Result 2 is the one that we use to estimate the size of a C.I.

• There are many ways to construct a C.I. using bootstrapping. The easier one is the one described above. Just use the distribution of the $\hat{\theta}^*$'s to compute directly a C.I. This is the *bootstrap percentile method*.

The percentile method uses the distribution of $\hat{\theta}^*$ as an approximation to the distribution of $\hat{\theta}$.

C.I. Application: Bootstrap Percentile Method

• Technical Note: The bootstrap delivers consistent results only.

Example: We construct a 95% C.I. for the variance of S&P 500 monthly returns (continuation of previous example). Set B = 1000. Using boot.ci function, with type=**perc**, from **boot** package (Need to install it first):

```
boot.samps <- boot(data=lr, statistic=var_p, R=sim_size)
boot.ci(boot.samps, type = "perc")
```

```
> boot.ci(boot.samps, type = "perc")
BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 1000 bootstrap replicates

CALL :
boot.ci(boot.out = boot.samps, type = "perc")

Intervals :
Level     Percentile
95%   ( 0.0014,  0.0020 )
Calculations and Intervals on Original Scale
```

C.I. Application: Bootstrap Percentile Method

Example (continuation):

• Check results by sorting boot.samps$t.

```
> new <- sort(boot.samps$t)
> new[25]                              # CI's Lower Bound
[1] 0.001398215
> new[975]                             # CI's Upper Bound
[1] 0.001955096
```

Or for σ, taking square root of the above bounds, the 95% CI is given by (**3.74%**, **4.42%**).

Note: Compare with 95% C.I. calculated under different assumptions:

- Assuming returns are Normal: (**3.918%**, **4.182%**).

- Using the CLT (N is large): (**3.914%**, **4.178%**).