Statistical Inference and Constant Expected Return Model

Econ 147

UCLA

Version 1.1

Probability and Statistics

- Probability Models (rv's) live in abstract space (unobservable)
  - mechanism generating numbers from probability models is called *sampling*
  - realization of sample (random variables) is called *data* (observable)
- Statistical inference is to infer probability space from data
  - study unobservable based on observable
  - typically impose some assumption on sampling procedure, e.g., IID
- Reading Eric Zivot's book chapter on **Constant Expected Return Model**
- Optional: Chapter 5 (**Modeling Univariate Distributions**) and Chapter 7 (**Multivariate Statistical Models**) in Ruppert's book

Statistical Inference on Finance Model

- For example, consider Constant Expected Return (CER) Model

$$r_t \sim \text{iid } N(\mu, \sigma^2) \quad t = 1, \cdots, T.$$

  - one example of probability model used in finance
  - hence a finance model
- Random (or IID) Sampling
- Parameters of interest: (μ, σ^2) or $f\left(\mu, \sigma^2\right)$

Statistical Inference

- Point Estimation
  - propose one number for our parameter of interest
- Interval Estimation (Confidence Interval)
  - propose an interval for our parameter of interest
- Hypothesis Testing
  - make a (statistical) judge for a given hypothesis

Parameter

- From CER model:

$$\mu = E[r_t]$$
$$\sigma^2 = \mathrm{var}(r_t)$$

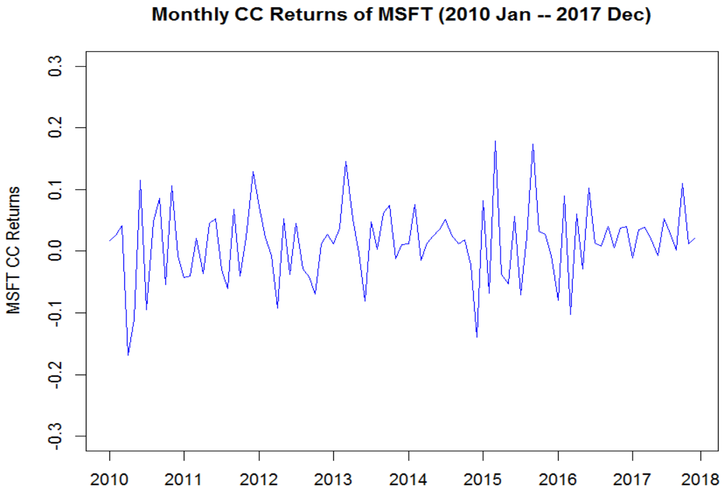
- Not known with certainty
- Estimate using observed sample (e.g., historical monthly returns)

Point Estimation: Mean

- Sample mean (or average) as estimator of $E[r_t] = \mu$

$$
\begin{aligned}
\{r_1, \ldots, r_T\} &= \text{ collection of random variables} \\
\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} r_t &= \text{ sample mean} \\
&= \text{ random variable}
\end{aligned}
$$

- $\hat{\mu}$ is a (point) estimator of μ

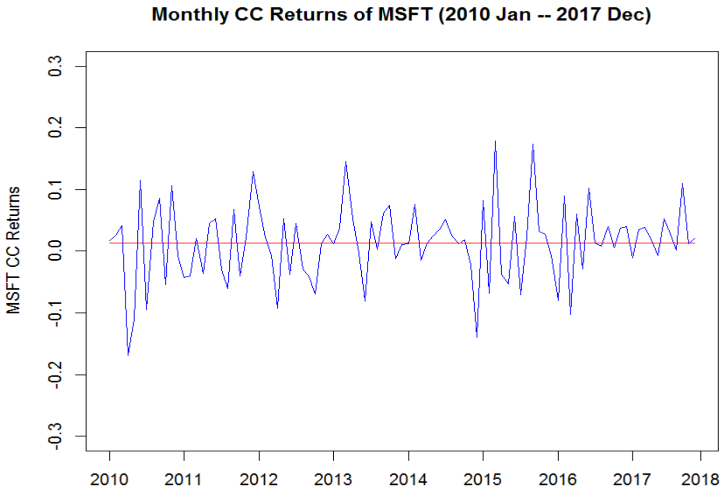

Monthly CC Returns of MSFT (2010 Jan -- 2017 Dec)
MSFT CC Returns
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
2010
2011
2012
2013
2014
2015
2016
2017
2018

Point Estimation: Mean

- Sample mean from an observed sample
- Example: MSFT cc return between January 2010 - December 2017 ($T = 95$)

$$
\begin{aligned}
\{r_1 = .017, r_2 = .026, \ldots, r_{95} = .021\} &= \text{observed sample} \\
\hat{\mu} &= \frac{1}{95}(.017 + .026 + \cdots + .021) \\
&= \text{ number } = 0.0139
\end{aligned}
$$

- 0.0139 is an estimate (or estimated value) of μ


Monthly CC Returns of MSFT (2010 Jan -- 2017 Dec)
MSFT CC Returns
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
2010
2011
2012
2013
2014
2015
2016
2017
2018

Properties of Estimators

- Unbiasedness
- Efficiency
- Mean Squared Error (MSE)
- Consistency

Properties of Estimators

- How good is 0.0139 (estimated value of $\hat{\mu}$) for μ?
- Estimation error:

$$error(\hat{\mu}, \mu) = \hat{\mu} - \mu$$

- Bias:

$$\text{bias}(\hat{\mu}, \mu) = E\left[error(\hat{\mu}, \mu)\right] = E\left[\hat{\mu}\right] - \mu$$

- $\hat{\mu}$ is *unbiased* if

$$E\left[\hat{\mu}\right] = \mu \Rightarrow \text{bias}(\hat{\mu}, \mu) = 0$$

- $\frac{1}{T}\sum_{t=1}^{T} r_t$ = sample mean is an *unbiased estimator* of μ. **Why?**

Properties of Estimators

- In general,

$$\theta = \text{parameter to be estimated}$$
$$\hat{\theta} = \text{estimator of } \theta \text{ from sample}$$

- Estimation error:
$$error(\hat{\theta}, \theta) = \hat{\theta} - \theta$$

- Bias:
$$\mathrm{bias}(\hat{\theta}, \theta) = E\left[error(\hat{\theta}, \theta)\right] = E\left[\hat{\theta}\right] - \theta$$

- $\hat{\theta}$ is *unbiased* if

$$E[\hat{\theta}] = \theta \Rightarrow \mathrm{bias}(\hat{\theta}, \theta) = 0$$

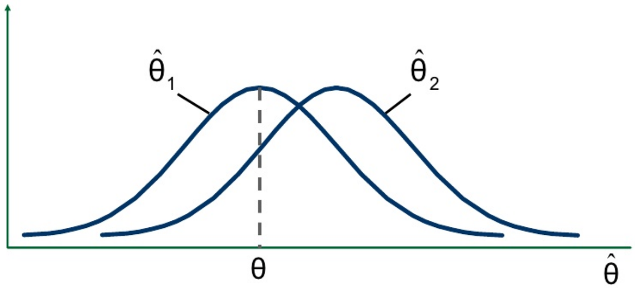
Properties of Estimators

- Unbiased estimator may be preferred to biased estimators.
- There might be an exception, however... see below!

Unbiasedness

(continued)

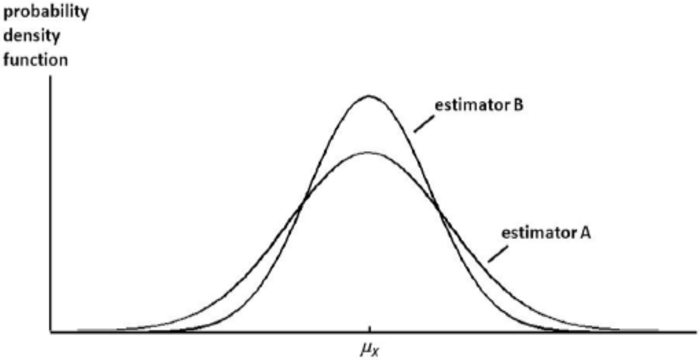
- $\hat{\theta}_1$ an unbiased estimator, $\hat{\theta}_2$ is biased:



Properties of Estimators

- If $\hat{\theta}_1$ and $\hat{\theta}_2$ are two unbiased estimators of θ, we prefer low variability
- *Relative efficiency:* $var\left(\hat{\theta}_1\right) < var\left(\hat{\theta}_2\right)$.

UNBIASEDNESS AND EFFICIENCY



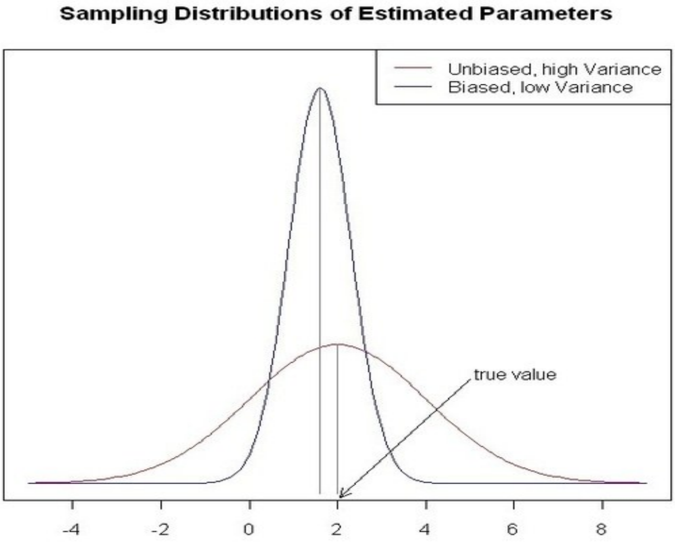
Properties of Estimators

- Precision by *Mean Squared Error*:

$$mse(\hat{\theta},\theta) = E\left[error(\hat{\theta},\theta)^2\right] = E\left[\left(\hat{\theta}-\theta\right)^2\right]$$
$$= \mathrm{bias}(\hat{\theta},\theta)^2 + \mathrm{var}(\hat{\theta})$$
$$\mathrm{var}(\hat{\theta}) = E[(\hat{\theta}-E[\hat{\theta}])^2$$

- If $\mathrm{bias}(\hat{\theta},\theta) \approx 0$ then precision is typically measured by the *standard error* of $\hat{\theta}$ defined by

$$\mathrm{SE}(\hat{\theta}) = \text{ standard error of } \hat{\theta}$$
$$= \sqrt{\mathrm{var}(\hat{\theta})} = \sqrt{E[(\hat{\theta}-E[\hat{\theta}])^2]}$$
$$= \sigma_{\hat{\theta}}$$


Sampling Distributions of Estimated Parameters
Unbiased, high Variance
Biased, low Variance
true value
-4
-2
0
2
4
6
8

Point Estimators

- Plug-in principle: Estimate parameters using appropriate sample statistics

$$\mu = E[r_t] : \hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} r_t$$

$$\sigma^2 = E[(r_t - \mu)^2] : \hat{\sigma}^2 = \frac{1}{T-1}\sum_{t=1}^{T}(r_t - \hat{\mu})^2$$

where $\hat{\mu}$ and $\hat{\sigma}^2$ are called sample mean and sample variance respectively.

Bias of Estimates

- $\hat{\mu}$ and $\hat{\sigma}^2$ are unbiased estimators:

$$E\left[\hat{\mu}\right] = \mu \Rightarrow \mathrm{bias}(\hat{\mu}, \mu) = 0$$
$$E\left[\hat{\sigma}^2\right] = \sigma^2 \Rightarrow \mathrm{bias}(\hat{\sigma}^2, \sigma^2) = 0$$

- Proofs for unbiasedness.

- Standard Error formulas for $\hat{\mu}$ and $\hat{\sigma}^2$

$$
\begin{aligned}
\mathrm{SE}(\hat{\mu}) &= \frac{\sigma}{\sqrt{T}}, \\
\mathrm{SE}(\hat{\sigma}) &\approx \frac{\sqrt{2}\sigma^2}{\sqrt{T}}
\end{aligned}
$$

Note: "$\approx$" denotes "approximately equal to", where approximation error $\longrightarrow 0$ as $T \longrightarrow \infty$.

- Practically useful formulas replace unknown values with estimated values:

$$\widehat{\mathrm{SE}}(\hat{\mu}) = \frac{\hat{\sigma}}{\sqrt{T}}, \ \hat{\sigma} \text{ replaces } \sigma$$

$$\widehat{\mathrm{SE}}(\hat{\sigma}^2) \approx \frac{\sqrt{2}\hat{\sigma}^2}{\sqrt{T}}, \ \hat{\sigma}^2 \text{ replaces } \sigma^2$$

- Such estimators are called plug-in estimators

Properties of Estimators

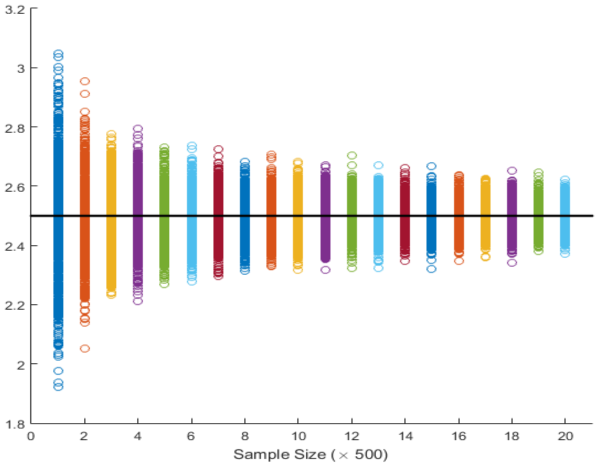
- Summary of desirable properties
  - unbiasedness
  - smaller MSE $(=\text{bias}(\hat{\theta}, \theta)^2 + \text{var}(\hat{\theta}))$
- Properties of Plug-in Estimates : good in general

Properties of Estimators

- Consistency
  - An estimator $\hat{\theta}$ is consistent for θ (converges in probability to θ) if for any $\varepsilon > 0$

    $$\lim_{T \to \infty} \Pr(|\hat{\theta} - \theta| > \varepsilon) = 0$$

    where n denotes the sample size.
  - Intuitively, as we get enough data then $\hat{\theta}$ will eventually equal θ.
- Consistency is an asymptotic property - it holds when we have an infinitely large sample.
- Graphical illustration?



- $\theta = 2.5$, $r_t \sim iid\ N(\theta, 4)$ and $\widehat{\theta} = T^{-1} \sum_{t=1}^{T} r_t$
- Sample sizes considered $T = 500, 1000, \ldots, 10000$
- For each sample size we consider 1000 simulated values for $\widehat{\theta}$

Properties of Estimators

- Fact: an estimator $\hat{\theta}$ is consistent for θ if
  - $\mathrm{bias}(\hat{\theta}, \theta) \to 0$ as $T \to \infty$ **AND** $\mathrm{SE}(\hat{\theta}) \to 0$ as $T \to \infty$
  - or, equivalently $mse(\hat{\theta}, \theta) \to 0$ as $T \to \infty$
  - proof by Markov ineq. (optional)
- Fact: the plug-in estimators $\hat{\mu}$, $\hat{\sigma}^2$ are consistent.
- Consistency of $\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} r_t$ is called *Law of Large Numbers* (LLN)

$$\lim_{T\to\infty} \Pr\left(\left| \frac{1}{T}\sum_{t=1}^{T} r_t - \mu \right| > \varepsilon \right) = 0$$

Estimation of VaR

- Let L_1 denote the profit of the investment:

$$L_1 = W_0(e^r - 1)$$

  where W_0 is the initial wealth and r is the cc return
- Then VaR_α is the α-quantile of the profit L_1

$$\text{VaR}_\alpha = W_0(e^{q_r^\alpha} - 1)$$

  where q_r^α is the α-quantile of r
- If $r \sim N(\mu, \sigma^2)$, and μ and σ^2 are known, then in R

$$q_r^\alpha = \textbf{qnorm(}\alpha, \mu, \sigma)$$

- When μ and σ^2 are unknown, we can estimate q_r^α as

$$\widehat{q}_r^\alpha = \textbf{qnorm(}\alpha, \widehat{\mu}, \widehat{\sigma})$$

  where $\widehat{\mu}$ is the sample mean and $\widehat{\sigma}^2$ is the sample variance.

Estimation of VaR

- The validity of the quantile estimator

$$\widehat{q}_r^\alpha = \textbf{qnorm}(\alpha, \widehat{\mu}, \widehat{\sigma})$$

  relies on the normal assumption. That is $r \sim N(\mu, \sigma^2)$
- Such estimator is called a parametric estimator because a parametric assumption (i.e., normal distribution) has been imposed on data
- In practice, nonparametric estimator (with does not assume a specific form for the pdf/pmf of data) is attractive
- the sample mean $\widehat{\mu}$ and sample variance $\widehat{\sigma}^2$ are nonparametric estimator (why?)
- Is there a easy to use nonparametric quantile estimator?

Estimation of VaR

- If we have data $\{r_1,\ r_2,\ \ldots,\ r_T\}$ from the same distribution (which is unknown), we can order them in the following way:

$$\{r_{[1]}, r_{[2]}, \ldots, r_{[T]}\}$$

  where $r_{[i]}$ is the i-th largest value in $\{r_1,\ r_2,\ \ldots,\ r_T\}$
- One popular nonparametric estimator of q_r^α is

$$\widehat{q}_r^\alpha = r_{[floor(T\alpha)]}$$

  where $floor(T\alpha)$ is the largest integer smaller than $T\alpha$.
- Therefore, a nonparametric estimator of VaR_α is

$$\text{VaR}_\alpha = W_0(e^{r_{[floor(T\alpha)]}} - 1)$$

- R function: **quantile(x, probs** $=$ α**, type** $=$ 1), where $\mathbf{x} =$ $\{r_1,\ r_2,\ \ldots,\ r_T\}$ is the data vector.

Estimation with MSFT Data

- Example: MSFT cc return (Jan 2010 - Dec 2017, $T = 95$)

$$\{r_1 = .017, r_2 = .026, \ldots, r_{95} = .021\} = \text{observed sample}$$

$$\hat{\mu} = \frac{1}{95}(.017 + .026 + \cdots + .021) = 0.0139;$$

$$\hat{\sigma}^2 = \frac{1}{95-1}\left((.017 - .0139)^2 + \cdots + (.021 - .0139)^2\right) = .00389$$

- The parametric estimator of $\mathrm{VaR}_{0.1}$ is

$$\widehat{\mathrm{VaR}}_{0.1}^{para} = W_0(e^{\mathbf{qnorm(}0.1, \widehat{\mu}, \widehat{\sigma})} - 1) = -639.4409$$

- The nonparametric estimator of $\mathrm{VaR}_{0.1}$ is

$$\widehat{\mathrm{VaR}}_{0.1}^{np} = W_0(e^{r_{[9]}} - 1) = -672.7413$$

- How good are these estimators?

Estimation with MSFT Data

- By definition, if L_1 is the profit of the 10,000 investment in the next month

$$\Pr\left(L_1 \leq \text{VaR}_{0.1}\right) = 0.1$$

- Therefore, we should have

$$\Pr\left(L_1 \leq \widehat{\text{VaR}}_{0.1}^{para}\right) \approx 0.1$$

and

$$\Pr\left(L_1 \leq \widehat{\text{VaR}}_{0.1}^{np}\right) \approx 0.1$$

- But how do we evaluate $\Pr\left(L_1 \leq \widehat{\text{VaR}}_{0.1}^{para}\right)$ and $\Pr\left(L_1 \leq \widehat{\text{VaR}}_{0.1}^{np}\right)$, given that L_1 is a random variable whose distribution is unknown?

Estimation with MSFT Data

- One idea is to calculate

$$L_{1t} = 10000 * \left(e^{r_t} - 1\right)$$

with the observed sample

$$\{r_1 = .017, r_2 = .026, \dots, r_{95} = .021\}$$

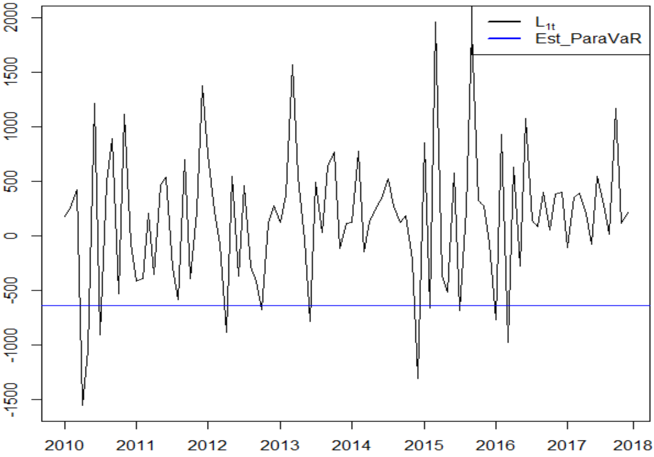
which will give us

$$\{L_{1,1} = 171.4532, L_{1,2} = 263.4095, \dots, L_{1,95} = 212.2205\}$$

- We count how many times that $L_{1,t}$ is smaller than $\widehat{\mathrm{VaR}}_{0.1}^{para}$ for $t = 1, \dots, 95$, let the number to be N_1
- Then we estimate $\Pr\left(L_1 \leq \widehat{\mathrm{VaR}}_{0.1}^{para}\right)$ using the relative frequency

$$\frac{N_1}{95} = \frac{11}{95} = 0.1158$$

$L_{1,t}$ $(t = 1, \dots, 95)$ and $\widehat{\text{VaR}}_{0.1}^{para}$



Estimation with MSFT Data

- Calculate

$$L_{1t} = 10000 * \left(e^{r_t} - 1\right)$$

with the observed sample

$$\{r_1 = .017, r_2 = .026, \ldots, r_{95} = .021\}$$

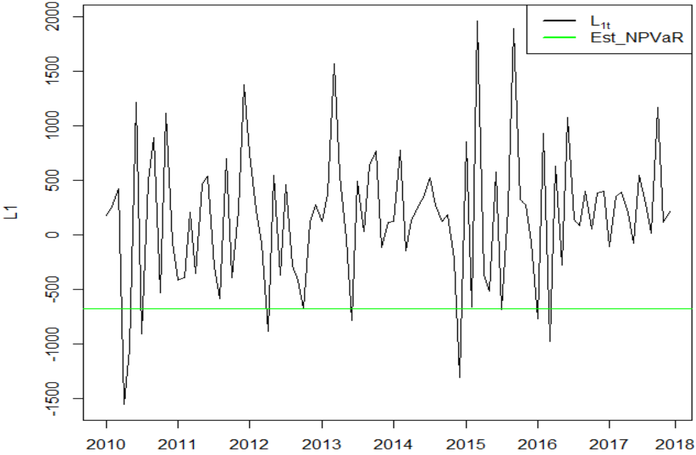
which will give us

$$\{L_{1,1} = 171.4532, L_{1,2} = 263.4095, \ldots, L_{1,95} = 212.2205\}$$

- We count how many times that $L_{1,t}$ is smaller than $\widehat{\text{VaR}}_{0.1}^{np}$ for $t = 1, \ldots, 95$, let the number to be N_2
- Then we estimate $\Pr\left(L_1 \leq \widehat{\text{VaR}}_{0.1}^{np}\right)$ using the relative frequency

$$\frac{N_2}{95} = \frac{9}{95} = 0.0947$$

$L_{1,t}$ $(t = 1, \ldots, 95)$ and $\widehat{\text{VaR}}_{0.1}^{np}$



Confidence Interval

- Point estimator:

$$\begin{aligned}\hat{\theta} &= \text{estimate of } \theta \\ &= \text{best guess for unknown value of } \theta\end{aligned}$$

- A confidence interval for θ is an interval estimate of θ that covers θ with a stated probability
- Need an (exact or asymptotic) distribution of $\hat{\theta}$ around θ

Confidence Interval for Mean

- Distribution of $\hat{\mu}$ with iid sampling from normal distribution

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} r_t, \;\; r_t \sim \text{iid } N(\mu, \sigma^2)$$

- Result: $\hat{\mu}$ is $\frac{1}{T}$ times the sum of T normally distributed random variables $\Rightarrow \hat{\mu}$ is also normally distributed with

$$E[\hat{\mu}] = \mu, \;\; \text{var}(\hat{\mu}) = \frac{\sigma^2}{T}$$

- That is,

$$\hat{\mu} \sim N\left(\mu, \frac{\sigma^2}{T}\right)$$

$$f(\hat{\mu}) = (2\pi\sigma^2/T)^{-1/2} \exp\left\{-\frac{1}{2\sigma^2/T}(\hat{\mu}-\mu)^2\right\}$$

Confidence Interval for Mean

- Since we know

$$\hat{\mu} \sim N\left(\mu, \frac{\sigma^2}{T}\right)$$

- By standardization,

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

- Thus,

$$\Pr\left(-1.96 \leq \frac{\hat{\mu} - \mu}{\frac{\sigma}{\sqrt{T}}} \leq 1.96\right) = 0.95$$

Confidence Interval for Mean

- By reverting $\frac{\hat{\mu}-\mu}{\frac{\sigma}{\sqrt{T}}}$ around μ, we get 95% confidence interval (CI) for μ

$$\left[\hat{\mu} - 1.96 \cdot \frac{\sigma}{\sqrt{T}},\ \hat{\mu} + 1.96 \cdot \frac{\sigma}{\sqrt{T}}\right]$$

- Plug-in:

$$\left[\hat{\mu} - 1.96 \cdot \frac{\hat{\sigma}}{\sqrt{T}},\ \hat{\mu} + 1.96 \cdot \frac{\hat{\sigma}}{\sqrt{T}}\right]$$

- Sometime a rough value 2 is used:

$$\left[\hat{\mu} - 2 \cdot \frac{\hat{\sigma}}{\sqrt{T}},\ \hat{\mu} + 2 \cdot \frac{\hat{\sigma}}{\sqrt{T}}\right]$$

Confidence Interval for Mean

- For MSFT cc return, $\hat{\mu} = 0.0139$, and $\hat{\sigma} = 0.0624$ for MSFT cc return with $T = 95$

$$\left[\hat{\mu} - 1.96 \cdot \frac{\hat{\sigma}}{\sqrt{T}},\ \hat{\mu} + 1.96 \cdot \frac{\hat{\sigma}}{\sqrt{T}}\right] = [0.0013,\ 0.0264]$$

Confidence Interval

- In reality, the exact distributions (for finite sample size T) of $\frac{\hat{\mu}-\mu}{\hat{\sigma}/\sqrt{T}}$ are not normal.
- However, as the sample size T gets large the exact distribution of $\frac{\hat{\mu}-\mu}{\hat{\sigma}/\sqrt{T}}$ get closer and closer to the normal distribution.
- This is due to the famous *Central Limit Theorem*

Theorem (Central Limit Theorem: CLT)

Let $X_1, \ldots, X_T$ be a iid random variables with $E[X_t] = \mu$ and $\mathrm{var}(X_t) = \sigma^2$. Then

$$\begin{aligned}\frac{\bar{X} - \mu}{\hat{\sigma}/\sqrt{T}} &\approx \frac{\bar{X} - \mu}{\sigma/\sqrt{T}} = \frac{\bar{X} - \mu}{\mathrm{SE}(\bar{X})} \\ &= \sqrt{T}\left(\frac{\bar{X} - \mu}{\sigma}\right) \sim N(0,1) \text{ as } T \to \infty\end{aligned}$$

- We say that $\bar{X}$ is asymptotically normally distributed with mean μ and variance $\mathrm{SE}(\bar{X})^2$.

Confidence Interval

Definition

An estimator $\hat{\theta}$ is asymptotically normally distributed if

$$\frac{\hat{\theta} - \theta}{\mathrm{SE}(\hat{\theta})} \text{ is approximately } N(0, 1), \text{ for large enough } T$$

- An implication of the CLT is that the estimators $\hat{\mu}$, $\hat{\sigma}^2$ are asymptotically normally distributed under iid sampling (but without assuming population distribution).

Confidence Interval

- Let $\hat{\theta}$ be an asymptotically normal estimator for θ.
- An approximate (or asymptotic) 95% confidence interval for θ is an interval estimate of the form

$$\left[\hat{\theta} - 2 \cdot \widehat{\mathrm{SE}}\left(\hat{\theta}\right),\ \hat{\theta} + 2 \cdot \widehat{\mathrm{SE}}\left(\hat{\theta}\right)\right]$$

$$\hat{\theta} \pm 2 \cdot \widehat{\mathrm{SE}}\left(\hat{\theta}\right)$$

that covers θ with probability approximately equal to 0.95. That is

$$\Pr\left\{\hat{\theta} - 2 \cdot \widehat{\mathrm{SE}}\left(\hat{\theta}\right) \leq \theta \leq \hat{\theta} + 2 \cdot \widehat{\mathrm{SE}}\left(\hat{\theta}\right)\right\} \approx 0.95$$

Confidence Interval

- An approximate 99% confidence interval for θ is an interval estimate of the form

$$\left[\hat{\theta} - 3 \cdot \widehat{\mathrm{SE}}\left(\hat{\theta}\right),\ \hat{\theta} + 3 \cdot \widehat{\mathrm{SE}}\left(\hat{\theta}\right)\right]$$
$$\hat{\theta} \pm 3 \cdot \widehat{\mathrm{SE}}\left(\hat{\theta}\right)$$

  that covers θ with probability approximately equal to 0.99.
- 99% confidence intervals are wider than 95% confidence intervals

Confidence Interval Parameters

- In iid sampling, (asymptotic) 95% confidence Interval for μ is

$$\hat{\mu} \pm 1.96 \cdot \frac{\hat{\sigma}}{\sqrt{T}}$$

$$\hat{\sigma}^2 \pm 1.96 \cdot \frac{\sqrt{2}\hat{\sigma}^2}{\sqrt{T}}$$

Confidence Interval for CER parameters

- From MSFT cc return,

$$\hat{\sigma}^2 \pm 1.96 \cdot \frac{\sqrt{2}\hat{\sigma}^2}{\sqrt{T}} = [0.00279,\ 0.00500]$$

where $\hat{\sigma}^2 = 0.00389$

Hypothesis Testing in IID sampling

$$r_{it} = \mu_i + \epsilon_{it} \quad t = 1, \cdots, T; \quad i = 1, \cdots N$$
$$\epsilon_{it} \sim \text{iid } N(0, \sigma_i^2)$$

- We want to know, for example:
  - Is $\mu_i = 0$ (expected return is zero) or $\mu_i > 0$ (positive expected return) ?
  - Is r_{it} really normally distributed?

Hypothesis Testing in IID Sampling

- Test for specific value

$$H_0 : \mu_i = \mu_i^0 \text{ vs. } H_1 : \mu_i \neq \mu_i^0$$

- Test for sign

$$H_0 : \mu_i = 0 \text{ vs. } H_1 : \mu_i > 0 \text{ or } \mu_i < 0$$

- Test for normal distribution

$$H_0 : r_{it} \sim \text{iid } N(\mu_i, \sigma_i^2)$$
$$H_1 : r_{it} \sim \text{not normal}$$

Hypothesis Testing

1. Specify hypothesis to be tested

$$H_0 : \text{ null hypothesis versus. } H_1 : \text{ alternative hypothesis}$$

2. Specify significance level of test

$$\text{level} = \Pr(\text{Reject } H_0 | H_0 \text{ is true})$$

3. Construct test statistic, T, from observed data
4. Use test statistic T to evaluate data evidence regarding H_0

$$\begin{aligned} T \text{ is big} &\Rightarrow \text{ evidence against } H_0 \\ T \text{ is small} &\Rightarrow \text{ evidence in favor of } H_0 \end{aligned}$$

Hypothesis Testing

- Decide to reject H_0 at specified significance level if value of T falls in the rejection region

$$T \in \text{ rejection region } \Rightarrow \text{ reject } H_0$$

- Usually the rejection region of T is determined by a critical value, cv, such that

$$\begin{aligned} T &> cv \Rightarrow \text{reject } H_0 \\ T &\leq cv \Rightarrow \text{ do not reject } H_0 \end{aligned}$$

Hypothesis Testing

	Reality	
Decision	H_0 is true	H_0 is false
Reject H_0	Type I error	No error
Do not reject H_0	No error	Type II error

- Significance Level of Test

$$\text{level} = \Pr(\text{Type I error}) = \Pr(\text{Reject } H_0 | H_0 \text{ is true})$$

  - Goal: construct test to have a specified small significance level: 5% or 1%

- Power of Test

$$1 - \Pr(\text{Type II error}) = \Pr(\text{Reject } H_0 | H_0 \text{ is false})$$

  - Goal: construct test to have high power

Hypothesis Testing for Mean

$$H_0 : \mu_i = \mu_i^0 \text{ vs. } H_1 : \mu_i \neq \mu_i^0$$

- Test statistic

$$t_{\mu_i = \mu_i^0} = \frac{\left| \hat{\mu}_i - \mu_i^0 \right|}{\widehat{\mathrm{SE}}(\hat{\mu}_i)}$$

  - If $t_{\mu_i = \mu_i^0} \approx 0$ then $\hat{\mu}_i \approx \mu_i^0$, and $H_0 : \mu_i = \mu_i^0$ should not be rejected
  - If $t_{\mu_i = \mu_i^0} > 2$, say, then $\hat{\mu}_i$ is more than 2 values of $\widehat{\mathrm{SE}}(\hat{\mu}_i)$ away from μ_i^0. This is very unlikely if $\mu_i = \mu_i^0$, so $H_0 : \mu_i = \mu_i^0$ should be rejected.

$$H_0 : \mu_i = \mu_i^0 \text{ vs.} H_1 : \mu_i \neq \mu_i^0$$

- Under the assumptions of iid normal sampling, and $H_0 : \mu_i = \mu_i^0$, the ratio
$$\frac{\hat{\mu}_i - \mu_i^0}{\widehat{\mathrm{SE}}(\hat{\mu}_i)} \sim t_{T-1}$$
where t_{T-1} is Student-t with $T-1$ degrees of freedom,
$$\hat{\mu}_i = \frac{1}{T}\sum_{t=1}^{T} r_{it},\ \widehat{\mathrm{SE}}(\hat{\mu}_i) = \frac{\hat{\sigma}_i}{\sqrt{T}},$$
$$\hat{\sigma}_i = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}(r_{it} - \hat{\mu}_i)^2}$$

- Under H_0,

$$\hat{\mu}_i = \frac{1}{T}\sum_{t=1}^{T} r_{it} \sim N\left(\mu_i^0, \frac{\sigma_i^2}{T}\right),$$

$$\frac{\hat{\mu}_i - \mu_i^0}{\frac{\sigma_i}{\sqrt{T}}} \sim N(0,1) := Z$$

- We can show,

$$(T-1)\frac{\hat{\sigma}_i^2}{\sigma_i^2} \sim \chi^2(T-1) := X$$

  and Z and X are independent.
- Therefore,

$$\frac{\frac{\hat{\mu}_i - \mu_i^0}{\frac{\sigma_i}{\sqrt{T}}}}{\sqrt{\frac{(T-1)\frac{\hat{\sigma}_i^2}{\sigma_i^2}}{(T-1)}}} = \frac{\hat{\mu}_i - \mu_i^0}{\frac{\hat{\sigma}_i}{\sqrt{T}}} = \frac{\hat{\mu}_i - \mu_i^0}{\widehat{\mathrm{SE}}(\hat{\mu}_i)} \sim t_{T-1}$$

Hypothesis Testing for Mean

1. Hypothesis to be tested

$$H_0 : \mu_i = 0 \text{ vs. } H_1 : \mu_i \neq 0$$

2. Significance level of test $= 5\%$
3. Test Statistic:

$$t_{\mu_i=0} = \left| \frac{\hat{\mu}_i - 0}{\widehat{\mathrm{SE}}(\hat{\mu}_i)} \right| = \frac{|\hat{\mu}_i|}{\widehat{\mathrm{SE}}(\hat{\mu}_i)}$$

where $\hat{\mu}_i / \widehat{\mathrm{SE}}(\hat{\mu}_i) \sim t_{T-1}$ under H_0

4. Decision rule:

$$\text{reject } H_0 : \mu_i = \mu_i^0 \text{ in favor of } H_1 : \mu \neq \mu_i^0 \text{ if}$$
$$t_{\mu_i=0} > cv = q_{.975}^{t_{T-1}} = \mathbf{qt}(.975,\ T-1)$$

Hypothesis Testing for Mean

We know, $\hat{\mu}_i = 0.0139$, and $\widehat{\mathrm{SE}}(\hat{\mu}_i) = 0.0624$ for MSFT cc return ($T = 95$)

1. Is expected MSFT cc return non-zero?
$$H_0 : \mu_i = 0 \text{ vs. } H_1 : \mu_i \neq 0$$
2. Significance level of test $= 5\%$
3. Test Statistic:
$$t_{\mu_i=0} = \left| \frac{\hat{\mu}_i - 0}{\widehat{\mathrm{SE}}(\hat{\mu}_i)} \right| = 2.1674 \text{ (t-ratio)}$$
4. Decision rule:
$$cv = q^{t_{94}}_{.975} = \mathbf{qt}(.975, 94) = 1.9855$$
$$\text{reject } H_0 : \mu_i = 0 \text{ in favor of } H_1 : \mu \neq 0$$
$$\text{because } t_{\mu_i=0} = 2.1674 > 1.9855 = cv$$

Tests based on CLT

Let $\hat{\theta}$ denote an estimator for θ. In many cases the CLT justifies the asymptotic normal distribution

$$\frac{\hat{\theta}-\theta}{\mathrm{SE}(\hat{\theta})} \text{ is approximately } N(0,1), \text{ for large enough } T$$

Consider testing

$$H_0 : \theta = \theta_0 \text{ vs. } H_1 : \theta \neq \theta_0$$

Result: Under H_0,

$$t_{\theta=\theta_0} = \left| \frac{\hat{\theta}-\theta}{\mathrm{SE}(\hat{\theta})} \right|$$

where

$$\frac{\hat{\theta}-\theta}{\mathrm{SE}(\hat{\theta})} \sim N(0,1)$$

for large sample sizes.

Testing for Sign

$$H_0 : \mu_i = 0 \text{ vs. } H_1 : \mu_i > 0$$

1. Test statistic

$$t_{\mu_i=0} = \frac{\hat{\mu}_i}{\widehat{\mathrm{SE}}(\hat{\mu}_i)}$$

- If $t_{\mu_i=\mu_i^0} \approx 0$ then $\hat{\mu}_i \approx 0$, and $H_0 : \mu_i = 0$ should not be rejected
- If $t_{\mu_i=\mu_i^0} >> 0$, then this is very unlikely under $\mu_i = 0$. So $H_0 : \mu_i = 0$ vs. $H_1 : \mu_i > 0$ should be rejected.

Testing for Sign

2. Set significance level and determine critical value

$$\Pr(\text{Type I error}) = 5\%$$

One-sided critical value cv is determined using

$$\Pr(t_{T-1} > cv_{.05}) = 0.05$$
$$\Rightarrow cv_{.05} = q_{.95}^{t_{T-1}}$$

where $q_{.95}^{t_{T-1}} = 95\%$ quantile of Student-t distribution with $T-1$ degrees of freedom.

3. Decision rule:

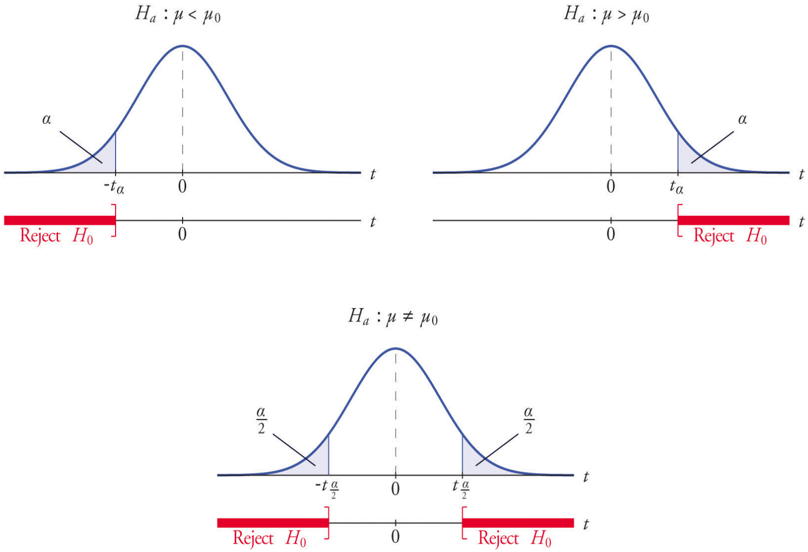
$$\text{Reject } H_0 : \mu_i = 0 \text{ in favor of } H_1 : \mu_i > 0$$
$$\text{at 5\% level if } t_{\mu_i=0} > q_{.95}^{t_{T-1}}$$

Testing for Sign

Useful Rule of Thumb:

4. If $T \geq 60$ then $q_{.95}^{t_{T-1}} \approx q_{.95}^{Z} = 1.645$ and the decision rule is

$$\text{Reject } H_0 : \mu_i = 0 \text{ in favor of } H_1 : \mu_i > 0$$
$$\text{at 5\% level if } t_{\mu_i = 0} > 1.645$$



$H_a : \mu < \mu_0$
α
$-t_\alpha$
0
t
Reject H_0
$H_a : \mu > \mu_0$
α
t_α
Reject H_0
$H_a : \mu \neq \mu_0$
$\frac{\alpha}{2}$
$\frac{\alpha}{2}$
$-t_{\frac{\alpha}{2}}$
$t_{\frac{\alpha}{2}}$
Reject H_0
Reject H_0

Testing for Normal Distribution

$$H_0 : r_t \sim \text{iid } N(\mu, \sigma^2)$$
$$H_1 : r_t \sim \text{ not normal}$$

1. Test statistic (Jarque-Bera statistic)

$$\text{JB} = \frac{T}{6}\left(\widehat{\text{skew}}^2 + \frac{(\widehat{\text{kurt}} - 3)^2}{4}\right)$$

See R package tseries function `jarque.bera.test`

- If $r_t \sim \text{iid } N(\mu, \sigma^2)$ then $\widehat{\text{skew}}(r_t) \approx 0$ and $\widehat{\text{kurt}}(r_t) \approx 3$ so that $\text{JB} \approx 0$
- If $\widehat{\text{skew}}(r_t) \neq 0$ and/or $\widehat{\text{kurt}}(r_t) \neq 3$ so that $\text{JB} >> 0$, then r_t must not be normally distributed

Testing for Normal Distribution

Distribution of JB under H_0
If $H_0 : r_t \sim \text{iid } N(\mu, \sigma^2)$ is true then

$$\text{JB} \sim \chi^2(2)$$

where $\chi^2(2)$ denotes a chi-square distribution with 2 degrees of freedom (d.f.)

Testing for Normal Distribution

Distribution of JB under H_0

2. Set significance level and determine critical value

$$\Pr\left(\text{Type I error}\right) = 5\%$$

Critical value cv is determined using

$$\Pr(\chi^2(2) > cv) = 0.05$$
$$\Rightarrow cv = q_{.95}^{\chi^2(2)} \approx 6$$

where $q_{.95}^{\chi^2(2)} \approx 6 \approx 95\%$ quantile of chi-square distribution with 2 degrees of freedom.

3. Decision rule:

$$\text{Reject } H_0 : r_t \sim \text{iid } N(\mu, \sigma^2) \text{ in favor of } H_1 : r_t \text{ is not normal}$$
$$\text{at } 5\% \text{ level if JB} > 6$$

Hypothesis Testing for Normal Distribution

Using the MSFT **monthly** cc returns ($T = 95$), we can calculate

$$\widehat{\text{skew}} = -0.1308 \text{ and } \widehat{\text{kurt}} = 3.654362$$

1. the JB test statistic

$$\text{JB} = \frac{T}{6}\left(\widehat{\text{skew}}^2 + \frac{(\widehat{\text{kurt}} - 3)^2}{4}\right) = 1.965725$$

2. Significance level of test $= 5\%$ and hence $cv = q_{.95}^{\chi^2(2)} \approx 6$
3. Decision rule:

$$\text{reject } H_1 : r_t \sim \text{ not normal}$$
$$\text{in favor of } H_0 : r_t \sim \text{iid } N(\mu, \sigma^2)$$
$$\text{because JB} = 1.965725 < 6 = cv$$

Hypothesis Testing for Normal Distribution

Using the MSFT **daily** cc returns (from Jan 2017 to Dec 2017, $T = 2012$), we can calculate

$$\widehat{\text{skew}} = -0.1109 \text{ and } \widehat{\text{kurt}} = 10.7645$$

1. the JB test statistic

$$\text{JB} = \frac{T}{6}\left(\widehat{\text{skew}}^2 + \frac{(\widehat{\text{kurt}} - 3)^2}{4}\right) = 5058.223$$

2. Significance level of test = 5% and hence $cv = q_{.95}^{\chi^2(2)} \approx 6$
3. Decision rule:

$$\text{reject } H_0 : r_t \sim \text{iid } N(\mu, \sigma^2)$$
$$\text{in favor of } H_1 : r_t \sim \text{ not normal}$$
$$\text{because JB} = 5058.223 > 6 = cv$$

Conclusion and Next Topic

- Probability and Statistics are essential tools to study finance models.
  - so far, IID assumption is maintained.
- Is IID assumption reasonable for financial market data? Very likely no.
- Next topic: time series econometrics.
  - another essential tool to study financial market.