Quantitative Analysis

Exam weight:20%

Preface

- **课程概述**
  - 理科的知识、文科考法
- **知识框架**
  - 至少覆盖大学期间五门以上学科知识
- **学习方法**
  - 理解与记忆相结合
  - 理解定理结论与相关概念性质，忽略证明过程
  - 难点不要转牛角尖
  - 多做题，掌握计算题的定式

Study Guide Change

- 与2017年相比，2018年考纲没有任何变化
- 考纲在2015年发生过巨变

CONTENTS

PART 2

CONTENTS

PART 2

Learning objectives

- Describe and distinguish between continuous and discrete random variables.
- Define and distinguish between the probability density function, the cumulative distribution function, and the inverse cumulative distribution function.
- **Calculate** the probability of an event given a discrete probability function.
- Distinguish between independent and mutually exclusive events.
- Define joint probability, describe a probability matrix, **and calculate joint probabilities** using probability matrices.
- Define and **calculate** a conditional probability, and distinguish between conditional and unconditional probabilities.

Basic Concepts

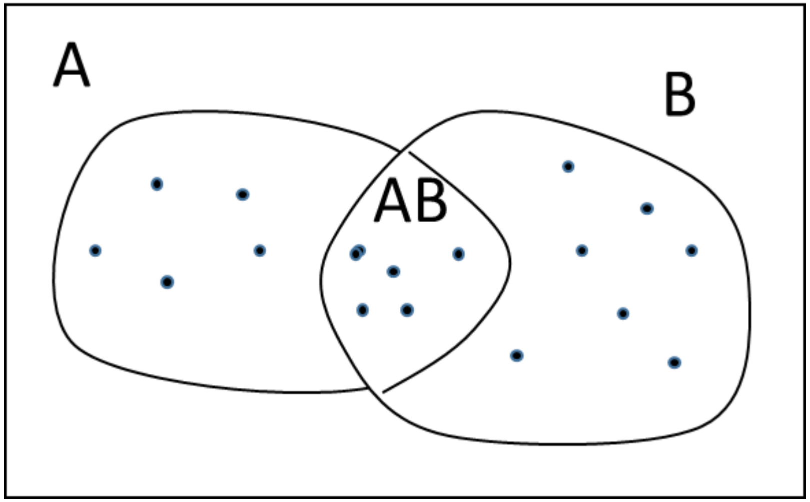
- **Random variable(随机变量)**
  - ✓An uncertain quantity/number.
- **Outcomes(结果)**
  - ✓An observed value of a random variable.
- **Event(事件)**
  - ✓A single outcome or a set of outcomes.

Relationship among events

- **Mutually exclusive events(互斥事件)**
  - ✓Events that cannot happen at the same time.
- **Exhaustive events(遍历事件)**
  - ✓Include all possible outcomes.
- **Two defining properties of a probability**
  - ✓The probability for any event E is: $0 \le P(E) \le 1$;
  - ✓For a set of events that are mutually exclusive and exhaustive, the sum of probabilities is 1: $\Sigma P(E_i) = 1$.

Two Important Rules

- Unconditional probability 无条件概率： P(A)
- Conditional probability 条件概率: P(A|B)
- Joint probability 联合概率: P(AB)



Distinguish and calculate (☆☆☆)

Two Important Rules

- **Multiplication rule**: P(AB)=P(A|B)×P(B)
  - ✓If A and B are mutually exclusive events, then:

$$P(AB)=P(A|B)=P(B|A)=0$$

- **Addition rule**: P(A or B)=P(A)+P(B)-P(AB)
  - ✓If A and B are mutually exclusive events, then:

$$P(A \text{ or } B)=P(A)+P(B)$$

Independent events

- **Definition:** The occurrence of B has no influence of on the occurrence of A

✓ $P(A|B)=P(A)$ or $P(B|A)=P(B)$

✓ $P(AB)=P(A)\times P(B)$

✓ $P(A \text{ or } B)=P(A)+P(B)-P(A)\times P(B)$

Example

If events A and B are mutually exclusive, then:

A. $P(A|B)=P(A)$

B. $P(A|B)=P(B)$

C. $P(AB)=P(A) \times P(B)$

D. $P(A \text{ or } B)=P(A)+P(B)$

Answer: D

Example

Two events are said to be independent if the occurrence of one event:

A. means the second event cannot occur.

B. means the second event is certain to occur.

C. affects the probability of the occurrence of the other event.

D. does not affect the probability of the occurrence of the other event.

Example

Answer: D

Two events are said to be independent if the occurrence of one event does not affect the probability of the occurrence of the other event.

Probability Matrix

- Joint probabilities of independent events can be conveniently summarized using a probability matrix (sometimes known as a probability table).
- The probability matrix in the table shows the joint and unconditional probabilities of these two variables.

		Interest Rates		
		Increase	No Increase	
Economy	Good	14%	6%	20%
	Normal	20%	30%	50%
	Poor	6%	24%	30%
		40%	60%	100%

Describe and distinguish (☆)

Discrete & Continuous Random Variable

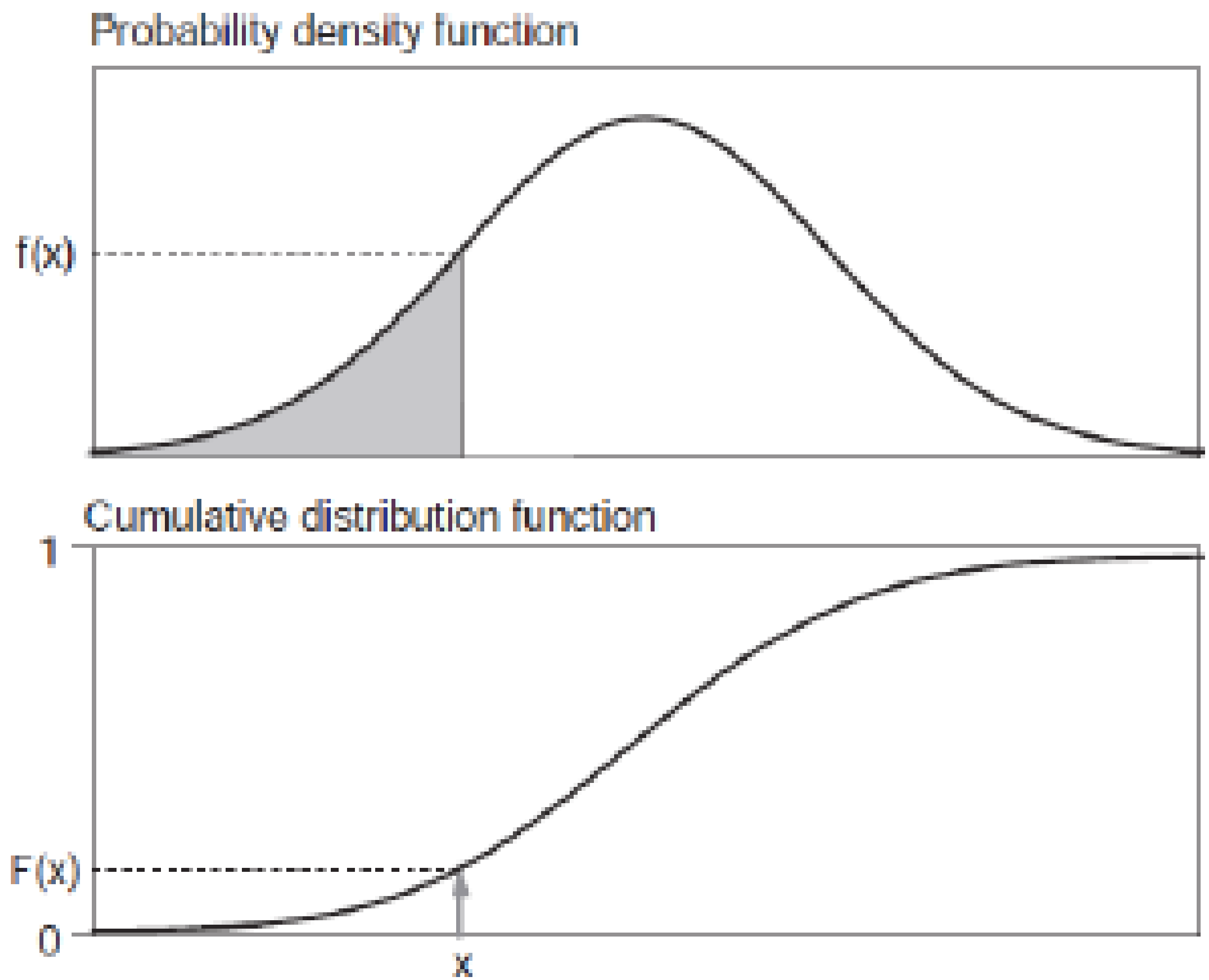
- **Discrete random variable**(离散随机变量):
  - ✓ Number of possible outcomes can be counted.
- **Continuous random variable**(连续随机变量):
  - ✓ Can take on any value within a given range
  - ✓ Outcomes is infinite even if lower and upper bounds exist.
  - ✓ Because outcomes is infinite, $P(X=x)=0$ even though x can occur.
  - ✓ 更关注区间概率—$P(x_1 \leq X \leq x_2) = P(x_1 < X < x_2)$

Probability distribution and function

- A **probability function, p(x),** gives the probability that a discrete random variable will take on the value x [e.g., $p(x) = x / 15$ for $X = \{1,2,3,4,5\} \rightarrow p(3) = 20\%$]
- A **probability density function (PDF),** f(x) can be used to generate the probability that outcomes of a continuous distribution **lie within a particular range of outcomes**.
- A **cumulative distribution function (CDF),** F(x), gives the probability that a random variable will **be less than or equal to a given value.**

Define and distinguish (☆☆)

PDF & CDF



Properties of PDF & CDF

- Properties of PDF
  - ✓ $P(r_1 \le X \le r_2)$（与PDF围成的面积即为概率）
  - ✓ $\int_{r_{min}}^{r_{max}} f(x)dx=1$， where r_{max} and r_{min} is the upper bound and lower bound of f(x)（面积的总和为1）

- Properties of CDF
  - ✓ $F(x)=\int_{-\infty}^{x} f(x)dx=P(X \le x) \rightarrow dF(x)/dx=f(x)$
  - ✓ F(x) is a non-decreasing function such that if x2 > x1 then F(x2)≥F(x1)——单调性
  - ✓ $F(-\infty)=0$ and $F(\infty)=1$——有界性

Inverse Cumulative Distribution Function

- **Inverse Cumulative Distribution Function**
  - ✓ $F(X)=x/100$, $(0<x<100)$
  - ✓ The inverse cumulative distribution function of F(X) will be:

$$F^{-1}(X) = 100y \quad (\text{assuming } y=F(X))$$

Summary

- **Compute probability**
  - ✓Mutually exclude & exhaustive event
  - ✓Two important rules： Multiplication & Addition rule
  - ✓Independent vs Mutually exclusive
  - ✓Probability Matrix
- **Distribution Function**
  - ✓Discrete vs continuous random variable
  - ✓PDF & CDF
  - ✓Inverse Cumulative Distribution Function

CONTENTS

PART 2

Learning objectives

- Describe Bayes' theorem and apply this theorem in the calculation of conditional probabilities.
- Compare the Bayesian approach to the frequentist approach.
- **Apply** Bayes' theorem to scenarios with more than two possible outcomes and calculate posterior probabilities.

Describe and distinguish (☆)

Bayesian approach vs. Frequentist approach

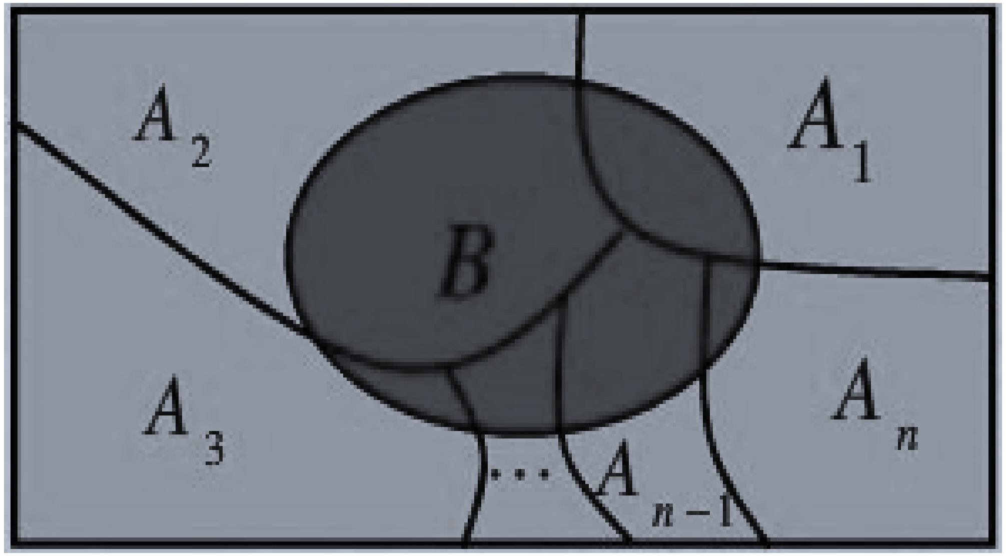
- **Frequentist approach**
  - ✓The conclusion is based only on the observed frequency of results
  - ✓Perform better for large data set
  - ✓Easier to calculate
- **Bayesian approach**
  - ✓starts with a prior belief about the probability
  - ✓Preferable for very little data
  - ✓Preferable for performance analysis and stress testing

Total Probability Theorem

- **Total Probability Theorem(全概率公式)**: Let A1,...,An be a partition of Ω. For any event B

$$P(B) = \sum_{j=1}^{n} P(A_j)P(B|A_j)$$

A_1, A_2, … … A_n are mutually exclusive and exhaustive



Describe and apply (☆☆☆)

Bayes' Theorem

- **Bayes' Formula:** Given a set of prior probabilities for an event of interest, if you receive new information, the rule for updating your probability of the event.

$$P(A|B)=\frac{P(B|A)}{P(B)}\times P(A)$$

P(A|B): updated probability （后验概率）

P(A): prior probability event （先验概率）

P(B) 用全概率公式计算

Example

- The lie detector can be used to detect if the suspect lies or not. It is known that the probability that the suspect lies is 0.7. If the suspect lies, the probability that the test result is “lied” is 0.9; if the suspect doesn’t lie, the probability that the test result is “lied” is 0.2. What is the probability that the suspect does lie given the test result is “lied”?

Example

Answer:

- Event L: the suspect lies, so $P(L) = 0.7$, $P(L^C) = 0.3$;
- Event T: the test result is "lied", so $P(T \mid L) = 0.9$, $P(T \mid L^C) = 0.2$. Using the total probability formula:

$$P(T) = P(T \mid L) \times P(L) + P(T \mid L^C) \times P(L^C)$$

$$= 0.9 \times 0.7 + 0.2 \times 0.3 = 0.69$$

- Then, the probability that the suspect does lie given the test result is "lied" is:

$$P(L \mid T) = \frac{P(T \mid L)}{P(T)} \times P(L) = \frac{0.9}{0.69} \times 0.7 = 0.913$$

Bayesian approach for three possible outcomes

Example

There are three types of managers: underperformers, in-line performers, and outperformers. The underperformers beat the market only 25% of the time, the in-line performers beat the market 50% of the time, and the outperformers beat the market 75% of the time. Initially we believe a manager has a 60% probability of being an in-line performer, a 20% chance of being an underperformer, and a 20% chance of being an outperformer.

Bayesian approach for three possible outcomes

Suppose the manager beats the market two years in a row.

What should our posterior belief that the manager is an underperformer be?

把已知条件写成事件的形式：

- ✓ Underperformers: P(U) = 20%
- ✓ In-line performers: P(I) = 60%
- ✓ Outperformers: P(O) = 20%

2B： represent beats the market two years in a row

- ✓ 事件：P(U|2B)?(题目要求的概率)

Bayesian approach for three possible outcomes

- The probability of beating the market two years in a row, for each type of manager:
  - ✓underperformers: P(2B|U)=0.25*0.25= 1/16
  - ✓in-line performers: P(2B|I)=0.5*0.5= 1/4
  - ✓outperformers: P(2B|O)=0.75*0.75= 9/16
- The unconditional probability of observing the manager beat the market two years in a row, given our prior beliefs about *p*, is:
  - ✓P(2B)= 20%*(1/16)+60%*(1/4)+20%*(9/16)=27.5%

Apply and calculate (☆☆☆)

Bayesian approach for three possible outcomes

- Posterior belief (the probability)that the manager is an underperformer is:

$$P(U|2B)=\frac{P(2B|U)}{P(2B)}\times P(U)=\frac{(\frac{1}{16})}{27.5\%}\times 20\%=4.55\%$$

Summary

- **Bayesian approach vs. Frequentist approach**
- **Total Probability Theorem**
- **Bayes' Formula:**
  - ✓Given a set of prior probabilities for an event of interest, if you receive new information, the rule for updating your probability of the event.

$$P(A|B)=\frac{P(B|A)}{P(B)}\times P(A)$$

- **Bayesian approach for multi possible outcomes**

CONTENTS

PART 2

CONTENTS

PART 2

1 . Probability Theory

2 . Statistics

3. Linear Regression

4. Time-Series Analysis

5. Simulation Method

2.1. Basic Statistics (☆)

2.2. Distributions

2.3. Hypothesis Testing and Confidence Intervals

Learning objectives

- Interpret and apply the mean, standard deviation, and variance of a random variable.
- **Calculate** the mean, standard deviation, and variance of a discrete random variable
- **Interpret and calculate** the expected value of a discrete random variable.
- **Calculate and interpret** the covariance and correlation between two random variables.
- **Calculate** the mean and variance of sums of variables.
- Describe the four central moments of a statistical variable or distribution: mean, variance, skewness and kurtosis.
- **Interpret** the skewness and kurtosis of a statistical distribution, and interpret the concepts of coskewness and cokurtosis.
- Describe and interpret the best linear unbiased estimator.

Descriptive & Inferential Statistics

- **Descriptive statistics** Describe the properties of a large data set.
- **Inferential statistics** uses a sample from a population to make probabilistic statements about the characteristics of a population.
- A **population** is a complete set of outcomes
- A **sample** is a subset of outcomes drawn from a population

Interpret and apply (☆☆)

Properties Of Return Distributions

- **Four Properties of Distributions**
  - ✓ **Central Tendency**

    Where returns are centered
  - ✓ **Dispersion**

    How far returns are dispersed from their center
  - ✓ **Skewness**

    Whether the distribution of returns is symmetrically shaped
  - ✓ **Kurtosis**

    Whether extreme outcomes are likely or whether fatty tails exist

Arithmetic Mean

- Arithmetic Mean: Equal to the sum of the observations divided by the number of the observations
  - ✓ Population Mean / Sample Mean

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N} \qquad \overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

  - ✓ Easy to work with mathematically
  - ✓ Uses all the information about the size and magnitude of the observations
  - ✓ Sensitive to extreme values
  - ✓ Arithmetic mean return focusing on average single-period performance

Median

- The value of the middle item of a set of items sorted into ascending or descending order.
  - ✓Odd number of n items, median occupies the (n+1)/2 position; even number of n items, median is equal to the mean of the items occupying the n/2 and (n+2)/2 positions.
  - ✓Advantage: not affected by extreme values (a.k.a., outliers) as arithmetic mean.
  - ✓Disadvantage: only one or two numbers considered, rest is to be ignored.

Mode

- Most frequently occurring value of the distribution.
  - ✓The distribution could have more than one mode, or even no mode (bimodal, trimodal, etc.);
  - ✓Mostly used with nominal data.
  - ✓Example: please find out the mode of following set of items: 2, 4, 5, 5, 7, 8, 8, 8, 10, 12.
  - ✓Answer: mode = 8
- The mode of a continuous random variable corresponds to the maximum of the density function.

Expected Value

- The expected value of a random variable X having possible values $x_1,x_2,x_3...,x_n$ is defined as:
  - ✓ $E(X)= x_1P(X=x_1) + x_2P(X=x_2) + ... + x_nP(X=x_n)$
  - ✓ If c is **any constant**, then $E(cx+a)= cE(X)+a$
  - ✓ $E(X+Y)= E(X)+E(Y)$
  - ✓ If X and Y are **independent** random variables, then $E(XY)= E(X)\times E(Y)$
  - ✓ $E(X^2)\neq[E(X)]^2$

Measures of Dispersion

- Variance: Equal to average of the sum of squared deviations around the mean.
  - ✓ Population Variance: $\sigma^2 = \dfrac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}$

  Where μ is the population mean and N is the size of population.
  - ✓ Sample Variance: $s^2 = \dfrac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$

  Where $\bar{X}$ is the sample mean and n is the sample size.

Interpret and calculate (☆☆☆)

Measures of Dispersion

- Positive squared root of variance(Same unit with mean)
  - ✓Population Standard Deviation:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(X_i-\mu)^2}{N}}$$

  - ✓Sample Standard Deviation:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1}}$$

Interpret and calculate (☆☆☆)

Measures of Dispersion

- **Variance**

✓Var (X)=E[(X-μ)^2]=E(X^2)-[E(X)]^2

Properties

✓If c is any constant, then Var (X+c)= Var(X)

✓If c is any constant, then Var (cX)= c^{2}Var(X)

✓If X and Y are independent random variables, then:

Var(X+Y)=Var(X)+Var(Y) or $\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2$

Var(X-Y)=Var(X)+Var(Y) or $\sigma_{X-Y}^2 = \sigma_X^2 + \sigma_Y^2$

Covariance and Correlation

- **Covariance:**

  ✓Covariance measures how one random variable moves with another random variable

$$Cov(X,Y)=E[X-E(X)][Y-E(Y)]=E(XY)-E(X)E(Y)$$

Properties:

✓Covariance ranges from negative infinity to positive infinity

✓If X ,Y are independent random variables, then: Cov (X,Y)=0

✓If X=Y, then $Cov(X,X) = E\{[X-E(X)][X-E(X)]\} = \sigma^2(X)$

✓$Cov(a+bX, c+dY)= bdCov(X,Y)$

✓$Var(X \pm Y)= Var(X)+ Var(Y) \pm 2\,Cov(X,Y)$

Covariance and Correlation

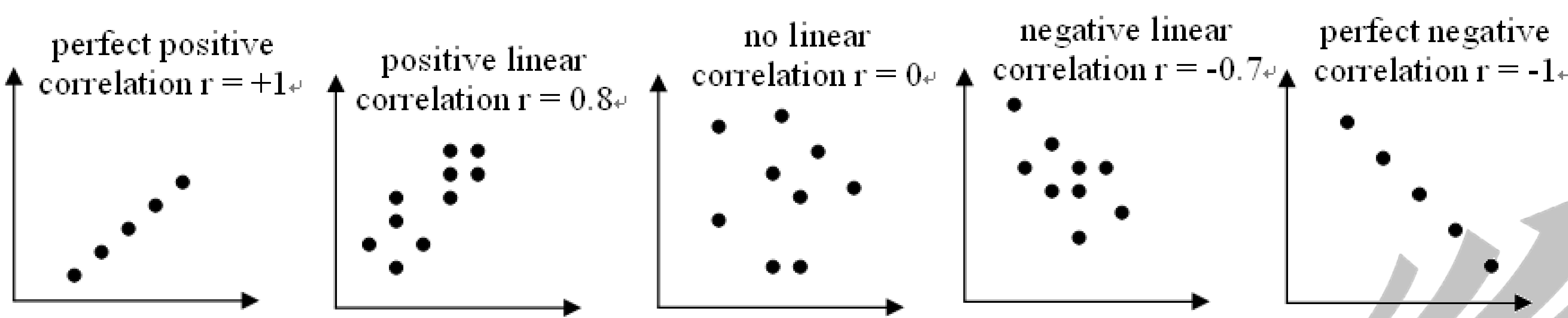
- **Correlation**
  - ✓A standardized measure of the linear relationship between two variables
  - ✓Values range from +1(perfect positive correlation) to –1, (perfect negative correlation), it has no units.
  - ✓A correlation of 0 (uncorrelated variables) indicates an absence of any linear(straight-line) relationship between the variables. The bigger the absolute value of correlation coefficient, the stronger linear relationship.

$$\rho_{X,Y}=\frac{Cov(X,Y)}{\sigma_X\sigma_Y}$$

Interpret and calculate (☆ ☆ ☆)

Interpretations of Correlation Coefficients

Correlation coefficient	Interpretation
$\rho = +1$	perfect positive linear correlation
$0 < \rho < +1$	positive linear correlation
$\rho = 0$	uncorrelated（两种情况）
$-1 < \rho < 0$	negative linear correlation
$\rho = -1$	perfect negative linear correlation



Skewness

- Symmetrical and nonsymmetrical distributions
- Positively skewed and negatively skewed

$$\text{Skewness}=\frac{E[(R-\mu)^3]}{\sigma^3}$$

Skewness= 0 $\rightarrow$ Symmetrical distribution

Skewness > 0 $\rightarrow$ Positively skewed distribution

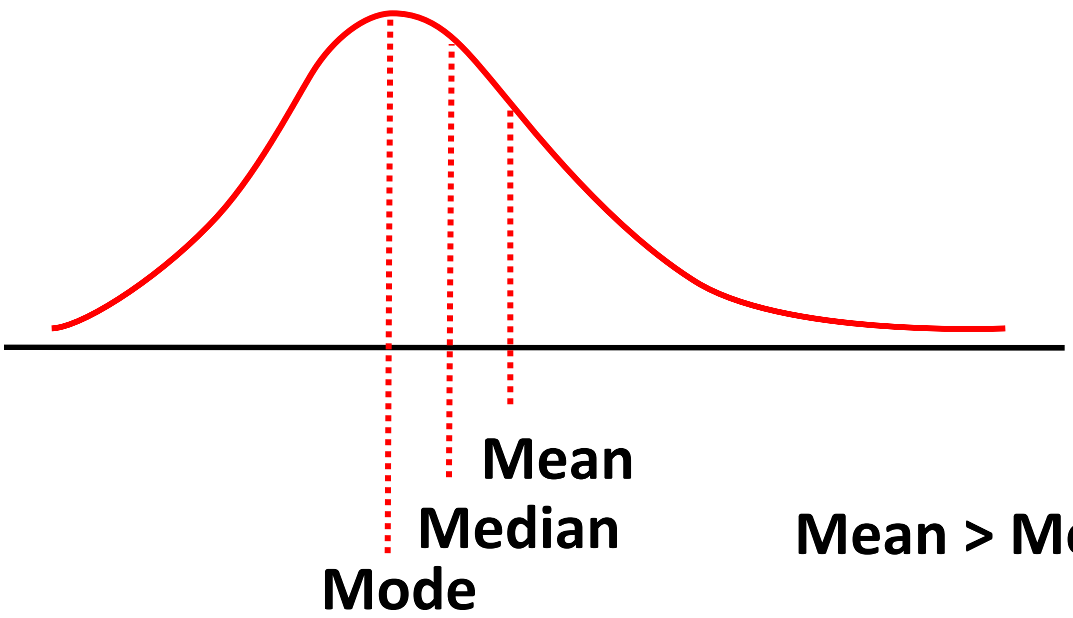
Skewness < 0 $\rightarrow$ Negatively skewed distribution

Interpret (☆ ☆)

Skewness

Positive Skew = Right Skew

- Positive skew has outliers in the right tail

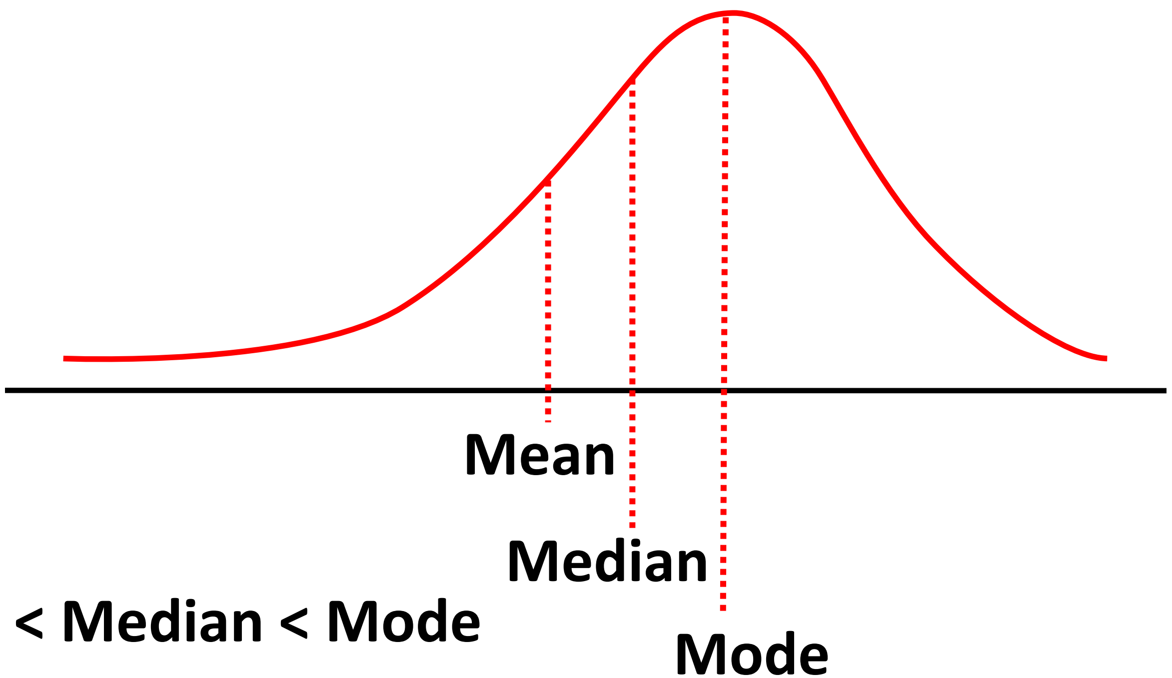


Mean > Median > Mode

Skewness

Negative Skew = Left Skew

- Negative skew has outliers in the left tail



Mean < Median < Mode

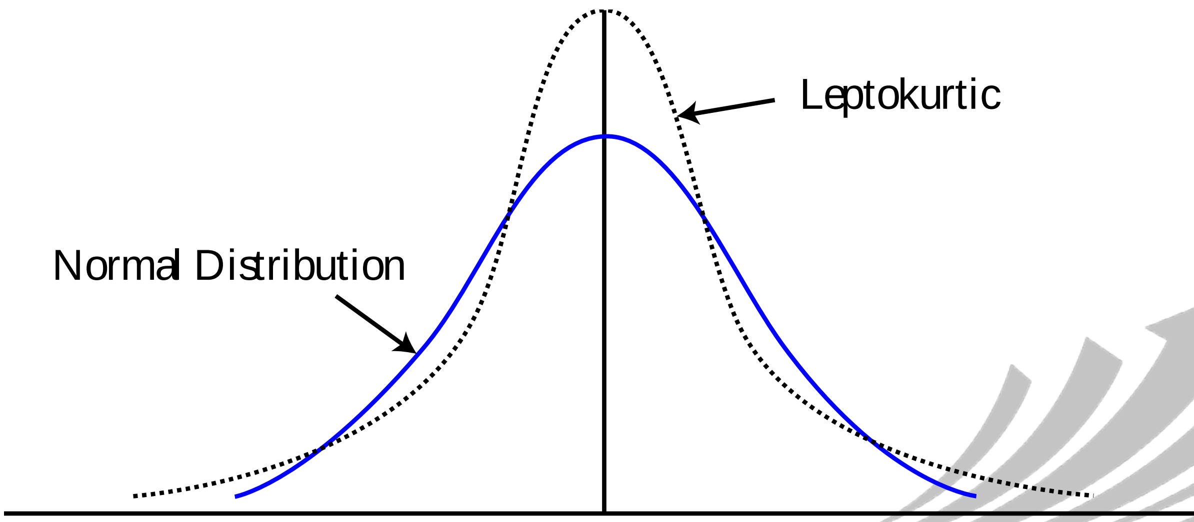
Interpret (☆☆)

Kurtosis

- A measure of the degree to how spread out a random variable is, but it puts more weight on extreme points.
- Leptokurtic vs. platykurtic

	leptokurtic	mesokurtic	platykurtic
Sample kurtosis	>3	=3	<3
Excess kurtosis	>0	=0	<0

Kurtosis$=\frac{E[(R-\mu)^4]}{\sigma^4}$



Example

Which type of distribution produces the lowest probability for a variable to exceed a specified extreme value which is greater than the mean, assuming the distributions all have the same mean and variance?

A. A leptokurtic distribution with a kurtosis of 4

B. A leptokurtic distribution with a kurtosis of 8

C. A normal distribution

D. A platykurtic distribution

高顿财经 GOLDEN FINANCE

Example

Answer: D

A platykurtic distribution has kurtosis less than 3, less than the normal p.d.f. because all other answers have higher kurtosis, this produces the lowest extreme values.

Example

A distribution of returns that has a greater percentage of small deviations from the mean and a greater percentage of extremely large deviations from the mean:

A. is positively skewed.

B. is a symmetric distribution.

C. has positive excess kurtosis.

D. has negative excess kurtosis.

Example

Answer: C

A distribution that has a greater percentage of small deviations from the mean and a greater percentage of extremely large deviations from the mean will be leptokurtic and will exhibit excess kurtosis (positive). The distribution will be taller and have fatter tails than a normal distribution.

Moments and Central Moments

- **Raw moments** are measured relative to an expected value raised to the appropriate power.

$$E\left(R^k\right) = \mu = \sum_{i=1}^{n} p_i R_i^k$$

- **Central moments** are measured relative to the mean (i.e., central around the mean). The kth central moment is defined as:

$$E(R-\mu)^k = \sum_{i=1}^{n} p_i (R_i - \mu)^k$$

  - ✓ Mean: first raw moment;
  - ✓ Variance: second central moment
  - ✓ Skewness: standardized third central moment
  - ✓ Kurtosis: standardized fourth central moment

Moments and Central Moments

- Cross moments
  - ✓Covariance(the second cross moment)
  - ✓Coskewness(the third cross moment)
  - ✓Cokurtosis(the fourth cross moment)

Coskewness and Cokurtosis

Example

- Assume four series of fund returns(A、B、C、D) where the mean, standard deviation, skew, and kurtosis all the same, but only the order of returns is different:

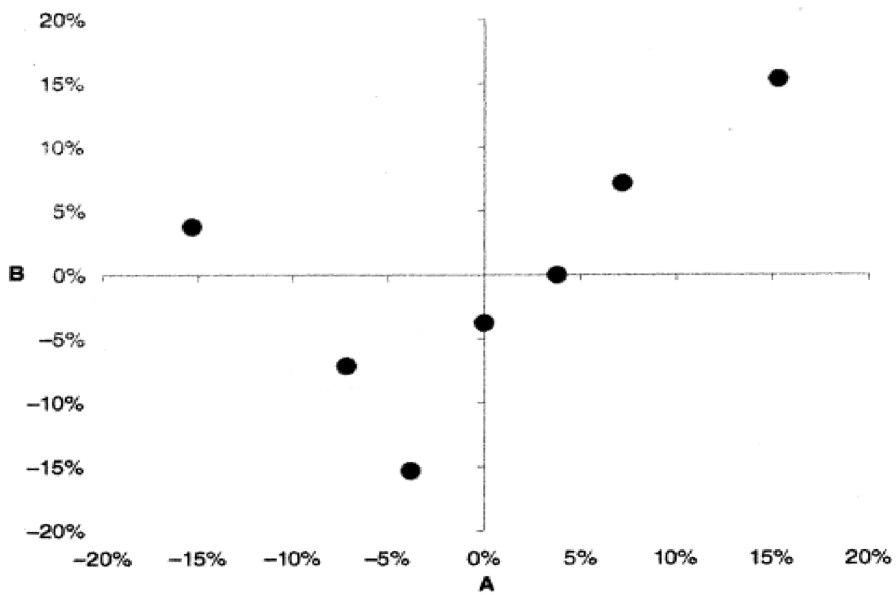
Time	A	B	C	D
1	0.00%	-3.80%	-15.30%	-15.30%
2	-3.80%	-15.30%	-7.20%	-7.20%
3	-15.30%	3.80%	0.00%	-3.80%
4	-7.20%	-7.20%	-3.80%	15.30%
5	3.80%	0.00%	3.80%	0.00%
6	7.20%	7.20%	7.20%	7.20%
7	15.30%	15.30%	15.30%	3.80%

Time	A+B	C+D
1	-1.90%	-15.30%
2	-9.50%	-7.20%
3	-5.80%	-1.90%
4	-7.20%	5.80%
5	1.90%	1.90%
6	7.20%	7.20%
7	15.30%	9.50%

- The two portfolios (A+B and C+D) have the same mean and standard deviation, but the skewness of the portfolios are **different.**

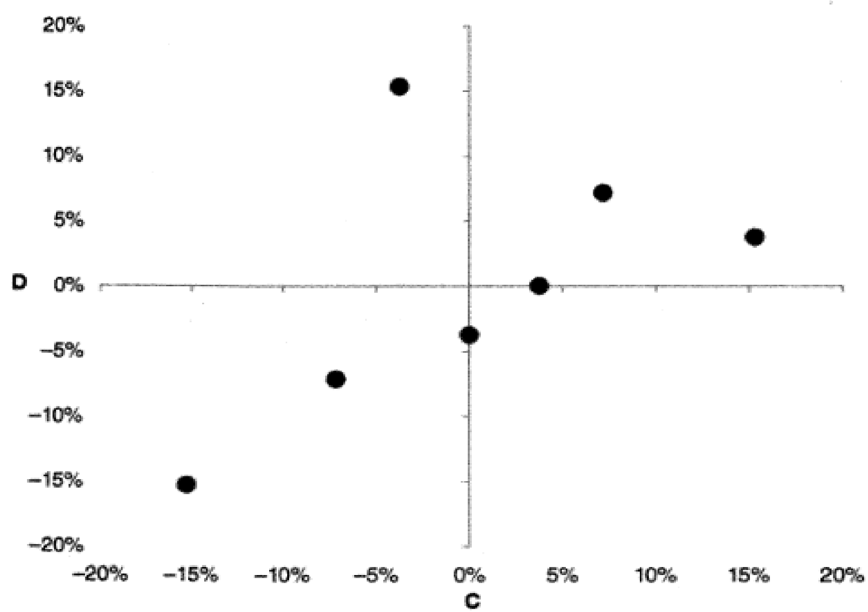
Interpret (☆)

Coskewness and Cokurtosis



- A and B: their best *positive returns occur during the same time period*, but their *worst negative returns occur in different periods*. This causes the distribution of points to be *skewed toward the top-right of the chart*.

Coskewness and Cokurtosis



- C and D: their *worst negative returns occur in the same period*, but their *best positive returns occur in different periods*. In the above chart, the points are *skewed toward the bottom-left of the chart*.

Interpret (☆)

Coskewness and Cokurtosis

- The reason is coskewness between the managers in each of the portfolios is different.
- For managers A and B, we have

$$S_{AAB} = E[(A-\mu_A)^2(B-\mu_A)]/\sigma_A^2\sigma_B$$

$$S_{ABB} = E[(A-\mu_A)(B-\mu_A)^2]/\sigma_A\sigma_B^2$$

Time	A+B	C+D
S_{XXY}	0.99	-0.58
S_{XYY}	0.58	-0.99

- Disadvantage: as the number of variables increase, the number of coskewness and cokurtosis terms will increase rapidly.

Summary

- **Descriptive statistics**
  - ✓**Central Tendency**

    mean, median, mode
  - ✓**Dispersion**

    variance, standard deviation

    Covariance and Correlation
  - ✓**Skewness**
  - ✓**Kurtosis**
  - ✓**Coskewness and Cokurtosis**

CONTENTS

PART 2

Learning objectives

- Distinguish **the key properties** among the following distributions: uniform distribution, Bernoulli distribution, Binomial distribution, Poisson distribution, normal distribution, lognormal distribution, Chi-squared distribution, Student’s t, and F-distributions, and identify common occurrences of each distribution.
- Describe the **central limit theorem** and the implications it has when combining independent and identically distributed (i.i.d.) random variables.
- Describe i.i.d. random variables and the implications of the i.i.d. assumption when combining random variables.
- Describe a mixture distribution and explain the creation and characteristics of mixture distributions.

Distinguish (☆☆)

Distribution

- **Discrete Probability Distribution**
  - ✓Bernoulli Distribution
  - ✓Binomial Distribution
  - ✓Poisson Distribution
- **Continuous Probability Distribution**
  - ✓Continuous Uniform Distribution
  - ✓Normal Distribution
    - •Standard Normal Distribution
  - ✓Lognormal Distribution

Distribution

- **Other Commonly Used Probability Distributions**
  - ✓Chi-Square Distribution
  - ✓t-Distribution
  - ✓F Distribution

（三大抽样分布）

Bernoulli Distribution

- **Bernoulli random variable & Bernoulli distribution**

  ✓ $P(X=1)=p \quad P(X=0)=1-p$

  ✓ The trial produces one of two outcomes, one representing success, denoted 1, the other representing failure, denoted 0

Binomial Distribution

- **Binomial random variable & Binomial distribution**
  - ✓The distribution of binomial random variable which is defined as the number of successes in **n** Bernoulli trials
  - ✓The assumptions for binomial distribution
    - •The probability, p, of success is constant for all trials
    - •The trails are all independent
  - ✓Expected value for binomial random variable = np
  - ✓Variance for binomial random variable = np(1-p)

Binomial Distribution

- **The probability of binomial random variable**

$$p(x) = P(X = x) = \frac{n!}{(n-x)!\,x!}p^x(1-p)^{n-x}$$

- **Example**

What is the probability of drawing exactly two white marbles from a bowl of white and black marbles in six tries if the probability of selecting white is 0.4 each time?

$$P(2)=\frac{6!}{(6-2)!2!}(0.4)^2(1-0.4)^{6-2}=0.31$$

Bernoulli Distribution & Binomial Distribution

➢ Expectations and variances

	Expectation	Variance
Bernoulli random variable (Y)	p	p(1-p)
Binomial random variable (X)	np	np(1-p)

Distinguish and identify (☆ ☆)

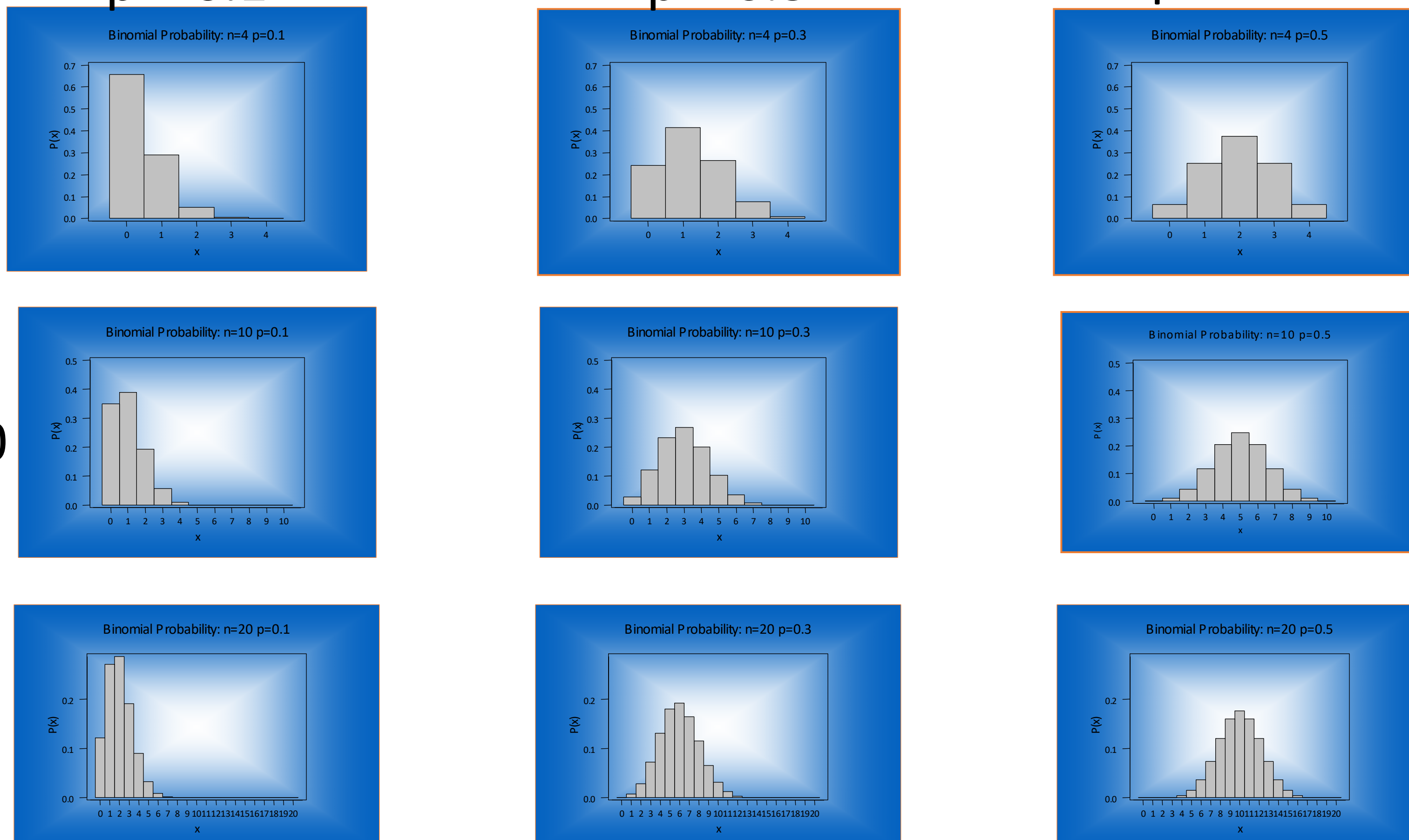
Shape of the Binomial Distribution

p = 0.1 | p = 0.3 | p = 0.5

n = 4

n = 10

n = 20



Binomial distributions become more symmetric as n increases and as **p→0.5**.

Example

On a multiple-choice exam with four choices for each of six questions, what is the probability that a student gets fewer than two questions correct simply by guessing?

A. 0.46%

B. 23.73%

C. 35.60%

D. 53.39%

Answer: D

$$C_6^0 \times 0.25^0 \times 0.75^6 + C_6^1 \times 0.25^1 \times 0.75^5 = 0.5339$$

Distinguish and identify (☆☆)

Poisson Distribution

- **Poisson random variable X**
  - refers to the number of successes occurring per unit, the parameter lamda (λ) refers to the average or expected number of successes per unit.
  - The probability function

$$f(x) = P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} \quad , \quad x = 0,1,2,\ldots\ldots$$

  - mean=variance= λ

Example

A call center receives an average of two phone calls per hour. The probability that they will receive 20 calls in an 8-hour day is closest to:

A. 5.59%

B. 16.56%

C. 3.66%

D. 6.40%

Example

Answer: A

To solve this question, we first need to realize that the expected number of phone calls in an 8-hour day is $\lambda = 2 \times 8 = 16$. Using the Poisson distribution, we solve for the probability that X will be 20.

$P(X = x) = P(X = 20) = 0.0559 = 5.59\%$

Continuous Uniform Distribution

- **Probability density function**

$$f(x)=\begin{cases}\dfrac{1}{b-a} & a\le x\le b\\ 0 & \text{otherwise}\end{cases}$$

- **Cumulative probability function**

$$F(x)=\begin{cases}0 \ \text{ for } \ x\le a\\ \dfrac{x-a}{b-a} \ \text{ for } \ a<x<b\\ 1 \ \text{ for } x\ge b\end{cases}$$

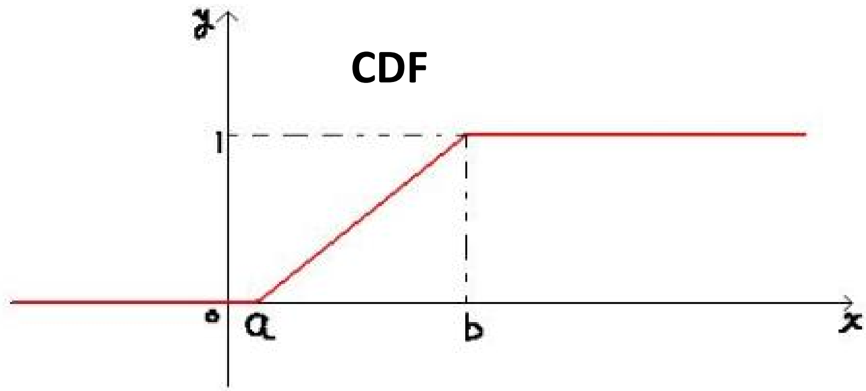
Distinguish and identify (☆ ☆)

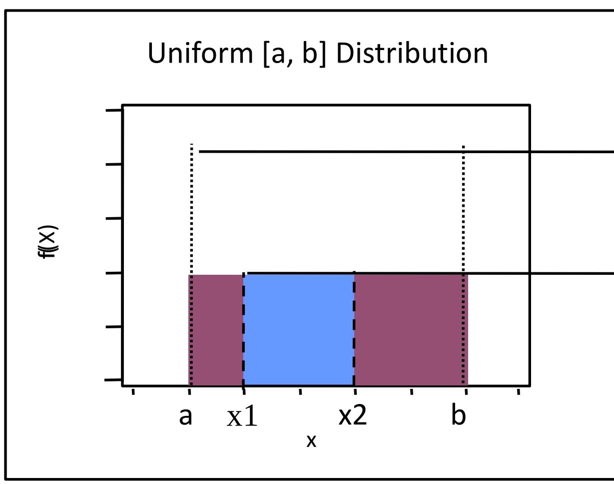
Continuous Uniform Distribution

- ✓ $E(X) = (a + b)/2$
- ✓ $Var(X) = (b - a)^2/12$
- ✓ For all $a \le x_1 < x_2 \le b$,

$$P(x_1 \le X \le x_2) = (x_2 - x_1)/(b - a)$$

CDF





PDF

The entire area under f(x) = 1/(b – a) * (b – a) = 1.00

The area under f(x) from a1 to b1 = P(x1£X£ x2)
= (x2 – x1)/(b – a)

Example

What is the probability of an outcome being between 15 and 25 for a random variable that follows a continuous uniform distribution over the range of 12 to 28 ?

A. 0.509

B. 0.625

C. 1.000

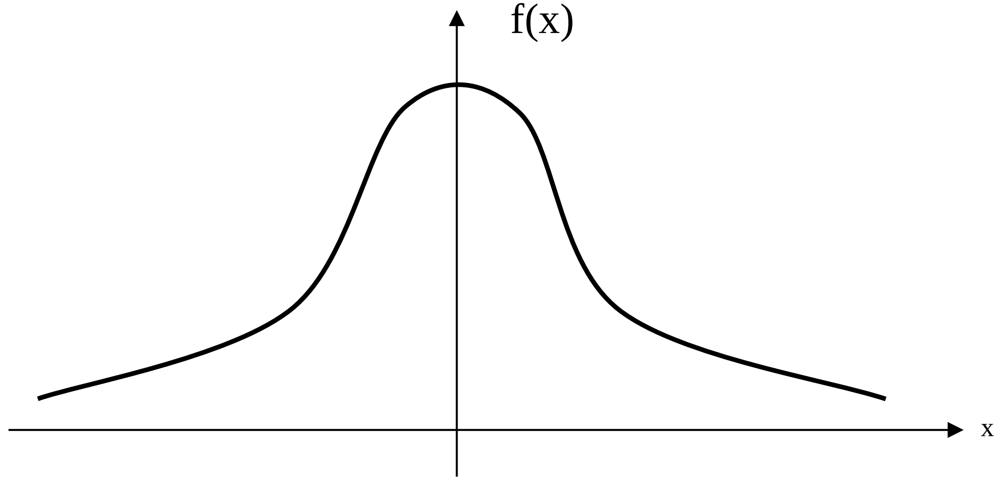
D. 1.600

Answer: B

(25-15)/(28-12) = 0.625

Normal Distribution

- **The shape of the density function**



- **Properties:**

✓X~N(μ , σ^2) $$X \sim f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad -\infty < x < \infty$$

✓Symmetrical distribution: skewness=0; kurtosis=3

Normal Distribution

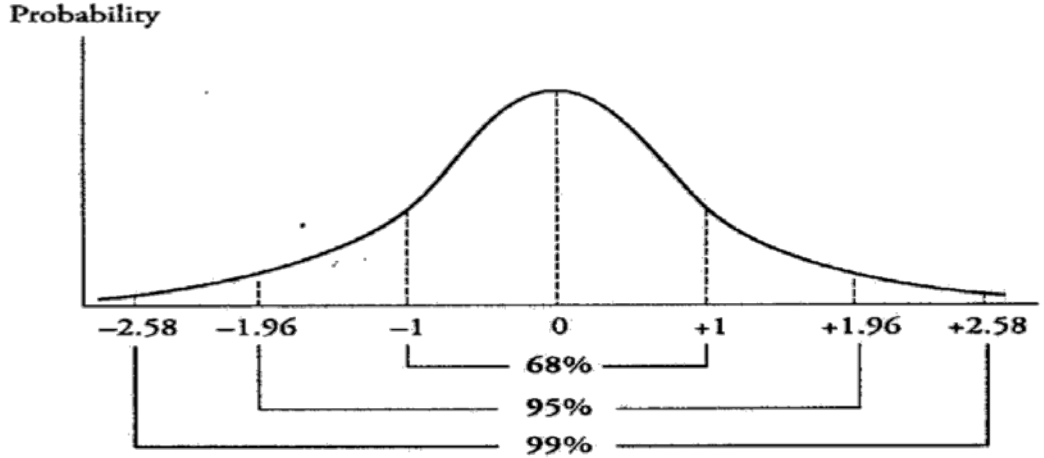
- **Properties of Normal Distribution**
  - ✓Completely Described by mean and variance
  - ✓Linear combination of normally distributed random variables is also normally distributed
  - ✓Probabilities decrease further from the mean, but the tails go on forever.

P.S.: Normal distribution can be used to model returns.

Distinguish and identify (☆☆☆)

Normal Distribution

- **The confidence intervals**
  - ✓68% confidence interval is $[\overline{x}-1s,\overline{x}+1s]$
  - ✓90% confidence interval is $[\overline{x}-1.65s,\overline{x}+1.65s]$
  - ✓95% confidence interval is $[\overline{x}-1.96s,\overline{x}+1.96s]$
  - ✓98% confidence interval is $[\overline{x}-2.33s,\overline{x}+2.33s]$
  - ✓99% confidence interval is $[\overline{x}-2.58s,\overline{x}+2.58s]$



Standard Normal Distribution

- Normal distribution defined by mean $\mu=0$, standard deviation $\sigma=1$.(z-distribution)
- Standardizing steps
  - ✓Step 1: Substracting the mean of random variable X
  - ✓Step 2: Dividing by standard deviation of random variable X
- The Z value calculated by standardization represents the number of the standard deviations from the mean.

$$z=\frac{X-\mu}{\sigma}$$

Standard Normal Distribution

➢ Example: Standard Normal Distribution

The EPS for a large group of firms are normally distributed and has μ = \$4.00 and σ = \$1.50. Find the probability that a randomly selected firm's earnings are less than \$3.70.

$$z=\frac{3.70-4.00}{1.50}=-0.20$$

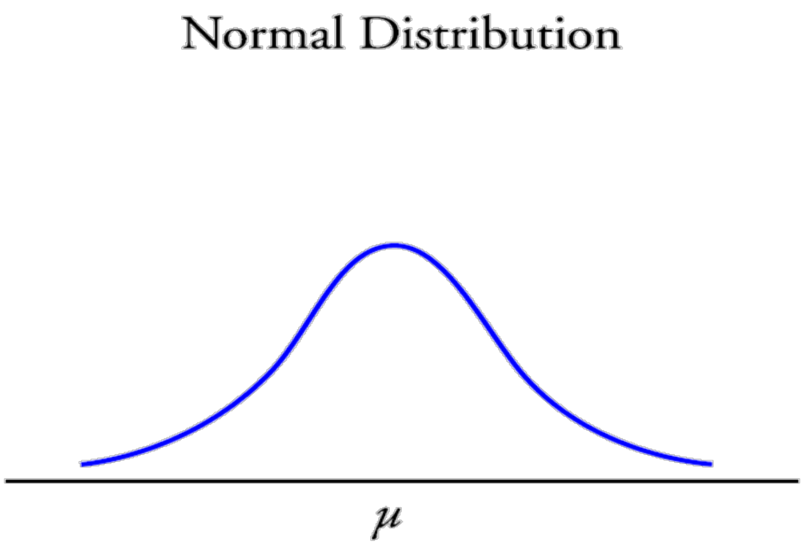
Z	0	0.01
0.1	0.5298	0.5438
0.2	0.5793	0.5832
0.3	0.6179	0.6217

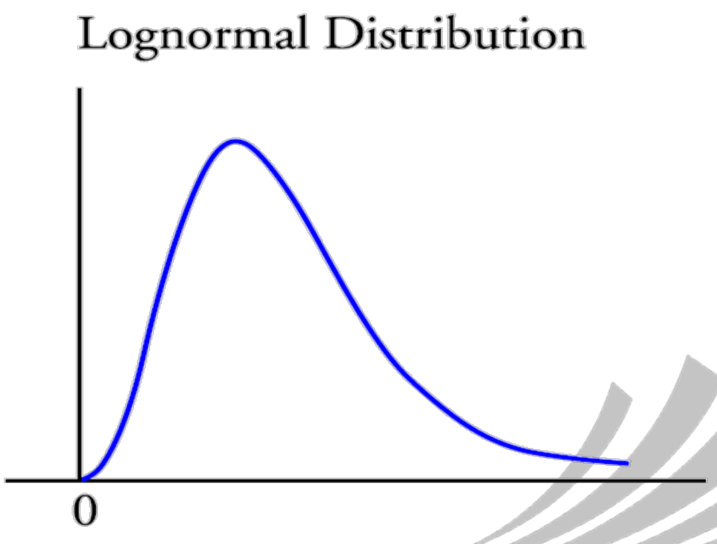
3.70 is 0.20 standard deviations below the mean of 4.00.

N(-0.2) = 1 – N(0.2) =1 - 0.5793 = 0.4207 = 42.07%

Lognormal Distribution

- The **Black-Scholes Model** assumes that the price of the underlying asset is lognormally distributed
- If lnX is normal, then X is lognormal
  - ✓Right skewed
  - ✓Bounded from below by zero
  - ✓It is used to model asset prices





Example

Which of the following statements are TRUE?

I. The sum of two random normal variables is also a random normal variable.

II. The product of two random normal variables is also a random normal variable.

III. The sum of two random lognormal variables is also a random lognormal variable.

Example

IV. The product of two random lognormal variables is also a random lognormal variable.

A. I and II only

B. II and III only

C. III and IV only

D I and IV only

- **Answer: D**

Chi-Square (χ2) Distribution

- If we have *k* independent standard normal variables, Z_1, Z_2, …, Z_k, then the sum of their squares, *S*, has a chi-squared distribution.

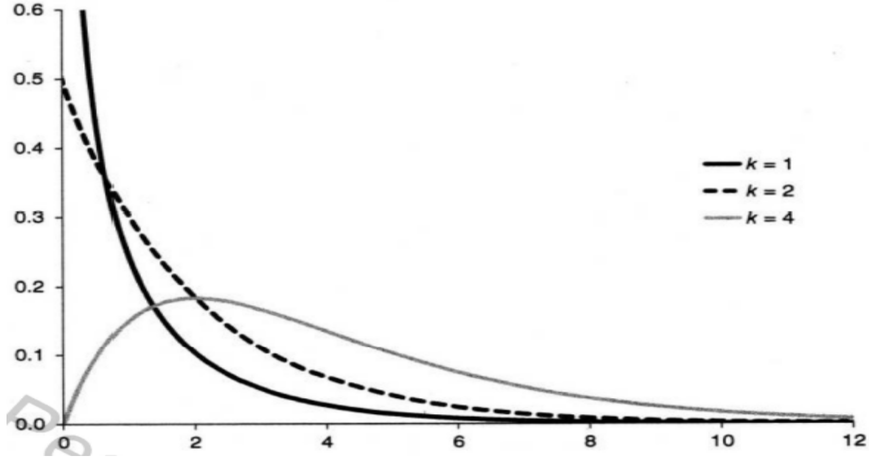
$$S = \sum_{i=1}^{k} Z_i^2 \sim \chi_k^2$$

- E(S)=k, Var(S)=2k

Describe and apply (☆)

Chi-Square (χ2) Distribution

- The chi-square distribution is asymmetrical, bounded below by zero, and approaches the normal distribution in shape as the degrees of freedom increase.
- The sum of two independent chi-squared variables, with k_1 and k_2 degrees of freedom, will follow a chi-squared distribution, with k_1+k_2 degrees of freedom.



Student's t-Distribution

- Definition: If Z is a standard normal variable and U is a chi-square variable with k degrees of freedom, which is independent of Z, then the random variable X follows a t distribution with k degrees of freedom.

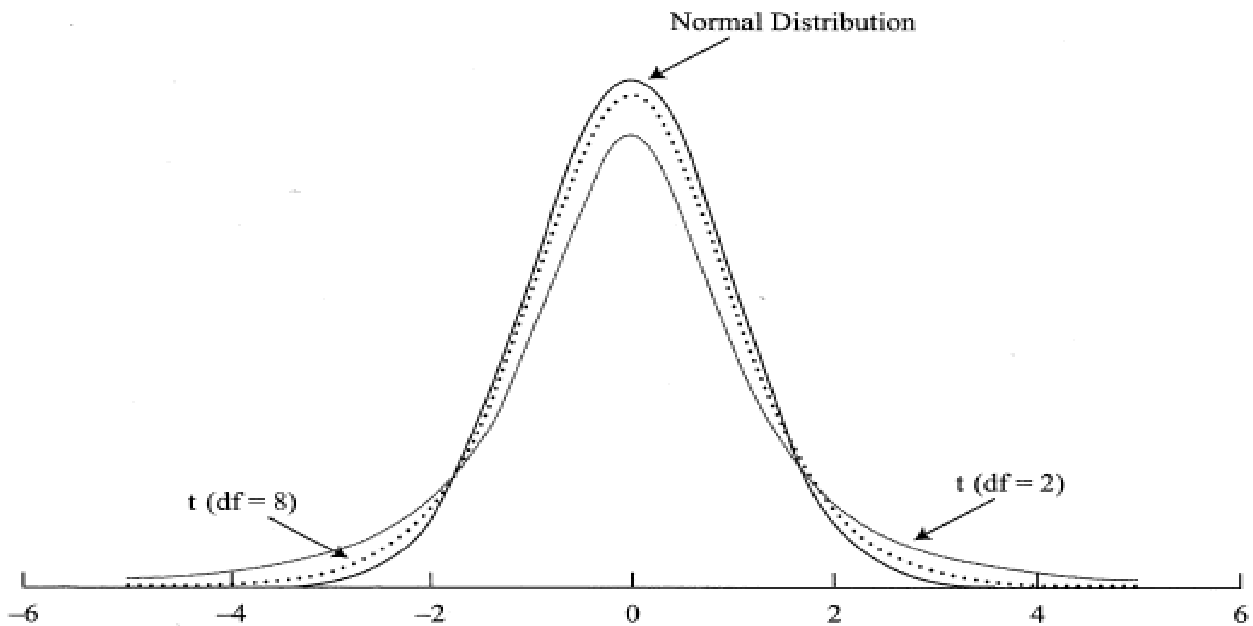
$$X = \frac{Z}{\sqrt{U/k}}$$

Student's t-Distribution

- Properties:
  - ✓ Defined by single parameter: degrees of freedom (df); and df = n – 1, where n is the sample size.
  - ✓ Symmetrical (bell shaped), skewness = 0;
  - ✓ Fatter tails than a normal distribution;
  - ✓ As df increase, t-distribution is approaching to standard normal distribution.

Student's t-Distribution

- **t-distribution vs. standard normal distribution**



✓Given a degree of confidence, t-distribution has a wider confidence interval than normal distribution.

✓As df increase, t-distribution is becoming more peaked with thinner tails, which means smaller probabilities for extreme values.

F-Distribution

- Definition: If U_1 and U_2 are two independent chi-squared distributions with k_1 and k_2 degrees of freedom, respectively, then X follows an F-distribution with parameters k_1 and k_2.

$$X = \frac{U_1/k_1}{U_2/k_2}$$

- Properties
  - ✓ The F-distribution approaches the normal distribution as the number of observations.
  - ✓ If X follows t(k), then X^2 has an F-distribution:

$$X^2 \sim F(1, k)$$

Distinguish (☆)

Parametric and Nonparametric Distribution

- **Parametric distribution**
  - ✓can be Described using a mathematical function
  - ✓easier to draw conclusions
  - ✓make restrictive assumptions, not supported by real-world
- **Nonparametric distribution**
  - ✓historical distribution
  - ✓can not be Described using a mathematical function
  - ✓fit the data perfectly
  - ✓Too specific, difficult to draw any conclusion

Mixure Distribution

- The distribution that results from a weighted average distribution of density functions is known as a mixture distribution.

$$f(x)=\sum_{i=1}^{n} w_i f_i(x), \sum_{i=1}^{n} w_i = 1$$

✓ w_i represent probability

✓ In a typical mixture distribution, the component distributions are parametric, but the weights are based on empirical data, which is non parametric.

Mixure Distribution

- there is a trade-off between using a low number and a high number of component distributions.

✓By adding more and more component distributions as inputs:

✓优点： we can approximate any data set with increasing precision;

✓缺点： The conclusions that we can draw tend to become less general in nature.

Describe and explain (☆)

Mixure Distribution

➢ Example

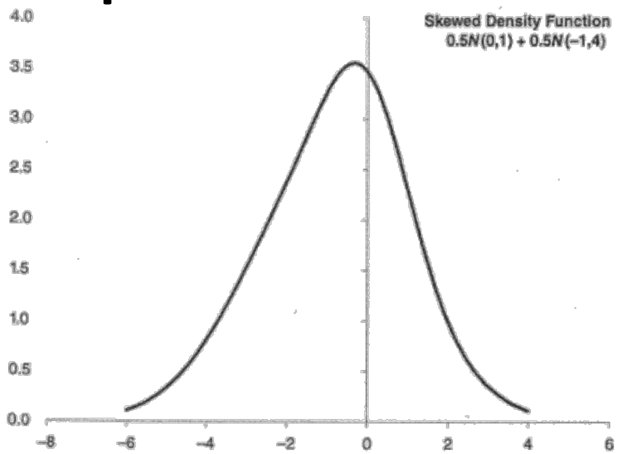


FIGURE 3-15 Skewed mixture distribution.

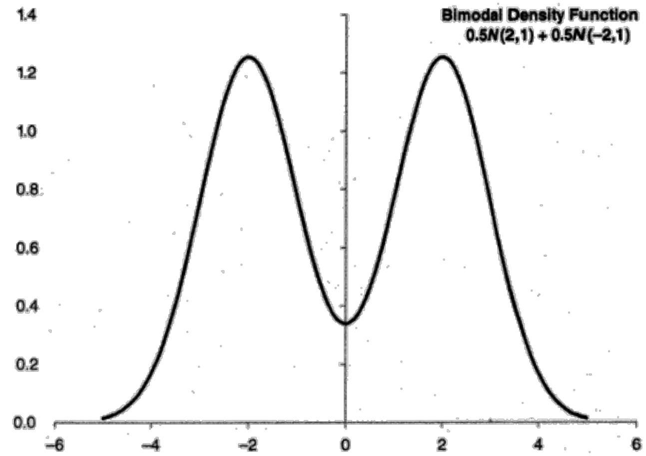


FIGURE 3-16 Bimodal mixture distribution.

- ✓ Skewness can be changed by combining distributions with different mean.
- ✓ Kurtosis can be changed by combining distributions with different variances.
- ✓ Combining distributions with significant different mean—— bimodal
- ✓ Incorporating the potential for low-frequency, high severity events.

Summary

- **Discrete Probability Distribution**
  - ✓Bernoulli Distribution & Binomial Distribution
  - ✓Poisson Distribution
- **Continuous Probability Distribution**
  - ✓Continuous Uniform Distribution
  - ✓Normal Distribution(Z-Distribution)
  - ✓Lognormal Distribution
  - ✓Chi-Square , t-Distribution, F Distribution
- **Mixture distribution**

CONTENTS

PART 2

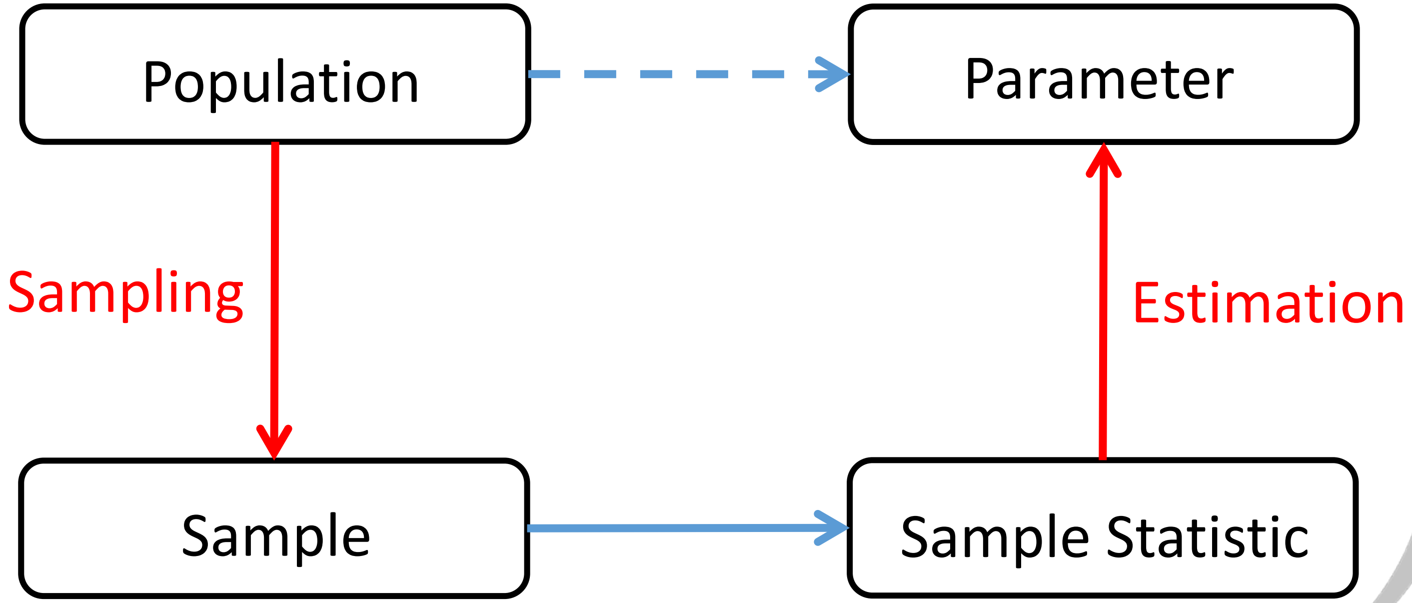
Learning objectives

- **Calculate** and interpret the sample mean and sample variance.
- **Construct and interpret** a confidence interval.
- **Construct** an appropriate null and alternative hypothesis, and calculate an appropriate test statistic.
- **Differentiate** between a one-tailed and a two-tailed test and identify when to use each test.
- **Interpret** the results of hypothesis tests with a specific level of confidence.
- **Demonstrate** the process of backtesting VaR by calculating the number of exceedances.

Statistical Inference, Sampling & Estimation

- **Statistical Inference**

✓Making forecasts, estimates or judgments about a population from the sample actually drawn from that population.



Sample Mean & Sample Variance

Sample Mean

✓It is an estimator of population mean.

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

（注意有时候$\bar{X}$也记为$\hat{\mu}$）

✓ Sample mean is actually a random variable itself.

$$E(\bar{X}) = \mu \qquad Var(\bar{X}) = \frac{\sigma^2}{n}$$

Sample Variance

✓It is an estimator of population variance

$$s^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$$

Central Limit Theorem & i.i.d

- For simple random samples of size n from a population with a mean μ and a variance σ^2, the sampling distribution of the sample mean approaches $N(\mu, \sigma^2/n)$ if the sample size is sufficiently large ($n \geq 30$).
- **Standard error of the sample mean**
  - ✓Known population variance

$$\sigma_{\overline{x}} = \sigma / \sqrt{n}$$

  - ✓Unknown population variance

$$S_{\overline{x}} = s / \sqrt{n}$$

Describe and interpret (☆☆☆)

Central Limit Theorem & i.i.d

- **Independent Identical Distribution**

✓ if we have **n** i.i.d. random variables, $X_1, X_2, \ldots, X_n$, each with mean μ and standard deviation σ, and we define $\boldsymbol{S_n}$ as the sum of those ***n*** variables, as n approaches infinity, the sum $\boldsymbol{S_n}$ converges to a normal distribution.

Best Linear Unbiased Estimator (BLUE)

- **Properties for BLUE:**
  - ✓Unbiased

    the expected value of the estimator is equal to the parameter you are trying to estimate.
  - ✓Efficient（Best）

    if the variance of its sampling distribution is smaller than all the other unbiased estimators
  - ✓Consistent

    the accuracy of the parameter estimate increases as the sample size increases
  - ✓Linearity

Point Estimate & Confidence Interval

- **Point Estimate**

  ✓The calculated value of the **sample statistic** in a given sample, used as an estimate of the **population parameter**.

- **Confidence Interval**

  ✓A range for which one can assert with a given probability 1-α, called the **degree of confidence**, that it will contain the parameter it is intended to estimate. (α is called **significance level)**

  ✓It is often referred to as the 100(1-α)% confidence interval for the parameter.

  ✓**Point estimate ± Reliability Factor x Standard Error.**

Construction of Confidence Interval of population mean

- Confidence Interval with **known** Population Variance

$$\overline{X} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

- Confidence Interval with unknown Population Variance

$$\overline{X} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

Where the number of degrees of freedom is n-1

Construct and Interpret (☆☆☆) Confidence Interval with Known and Unknown Population Variance

When sampling from a:		Reliability Factors	
Distribution	**Variance**	**Small Sample (n<30)**	**Large Sample (n>30)**
Normal	**Known**	**z-statistic**	**z-statistic**
Normal	**Unknown**	**t-statistic**	**t-statistic***
Non-normal	**Known**	**Not available**	**z-statistic**
Non-normal	**Unknown**	**Not available**	**t-statistic***

* *z*-statistic is theoretically acceptable here, but use of the *t*-statistic is more conservative

Confidence Interval with Known and Unknown Population Variance

➢ Example

The sample mean is 19.0, the sample standard error is 6.6, and n = 41. Establish a 90% confidence interval for the population mean.

t-distribution reliability factor = 1.684 (df=40, $\alpha/2=0.05$)

Confidence Interval for the population mean.

$$19 \pm 1.684 \frac{6.6}{\sqrt{41}}$$

$$17.27 < \text{mean} < 20.73$$

Construct and interpret (☆☆☆)

Construction of Confidence Interval of population mean

- **Reliability Factors for Confidence Intervals Based on the Standard Normal Distribution**
  - ✓90% confidence intervals: $Z_{0.05} = 1.65$
  - ✓95% confidence intervals: $Z_{0.025} = 1.96$
  - ✓99% confidence intervals: $Z_{0.005} = 2.58$

The Factors That Affect The Width of Confidence Interval

Change in Factors	Width of Confidence Interval (z-distribution)	Width of Confidence Interval (t-distribution)
Larger α	Smaller	Smaller
Larger n	Smaller	Smaller
Larger df	N/A	Smaller
Larger s	Larger	Larger

Describe (☆)

Subdivisions of Statistical Inference

- **Estimation**
  - ✓Addresses the questions such as “what is this parameter’s value”.
- **Hypothesis Testing**
  - ✓Addresses the questions such as “is the value of the parameter equal to a specific value”.

Steps for Hypothesis Testing

- **Hypothesis**
  - ✓A statement about one or more populations
- **Hypothesis Testing Steps**
  - ✓Stating the hypothesis—relation to be tested
  - ✓Selecting a test statistic
  - ✓Specifying the level of significance
  - ✓Stating the decision rule for the hypothesis
  - ✓Collecting the data and calculate the test statistic
  - ✓Making a decision about the hypothesis
  - ✓Making the economic or investment decision.

Differentiate and identify (☆☆☆)

Hypothesis Testing

- **Hypothesis**: a statement about the value of a population parameter to be tested
- The **null** hypothesis (H_0) and **alternative** hypothesis (H_a)
- **One-tailed** test vs. **Two-tailed** test
  - ✓One-tailed test

    $H_0: \mu \geq \mu_0$ $H_a: \mu < \mu_0$

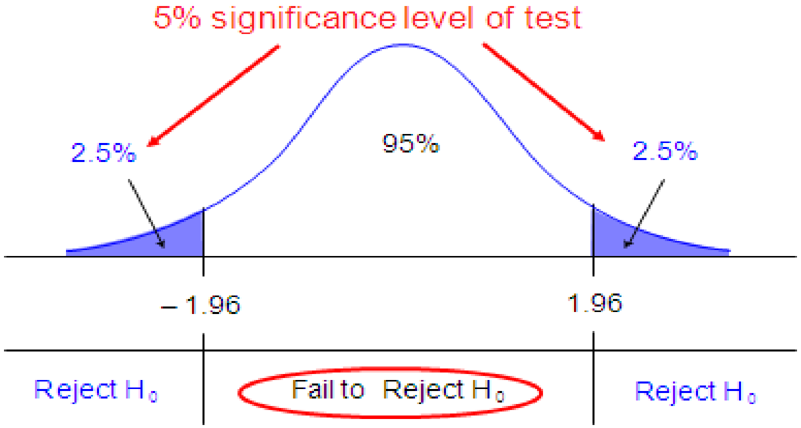
    $H_0: \mu \leq \mu_0$ $H_a: \mu > \mu_0$
  - ✓Two-tailed test

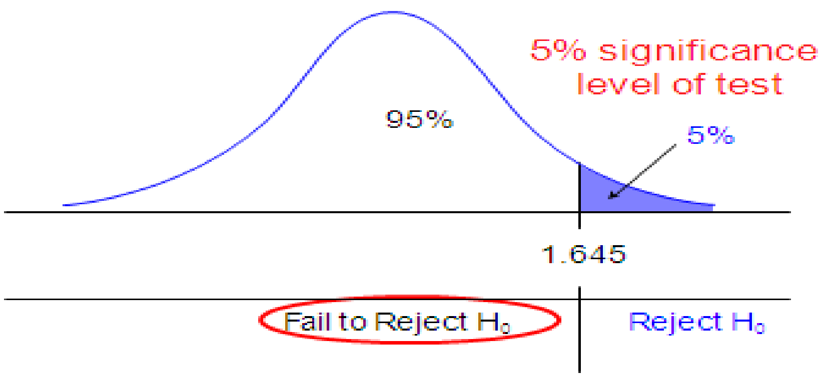
    $H_0: \mu = \mu_0$ $H_a: \mu \neq \mu_0$

Two-tailed Test and One-tailed Test

Example:

H_0: $\mu = \mu_0$ versus H_a: $\mu \neq \mu_0$ *(Two-tailed test)*





H_0: $\mu \leq \mu_0$ versus H_a: $\mu > \mu_0$ *(One-tailed test)*

Differentiate and identify (☆☆☆)

Type I Error, Type II Error, Level of Significance, Power of Test

- **Type I Error**
  - ✓Rejecting null hypothesis when it is true.
  - ✓The probability of making a Type I error is equal to α, also known as the level of significance of the test.
  - ✓P(Type I Error)= α.
- **Type II Error**
  - ✓Failing to reject the null hypothesis when it is false.
  - ✓The probability of making a Type II error is equal to β, P(Type II Error)= β.
- **Power of Test**
  - ✓Rejecting the null hypothesis when it is false.
  - ✓The probability of power of test is equal to (1-β),
  - ✓P(Power of Test)=1 - P(Type II Error).

Differentiate and identify (☆☆☆)

高顿财经 GOLDEN FINANCE

Type I Error, Type II Error, Level of Significance, Power of Test

Decision	True Situation	
	H_0 True	**H_0 False**
Do not reject H_0	Correct Decision	Type II Error (probability = β)
Reject H_0 (accept H_a)	Type I Error (probability = α)	Power of Test (probability = 1-β)

高顿财经 GOLDEN FINANCE

Test Statistic and Decision Rule

- **Test Statistic**
  - ✓A quantity calculated based on a sample.

$$\text{Test Statistic} = \frac{\text{Sample Statistic-Value of the population parameter under } H_0}{\text{Standard error of the sample statistic}}$$

- **Decision Rule**
  - ✓If: Test Statistic ≥ upper critical value or

    Test Statistic ≤ lower critical value
  - ✓Conclusion: Rejecting the null hypothesis.

P-value and Decision Rule

- **p-value**

  ✓The smallest significance level at which the null hypothesis can be rejected.

- **Decision Rule**

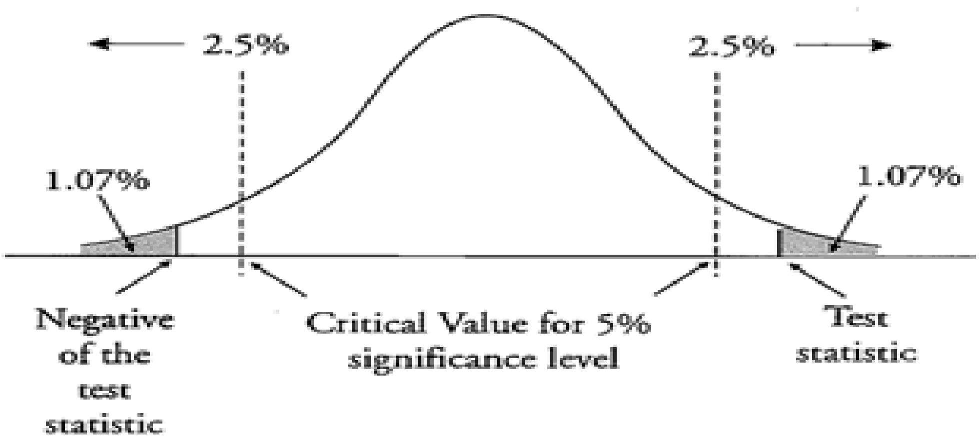
  ✓Rejecting the null hypothesis if the p-value is less or equal to the level of significance, α,or (p-value ≤ α).

Differentiate and identify (☆☆☆)

P-value and Decision Rule

Conclusion: Rejecting the null hypothesis.

Two-tailed Hypothesis Test with p-value = 2.14%



Test of Single Population Mean

- **H_0: $\mu=\mu_0$**
- **z-test vs. t-test**

	Normal Distribution n<30	**n≥30**
Known Variance	Z-statistic	Z-statistic
Unknown Variance	t-statistic	t-statistic or Z-statistic

- **z-statistic** $Z=\dfrac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ $\quad Z=\dfrac{\bar{x}-\mu_0}{s/\sqrt{n}}$
- **t-statistic** $t_{n-1}=\dfrac{\bar{x}-\mu_0}{s/\sqrt{n}}$

Example

Bob tests the null hypothesis that the population mean is less than or equal to 45. From a population size of 3,000,000 people, 81 observations are randomly sampled. The corresponding sample mean is 46.3 and sample standard deviation is 4.5. What is the value of the appropriate test statistic for the test of the population mean, and what is the correct decision at the 5 percent significance level?

Example

A. z = 0.29. and fail to reject the null hypothesis

B. z = 2.60. and reject the null hypothesis

C. t = 0.29. and accept the null hypothesis

D. t = 2.60. and neither reject nor fail to reject the null hypothesis

Answer: B

The population variance is unknown but the sample size is large (>30) The test statistics is: $z = (46.3- 45)/(4.5/(81)^{1/2}=2.60$, which exceeds the critical z-value of 1.65,then we reject the H_0.

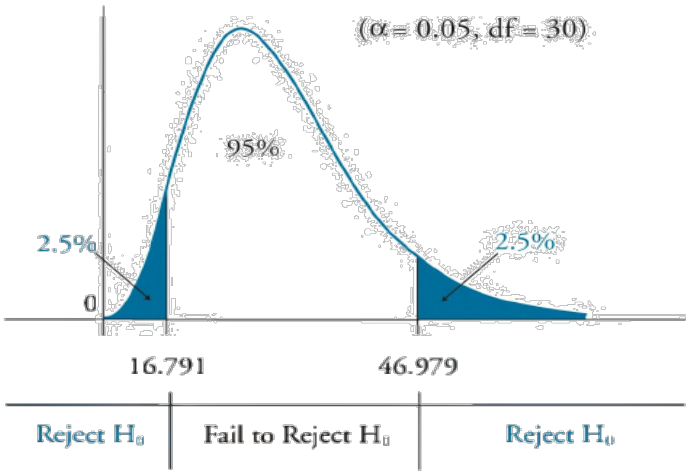
Test of Population Variance

Test of Single population variance

✓The chi-square test (χ^2-test)

- H_0: $\sigma^2=\sigma_0^2$

$$x^2=\frac{(n-1)s^2}{\sigma_0^2}$$ df=n-1



Degrees of Freedom	Probability in Right Tail					
	0.975	0.95	0.90	0.1	0.05	0.025
9	2.700	3.325	4.168	14.684	16.919	19.023
10	3.247	3.940	4.865	15.987	8.307	20.483
11	3.816	4.575	5.578	17.275	19.675	21.920
30	**16.791**	18.493	20.599	40.256	43.773	**46.979**

Test of Population Variance

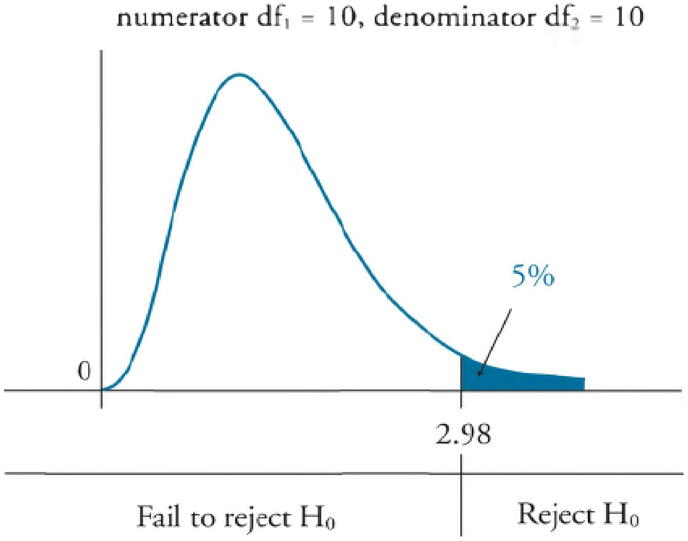
- **Test of Two population variance**

✓ **The F test**

- H_0: $\sigma_1^2=\sigma_2^2$

$$F = \frac{s_1^2}{s_2^2}$$

$df_1=n_1-1$; $df2=n2-1$



- Always put the larger variance in the numerator ($s_1^2>s_2^2$)
- The rejection region is always the right-side tail, no matter the test is one-tailed or two-tailed

Differentiate and identify (☆☆☆)

Summary of Hypothesis Testing

Test Type	Assumptions	H_0	Test Statistic	Distribution
Mean Hypothesis Testing	Normal distributed population, Known population variance	$\mu = \mu_0$	$Z = \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$	$N(0,1)$
	Normal distributed population, Unknown population variance	$\mu = \mu_0$	$t = \dfrac{\bar{x} - \mu_0}{s / \sqrt{n}}$	t_{n-1}
Variance Hypothesis Testing	Normal distributed population	$\sigma^2 = \sigma_0^2$	$x^2 = \dfrac{(n-1)s^2}{\sigma_0^2}$	$\chi^2\ (n-1)$
	Two independent normal distributed populations	$\sigma_1^2 = \sigma_2^2$	$F = \dfrac{s_1^2}{s_2^2}$	$F(n_1-1, n_2-1)$

Chebyshev's Inequality

- **Chebyshev's Inequality**
  - ✓For any distribution with finite variance, the minimum percentage of observations that lie within *k* standard deviations of the mean would be $1-1/k^2$, given $k>1$.
- **Example:**
  - ✓Minimum percentage of observations lie within 2 standard deviations of the mean would be $1-1/2^2=75\%$

Chebyshev's Inequality

- According to **Chebyshev's inequality**, the following relationships hold for any distribution. At least
  - ✓ 36% of observations lie within ±1.25 σ
  - ✓ 56% of observations lie within +1.50 σ
  - ✓ 75% of observations lie within ±2 σ
  - ✓ 89% of observations lie within ±3 σ
  - ✓ 94% of observations lie within ±4 σ

Backtesting VaR

Backtesting VaR

✓checking the predicted outcome of a model against actual data. Any model parameter can be backtested.

✓when assessing a VaR model, each period can be viewed as a Bernoulli trial. In the case of one-day 95% VaR, there is a 5% chance of an exceedance event each day, and a 95% chance that there is no exceedance.

✓Because exceedance events are independent, over the course of n days the distribution of exceedances follows a binomial distribution

Example

An analyst wants to test whether the standard deviation of return from pharmaceutical stocks is lower than 0.3. For this purpose, he obtains the following data from a sample of 30 pharmaceutical stocks. Mean return from pharmaceutical stocks = 8%. Standard deviation of return from pharmaceutical stocks = 12%. Mean return from the market = 12%. Standard deviation of return from the market = 16%. What is the appropriate test statistic for this test?

Example

A. t-statistic

B. z-statistic

C. F-statistic

D. χ2 statistic.

Answer: D

Tests of the variance (or standard deviation) of a population requires the chi-squared test.

Summary

- **Estimation**
  - ✓Point Estimate
  - ✓Confidence Interval
- **Properties for BLUE**
- **Hypothesis Testing**
  - ✓Steps(Hypothesis, Test Statistic, Decision Rule: critical value or p-value)
  - ✓Type(Mean , Variance)
  - ✓Type I Error, Type II Error
- **Chebyshev's Inequality**

CONTENTS

PART 2

CONTENTS

PART 2

Learning objectives

- **Explain** how regression analysis in econometrics measures the relationship between dependent and independent variables.
- **Interpret** a population regression function, regression coefficients, parameters, slope, intercept, and the error term.
- **Interpret** a sample regression function, regression coefficients, parameters, slope, intercept, and the error term.
- Describe the key properties of a linear regression.
- Define an ordinary least squares (OLS) regression and **calculate** the intercept and slope of the regression.

Learning objectives

- Describe the method and three key assumptions of OLS for estimation of parameters.
- Summarize the benefits of using OLS estimators.
- Describe the properties of OLS estimators and their sampling distributions, and explain the properties of consistent estimators in general.
- **Interpret** the explained sum of squares, the total sum of squares, the residual sum of squares, the standard error of the regression, and the regression R2.
- **Interpret** the results of an OLS regression.

Learning objectives

- **Calculate** and interpret confidence intervals for regression coefficients.
- **Interpret** the p-value.
- **Interpret** hypothesis tests about regression coefficients.
- **Evaluate** the implications of homoskedasticity and heteroskedasticity.
- Determine the conditions under which the OLS is the best linear conditionally unbiased estimator.
- **Explain** the Gauss-Markov Theorem and its limitations, and alternatives to the OLS.
- **Apply and interpret** the t-statistic when the sample size is small.

Interpret (☆☆)

The Linear Regression Model

- Study Question： Reduces the average class size, what will the effect be on standardized test scores in her district?

$$\beta_{classsize} = \frac{\text{change in testscores}}{\text{change in classsize}} = \frac{\Delta \text{TestScore}}{\Delta \text{ClassSize}}$$

$$\downarrow$$

$$\text{TestScore} = \beta_0 + \beta_{\text{ClassSize}} \times \text{ClassSize} + \text{other factors}$$

- $\beta_0 + \beta_{\text{ClassSize}} \times \text{ClassSize}$ represents the **average effect** of class size on scores in the population of school districts

Population Regression Function

- Dependent variable (Y) and independent variable (X)
- The simple linear regression model

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

✓Y_i = i_{th} observation of the dependent variable, Y

✓X_i = i_{th} observation of the independent variable, X（regressor）

✓β_0= regression intercept term

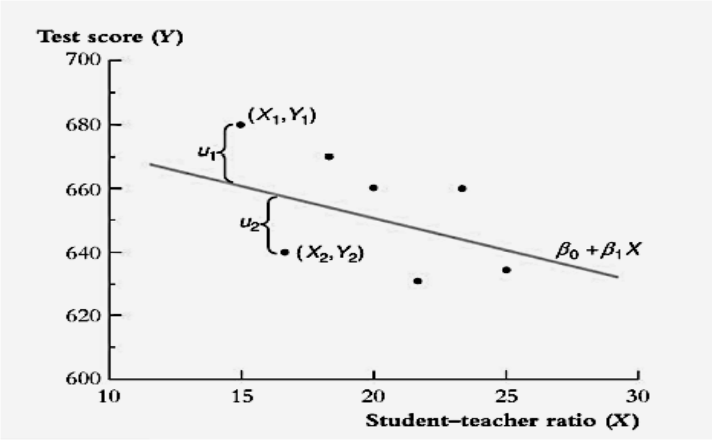
✓β_1= regression slope coefficient

✓u_I = error term

Interpret (☆☆)

Population Regression Function

- If you know the value of X, according to this population regression line you would predict that the value of the dependent variable Y is $Y_i = \beta_0 + \beta_1 X_i$
- In a practical situation, β_0 and β_1 is unknown, need to use sample data to estimate.
- The population and sample coefficients are almost always different.



Ordinary Least Squares (OLS)

➢ **Ordinary least squares (OLS)** estimation is a process that estimates the population parameters B; with corresponding values for b; that minimize the squared residuals (i.e., error terms).

$$\min_{\beta_0,\beta_1} \sum u_i^2 = \sum_{\beta_0,\beta_1} [Y_i - \beta_0 - \beta_1 X_i]^2$$

✓The estimated **slope coefficient (** $\hat{b}_1$ **)**

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{Cov(X,Y)}{Var(X)}$$

✓The **intercept term ($\hat{\boldsymbol{\beta}}_0$)** $\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}$

✓The **residual** $\hat{u}_i = Y_i - \hat{Y}_i$

Interpret and explain (☆ ☆ ☆)

高顿财经 GOLDEN FINANCE

Measures of Fit

- **Coefficient of determination (R^2)**
  - ✓ the R^2 is the ratio of the sample variance of $\hat{Y}_i$ to the sample variance of Y_i. $Y_i = \hat{Y}_i + \hat{u}_i$
  - ✓**Total sum of squares(TSS)**： Measures the total variation in the dependent variable.

$$TSS = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

  - ✓**Explained sum of squares (ESS)**： Measures the variation in the dependent variable that is explained by the independent variable.

$$ESS = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2$$

Measures of Fit

- **Residual Sum of Squared (SSR) :** It's also known as the sum of squared residuals or the residual sum of squares.

$$SSR=\sum_{i-1}^{n}(Y_i-\hat{Y})^2$$

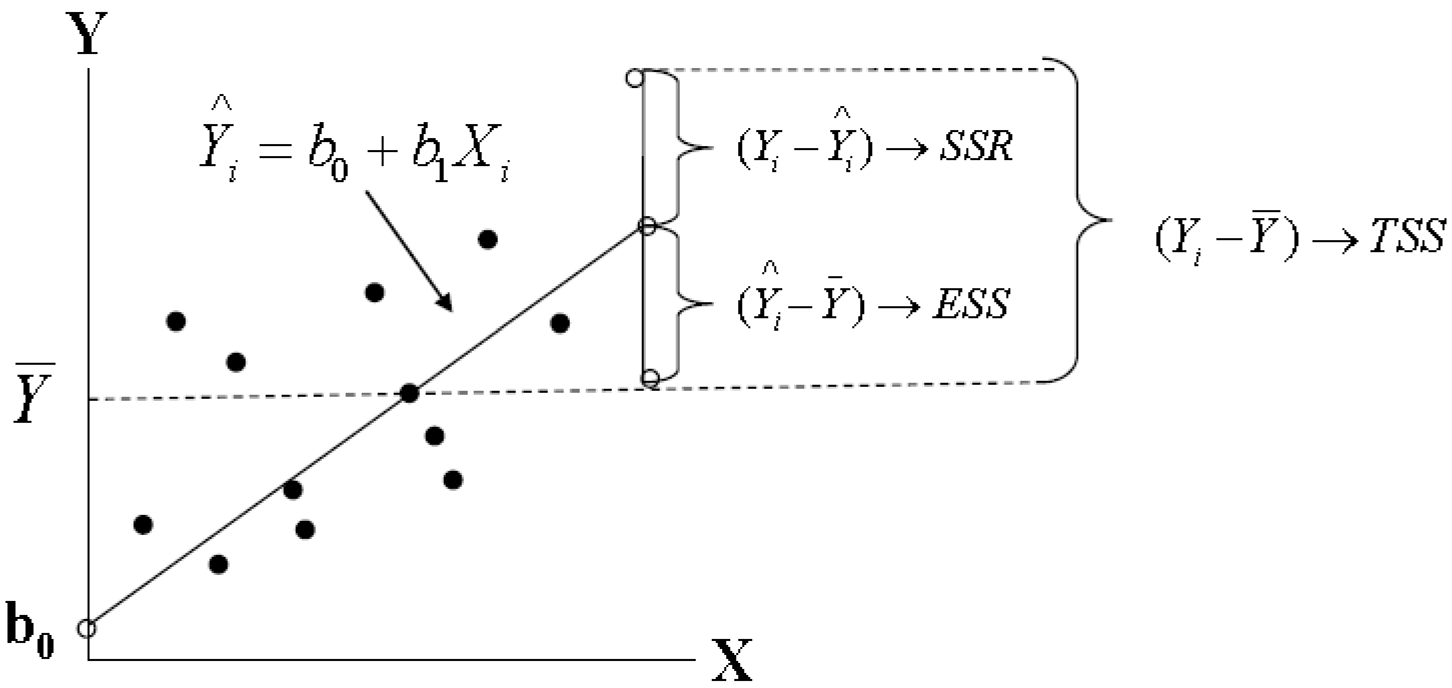
✓**Total variation** = Explained variation + Unexplained variation, or TSS = ESS + SSR

$$R^2=\frac{ESS}{TSS}=\frac{\sum(\hat{Y}_i-\bar{Y})^2}{\sum(Y_i-\bar{Y})^2}=1-\frac{SSR}{TSS}=1-\frac{\sum(Y_i-\hat{Y}_i)^2}{\sum(Y_i-\bar{Y})^2}$$

✓For simple linear regression (one independent variable), R^2 is computed by simply squaring the correlation coefficient, that is $R^2=\rho_{Y,X}^2$

Measures of Fit

Components of Variations



Standard Error of Regression

- **Standard error of regression**
  - ✓An estimator of the standard deviation of the regression error ui
  - ✓A measure of the spread of the observations around the regression line

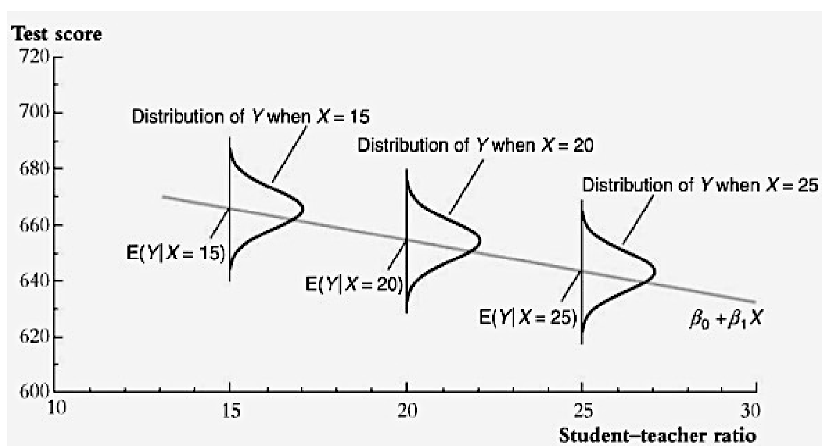
$$SER=\sqrt{\frac{SSR}{n-2}}=\sqrt{\frac{1}{n-2}\sum_{i=1}^{n}\hat{u}_i^2}$$

  - ✓because u_i are unobserved, the SER is computed using u_i hat
  - ✓ Gauges the "fit" of the regression line. The smaller the standard error, the better the fit.

Describe (☆ ☆)

Assumptions Underlying Linear Regression

- **Three key assumptions**
  - ✓ The expected value of the error term, conditional on the independent variable is zero. ($E(\varepsilon_i|X_i) = 0$)
    - Is equivalent to assuming that the population regression line is the conditional mean of Y_i given X_i
    - $E(u_i|X_i) = 0$ implies that $\text{corr}(u_i, X_i) = 0$,but does not go the other way.



Describe (☆☆)

Assumptions Underlying Linear Regression

- **Three key assumptions**
  - ✓ All (X, Y) observations are independent and identically distributed (i.i.d.)
  - ✓ It is unlikely that large outliers will be observed in the data.
    - Large outliers have the potential to create misleading regression results.

Summarize (☆☆)

Sampling Distribution of OLS estimator

- The larger variance of X, the smaller variance of $\hat{\beta}_1$

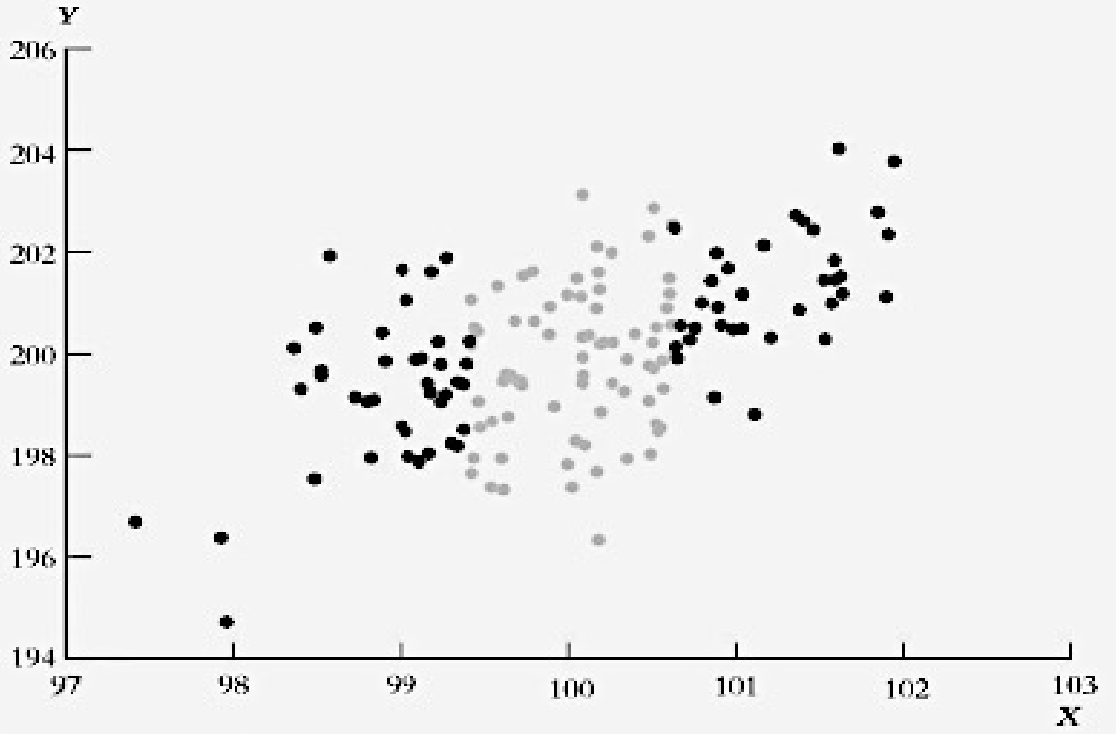


FIGURE 7-6 The variance of $\hat{\beta}_1$ and the variance of X.

Hypothesis Testing about the Regression Coefficient

- **Significance test for a regression coefficient**
  - ✓ H_0: b_1= the hypothesized value

$$t = \frac{\hat{b}_1 - b_1}{s_{\hat{b}_1}} \qquad df=n-2$$

  - ✓ Decision rule: Reject H_0 if $t > +t_{critical}$, or $t < -t_{critical}$
  - ✓ Rejection of null hypothesis means the regression coefficient is different from the hypothesized value given a level of significance α.

Regression coefficient confidence interval

- **Decision rule:** If the confidence interval at the desired level of significance does not include zero, the null is rejected, and the coefficient is said to be statistically different from zero.

$$\hat{b}_1 \pm t_c s_{\hat{b}_1}$$

✓Where t_c is the critical two-tailed t-value for the selected confidence level with the appropriate number of degrees of freedom, which is equal to the number of sample observations minus 2 (i.e., n - 2). $s_{\hat{b}_1}$ is the standard error of regression coefficient.

Example

Paul Graham, FRM® is analyzing the sales growth of a baby product launched three years ago by a regional company. He assesses that three factors contribute heavily towards the growth and comes up with the following results:

$$Y = b + 1.5\,X_1 + 1.2X_2 + 3X_3$$

Explained Sum of Squares [ESS] = 869.76

Sum of Squared Residuals [SSR] = 22.12

Example

Determine what proportion of sales growth is explained by the regression results.

A. 0.36

B. 0.98

C. 0.64

D. 0.55

Answer: B

Binary Variable

- **Binary Variable(Dummy Variable/Indicator variable)**

  There are occasions when the independent variable is binary in nature-it is either "on" or "off". Quantify the impact of qualitative events.

$$Y_i = \beta_0 + \beta_1 \mathrm{D}_i + u_i \quad D_i=0 \text{ or } 1$$

- indicates the difference in the dependent variable for the category represented by the dummy variable and the average value of the dependent variable for all classes except the dummy variable class.

β_1 indicates $E(Y|\mathrm{D}_i=1) - E(Y|\mathrm{D}_i=0)$

Evaluate (☆ ☆)

高顿财经 GOLDEN FINANCE

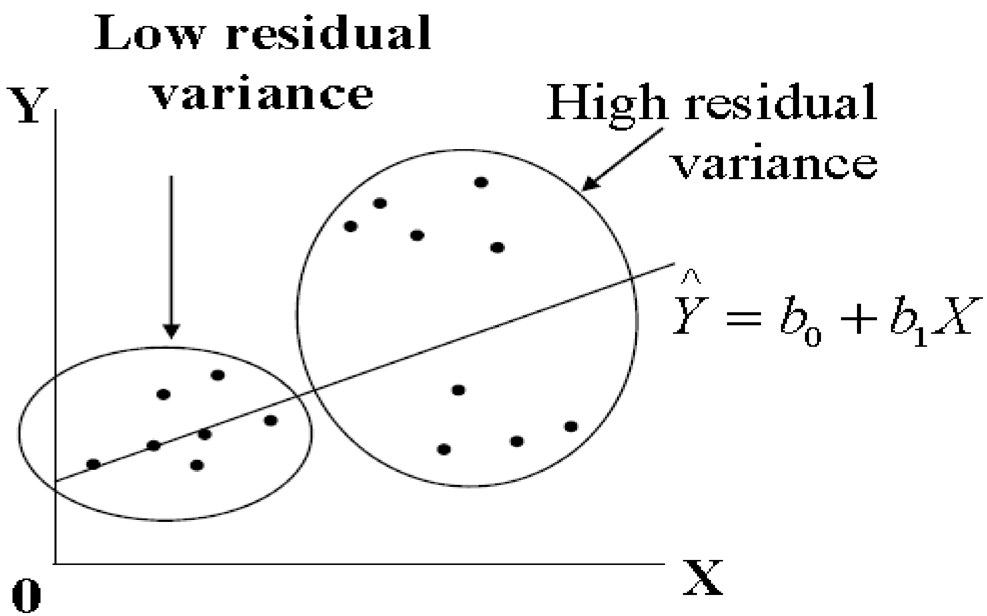
Heteroskedasticity

- **Homoskedastic**
  - The variance of the conditional distribution of u_i given X_i is constant for i=1,…,n
  - Is equivalent to say variance of Y_i is the same for X_i
- **Heteroskedasticity**
  - The variance of the conditional distribution of u_i given X_i is not constant

Heteroskedasticity

➢ Conditional heteroskedasticity

- ✓ The variance of the error terms increase(decrease) with increase(decrease) in the value of the independent variables.
- ✓ Conditional heteroskedasticity creates significant problems for statistical inference.



Evaluate (☆ ☆ ☆)

Heteroskedasticity

➤ **Effect of Heteroskedasticity on Regression Analysis**

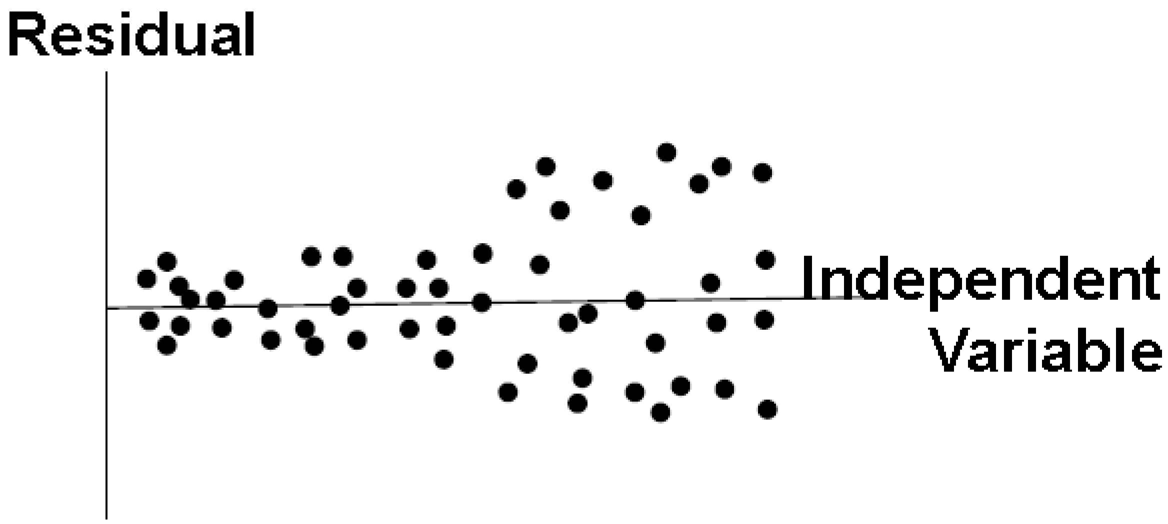
✓The standard errors are usually unreliable estimates.

✓The OLS estimators remain unbiased and consistent and asymptotically even if the error are heteroskedasticity. (但不包括有效性efficiency)

✓If the standard errors are too small, but the coefficient estimates themselves are not affected, the t-statistic will be too large and the null hypothesis of no statistical significance is rejected too often(***Type I error***). The opposite will be true if the standard errors are too large**(*Type II error*)**

Detecting Heteroskedasticity

- **Detecting heteroskedasticity with a residual plot**



✓Heteroskedasticity is not easy to correct

✓The most common remedy, however, is to calculate **robust standard errors.**

Gauss-Markov Theorem

Gauss-Markov Theorem

✓If the three least squares assumptions hold **and if the error is homoskedastic,** then the OLS estimator has the smallest variance and it is the Best Linear conditionally Unbiased Estimator （BLUE）.

$$\widetilde{\beta_1} = \sum_{i=1}^{n} a_i Y_i \; (\widetilde{\beta_1} \; is \;\; linear)$$

✓ The result extends to $\bar{Y}$ *is* the most efficient estimator of the population mean among the class of all estimators that are unbiased and are linear functions(weighted averages) of $Y_1,\ldots Y_n$

Gauss-Markov Theorem

- **Limitations**

  ✓Its conditions might not hold in practice. In particular, if the error term is heteroskedastic, then the OLS estimator is no longer BLUE. ***Alternative is to use weighted least squares estimator.***

  ✓Even if the conditions of the theorem hold, there are other candidate estimators that are not linear and conditionally unbiased; under some conditions, these other estimators are more efficient than OLS.

Example

The error term represents the portion of the:

A. dependent variable that is not explained by the independent variable(s) but could possibly be explained by adding additional independent variables.

B. dependent variable that is explained by the independent variable(s).

C. independent variables that are explained by the dependent variable.

D. dependent variable that is explained by the error in the independent variable(s).

Answer: A

The error term represents effects from independent variables not included in the model. It could be explained by additional independent variables.

Summary

- **Assumptions** (Heteroskedasticity)
- **Regression line** (Ordinary least squares ,OLS)
  - ✓Intercept
  - ✓Slope
- **Analysis of Variance (ANOVA)**
  - ✓Standard error of regression
  - ✓Coefficient of determination (R^2) : $R^2 = r^2$
- **Hypothesis Testing**
- **Gauss-Markov Theorem**

CONTENTS

PART 2

Learning objectives

- **Define and interpret** omitted variable bias, and describe the methods for addressing this bias.
- Distinguish between single and multiple regression.
- **Interpret** the slope coefficient in a multiple regression.
- Describe homoskedasticity and heterosckedasticity in a multiple regression.
- Describe the OLS estimator in a multiple regression.
- **Calculate and interpret** measures of fit in multiple regression.
- **Explain** the assumptions of the multiple linear regression model.
- **Explain** the concept of imperfect and perfect multicollinearity and their implications.

Learning objectives

- Construct, apply, and interpret hypothesis tests and confidence intervals for a single coefficient in a multiple regression.
- Construct, apply, and interpret **Joint hypothesis** tests and confidence intervals for multiple coefficients in a multiple regression.
- **Interpret** the F-statistic.
- Interpret tests of a single restriction involving multiple coefficients.
- Interpret confidence sets for multiple coefficients.
- Identify examples of omitted variable bias in multiple regressions.
- **Interpret** the R2 and adjusted R2 in a multiple regression.

Omitted variable bias

- **Definition**

✓the omitted variable is correlated with the movement of the independent variable in the model.

✓the omitted variable is a determinant of the dependent variable.

（注意：这两个条件同时成立才算遗漏变量，只满足一个不算遗漏变量！）

- **Example**

✓Percentage of English Learner(omitted variable bias)

✓Time of day the test(first condition does not hold)

✓Parking lot space per pupil(second condition does not hold)

✓莫扎特音乐能够提高智商吗？

Define and interpret (☆☆)

Omitted variable bias

- **Impact**
  - ✓Omitted variable bias occurs regardless of the size of the sample and will make OLS estimators biased and inconsistent.
  - ✓Violate $E(u_i|X_i)=0$, because $corr(u_i,X_i)\neq 0$
  - ✓The correlation between the omitted variable and the independent variable will determine the size of the bias and the direction of the bias.
  - ✓解决方法：把遗漏变量加入模型，即多元回归

The Basics of Multiple Regression

- Multiple regression is regression analysis with more than one independent variable

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i$$

✓Treat β_0 as the coefficient on a regressor that always equals 1

✓The population regression line is the relationship that holds between Y and X's on average in population

$$E(Y \mid X) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki}$$

✓Homoskedastic

✓OLS method β_i joint normally distributed and each normal distributed

Interpret (☆☆☆)

Slope Coefficient in Multiple Regression

- Partial slope coefficient
  - ✓The intercept term is the value of the dependent variable when the independent variables are all equal to zero.
  - ✓Each slope coefficient is the estimated change in the dependent variable for a one-unit change in that independent variable, holding the other independent variables constant.
  - ✓**Partial effect:** $\beta_1 = \frac{\Delta Y}{\Delta X_1}$, holding X_2, …, X_k constant, control variable

Measures of Fit

- Standard error of regression (SER)

$$SER = \sqrt{\frac{SSR}{n-k-1}} = \sqrt{\frac{\sum_{i=1}^{n} \hat{u}_i^2}{n-k-1}}$$

- Coefficient of determination (R^2)

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{SSR}{TSS}$$

✓ R^2 increases whenever a regressor is added , unless the estimated coefficient on the added regressor is exactly zero.

Measures of Fit

- Adjusted R² (or $\bar{R}^2$，注意别称！)

$$\text{Adjusted } R^2 = 1 - \left[\left(\frac{n-1}{n-k-1}\right) \times (1 - R^2)\right] = 1 - \frac{s_{\hat{\mu}}^2}{s_Y^2}$$

✓adjusted R^2 is 1 minus the ratio of the sample variance of the OLS residuals

✓As n-1/n-k-1 always greater than 1, so adjusted $R^2 \leq R^2$

✓adjusted R^2 may be less than zero

✓Adding a regressor has two opposite effect, adjusted R^2 can increases or decreases

Measures of Fit

- Potential pitfalls when using the R^2 or adjust R^2
  - ✓An increase in the R^2 or adjust R^2 does not necessarily mean that an added variable is statistically significant.
  - ✓A high R^2 or adjust R^2 does not mean that the regressors are a true cause of the dependent variable
  - ✓ A R^2 or adjust R^2 does not mean that there is no omitted variable bias.
  - ✓An increase in the R^2 or adjust R^2 does not necessarily mean you have the most appropriate set of regressors, nor does a low value mean an inappropriate set of regressors

Example

When interpreting the R^2 and adjusted R^2 measures for a multiple regression, which of the following statements incorrectly reflects a pitfall that could lead to invalid conclusions?

A. The R^2 measure does not provide evidence that the most or least appropriate independent variables have been selected.

B. If the R^2 is high, we have to assume that we have found all relevant independent variables.

C. If adding an additional independent variable to the regression improves the R^2, this variable is not necessarily statistically significant.

D. The R^2 measure may be spurious, meaning that the independent variables may show a high R2; however, they are not the exact cause of the movement in the dependent variable.

Example

Answer: B

If the R^2 is high, we cannot assume that we have found all relevant independent variables. Omitted variables may still exist, which would improve the regression results further.

Explain (☆☆)

Multiple Regression Assumptions

- **The assumptions of the multiple linear regression**
  - ✓ u_i has conditional mean zero given $X_{1i}, X_{2i}, \ldots, X_{ki}$, that is $E(u_i \mid X_{1i}, X_{2i}, \ldots, X_{ki}) = 0$
  - ✓ $(X_{1i}, X_{2i}, \ldots, X_{ki}, Y_i), i=1,\ldots,n$ are independently and identically distributed (i.i.d.) draw from their joint distribution.
  - ✓ Large outliers are unlikely: $X_{1i}, X_{2i}, \ldots, X_{ki}$ and Y_i have nonzero finite fourth moments.
  - ✓ There is no perfect multicollinearity.

Multicollinearity

- **Multicollinearity** refers to the condition when two or more of the independent variables, or linear combinations of the independent variables are highly correlated with each other.
  - ✓ **Perfect multicollinearity**
    - If one of the independent variable is a perfect linear combination of the other independent variables.
    - Produces division by zero in the OLS estimates
  - ✓ **Imperfect multicollinearity**
    - If two or more independent variables are highly correlated but less than perfectly correlated.

Explain (☆☆)

Multicollinearity

➢ Imperfect multicollinearity

✓At least one individual regressor will be imprecisely estimated, if assumptions hold, it is unbiased, but there is a greater probability that we will incorrectly conclude that a variable is not statistically significant.

✓Methods to correct multicollinearity:

- omit one or more of the correlated independent variables
- stepwise regression

Multicollinearity

- **Two methods to detect multicollinearity**

✓(1) t-tests indicate that none of the individual coefficients is significantly different than zero, while the F-test indicates overall significance and the R^2 is high.

✓(2) the absolute value of the sample correlation between any two independent variables is greater than 0.7 (i.e., $|r|>0.7$)

Explain (☆☆)

Dummy Variable Trap

- **Definition**
  - ✓ another possible source of perfect multicollinearity
  - ✓ if there are G binary variables, and there is an intercept in the regression, and if all G binary variables are included as regressors, then perfect multicollinearity happened.
  - ✓ as $G_1+G_2+G_3+G_4=1$ (注意：如果没有截距项，把所有类别包括进来就没有perfect multicollinearity)

Hypothesis Testing about the Regression Coefficient

- Significance test for a regression coefficient
  - H_0: $\beta_j = 0$

$$t = \frac{\hat{b}_j}{s_{\hat{b}_j}} \qquad df = n-k-1$$

- p-value: the smallest significance level for which the null hypothesis can be rejected
  - ✓Reject H_0 if p-value<α
  - ✓Fail to reject H_0 if p-value>α
- Regression coefficient confidence interval

$$\beta_j \pm \left(t_c \times s_{\hat{\beta}_j}\right)$$

Regression Coefficient F-test

- An F-test is used to test whether at least one slope coefficient is significantly different from zero

  H_0: $\beta_1 = \beta_2 = \beta_3 = \ldots = \beta_k = 0$

  H_a: at least one $b_j \neq 0$ (j = 1 to k)

- F-statistic
  - ✓The F-test here is always a one-tailed test
  - ✓The test assesses the effectiveness of the model as a whole in explaining the dependent variable
  - ✓**Can not use separate t test**

Interpret (☆☆)

Testing Single restrictions Involving Multipe Coefficients

- **Definition**
  - ✓ H0: $\beta_1=\beta_2$ vs H1: $\beta_1\neq\beta_2$
  - ✓ a single restriction so q=1, but involves multiple restrictions β_1 、 β_2
- **Two approach**
  - ✓ 1. Test the Restriction Directly

    Has an $F1,\infty$ distribution under the null hypothesis
  - ✓ 2. Transform the Regression, and test new coefficiant

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \mathrm{u}_i$$

$$Y_i = \beta_0 + (\beta_1 - \beta_2) X_{1i} + \beta_2 (X_{1i} + X_{2i}) + \mathrm{u}_i$$

Summary

- **Multiple regression assumptions(Multicollinearity)**
- **ANOVA Table**
  - ✓Standard error of regression (SER)
  - ✓adjusted R^2
- **Hypothesis Testing**
  - ✓ t-test
  - ✓ F-test
  - ✓Restrictions
- **Confidence Sets**

CONTENTS

PART 2

CONTENTS

PART 2

Learning objectives

- Describe linear and nonlinear trends.
- Describe trend models to estimate and forecast trends.
- Compare and evaluate model selection criteria, including mean squared error (MSE), s^2, the Akaike information criterion (AIC), and the Schwarz information criterion (SIC).
- Explain the necessary conditions for a model selection criterion to demonstrate consistency.

Describe(☆)

Modeling Trend

- **Trend** is slow, long-run evolution in the variables that we want to model and forecast.
  - ✓In business, finance, and economics, trend is produced by slowly evolving preferences, technologies, institutions, and demographics.
  - ✓Focus on **deterministic trend**: trend evolves in a perfectly predictable way
- Numerous series in diverse fields display trends.

Linear and Nonlinear Trend

- **Linear Trend**

$$T_t = \beta_0 + \beta_1 TIME_t$$

The variable TIME: a **time trend or time dummy**;

β_0 regression intercept, the value of trend at time 0;

β_1 regression slope, positive if trending is increasing.

- **Nonlinear Trend**

don't require that trend be linear, only that they be smooth

✓ **Quadratic trend**

$$T_t = \beta_0 + \beta_1 TIME_t + \beta_2 TIME_t^2$$

$$= \beta_2 \left(TIME + \frac{\beta_1}{2\beta_2}\right)^2 + C$$

Describe(☆)

Linear and Nonlinear Trend

note that when β_1>0, β_2>0 and β_1<0, β_2<0, trend is monotonically

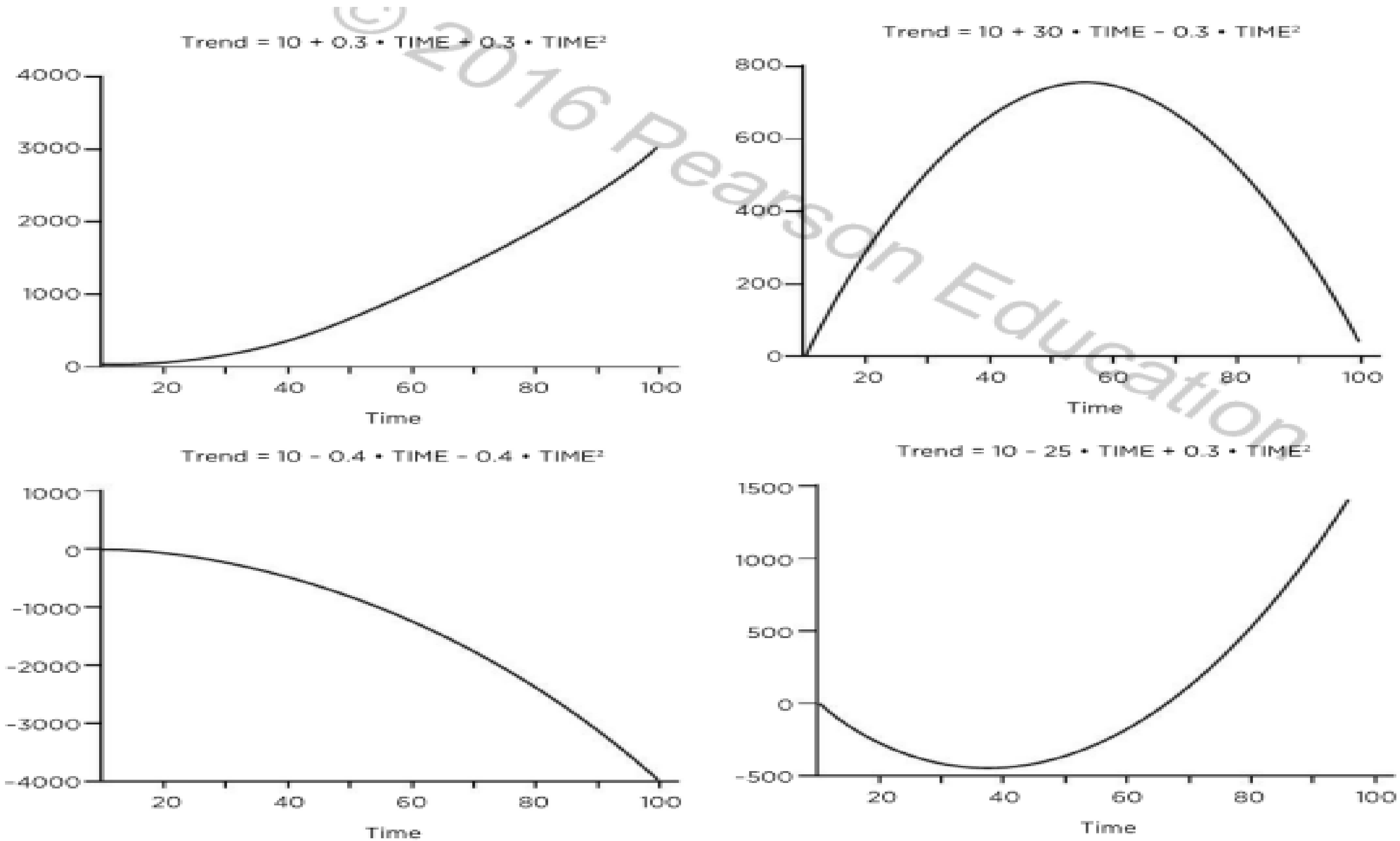


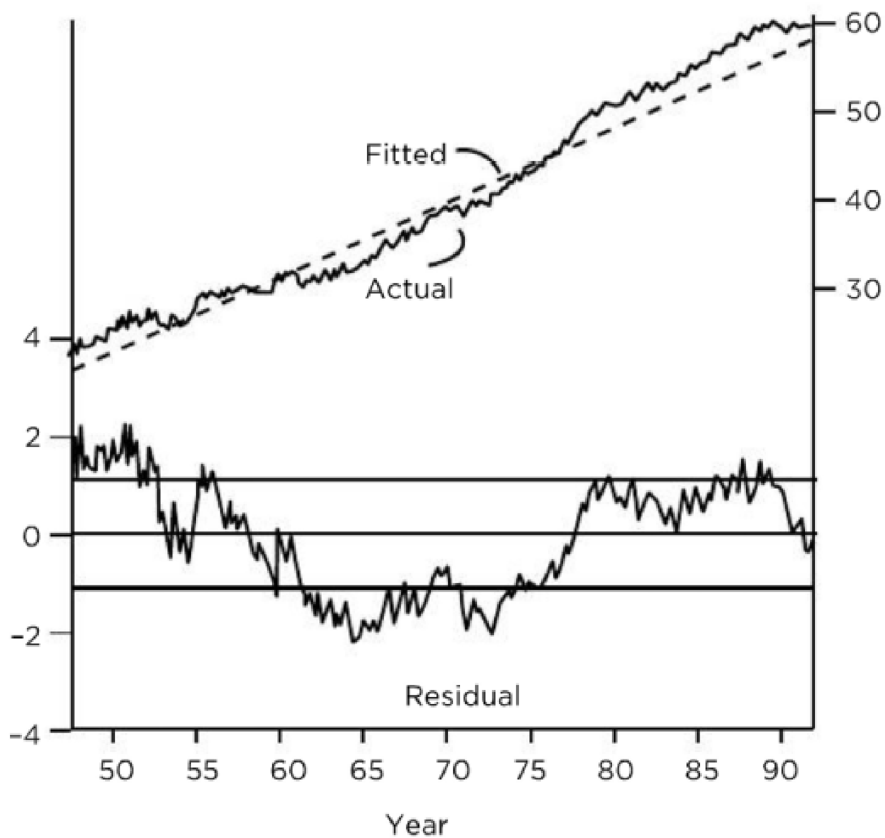
FIGURE 10-7 Various shapes of quadratic trends.

Describe(☆)

Linear and Nonlinear Trend

The quadratic trend is still a little more U-shaped than the volume data, but it still has some awkward features.

FIGURE 10-3 Labor force participation rate, males.



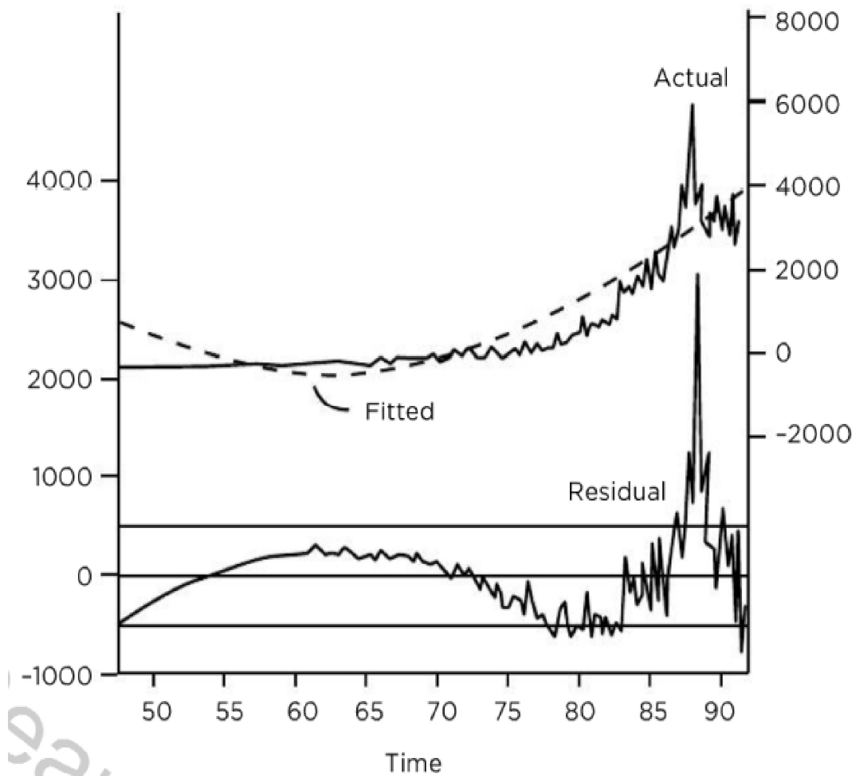


FIGURE 10-8 Quadratic trend, volume on the New York Stock Exchange.

Describe(☆☆)

Log-linear Trend

- **Exponential Trend(log-linear trend)**

  a trend appears nonlinear in levels but linear in logarithms.

  ✓very common in business, finance, and economics

  ✓Economic variable often display roughly constant growth rates （β_1）

$$T_t = \beta_0 e^{\beta_1 TIME_t}$$

$$\ln(T_t) = \ln(\beta_0) + \beta_1 TIME_t$$

- **Log-linear vs Quadratic**

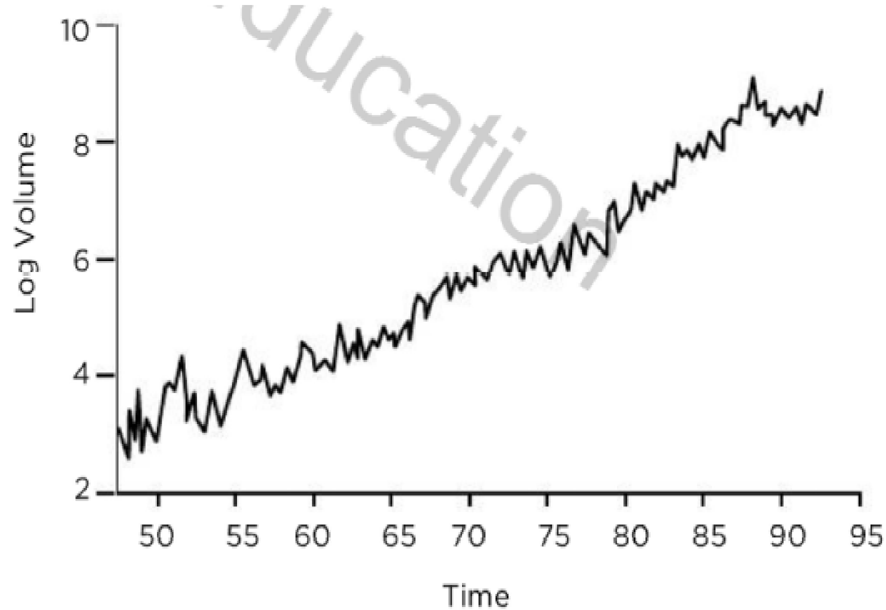


FIGURE 10-9 Log volume on the New York Stock Exchange.

Loglinear Trend

Describe(☆)

β_1表示增长率，β_0表示初始值

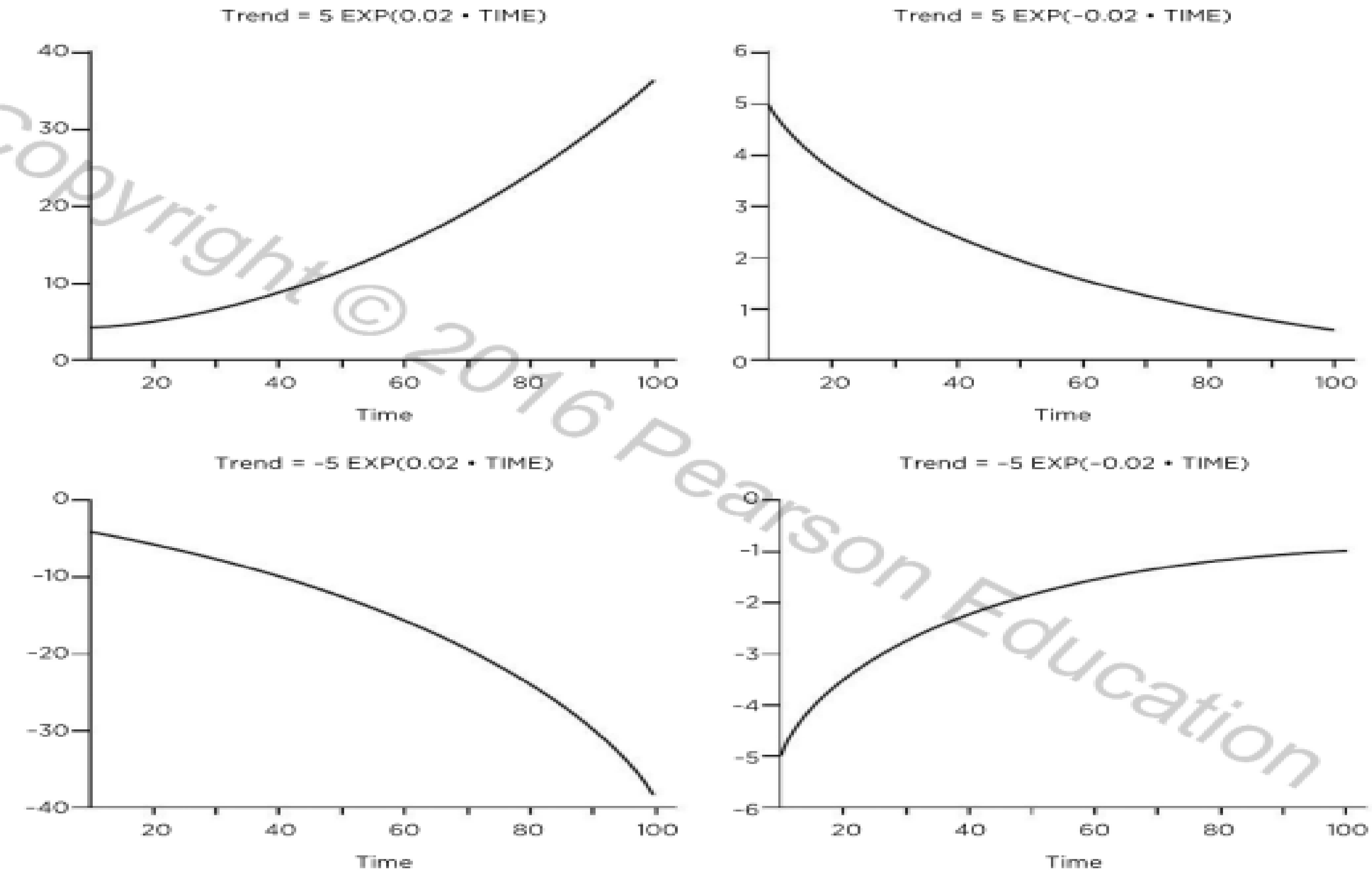


FIGURE 10-10 Various shapes of exponential trends.

Estimating and Forecasting Trend

- **Estimating**

  We fit our various trend models to data on a time series *y* *using* ***least squares*** **regression. That is, we use a computer to find**

$$(\hat{\beta}_0, \hat{\beta}_1) = \arg\min_{\beta_0, \beta_1} \sum_{t=1}^{T} (y_t - \beta_0 - \beta_1 TIME_t)^2$$

- **Forecasting**

  forecast the h-step-ahead value of a series y is made at time T

$$y_{T+h} = \beta_0 + \beta_1 TIME_{T+h} + \varepsilon_{T+h}$$

$$\hat{y}_{T+h} = \hat{\beta}_0 + \hat{\beta}_1 TIME_{T+h}$$

Forecast model selection criteria

- Most model selection criteria attempt to find model with the smallest out-of-sample 1-step-ahead mean square prediction error **(the smaller, the better).**

✓mean squared error (MSE)

✓mean squared error corrected for degrees of freedom (s^2)

✓Akaike information criterion (AIC)

✓the Schwarz information criterion (SIC)

Differences among criteria amount to different penalties for the number of degrees of freedom used in estimating the model

Define and explain (☆)

Forecast model selection criteria

- **Mean squared error (MSE)** is a statistical measure computed as the sum of squared residuals divided by the total number of observations in the sample

$$\mathrm{MSE} = \frac{\sum_{t=1}^{T} e_t^2}{T}$$

where: $e_t = y_t - \hat{y}_t$

T = total sample size

$y_t = \hat{\beta}_0 + \hat{\beta}_1 TIME_t$ (i.e., a regression model)

✓closely related to the coefficient of determination (R^2).

Forecast model selection criteria

- **s^2 measure**

an **unbiased estimate** of the regression disturbance because it corrects for degrees of freedom as follows:

$$s^2=\left(\frac{T}{T-k}\right)\frac{\sum_{t=1}^{T}e_t^2}{T} \quad \text{P.S.:}\ \frac{\sum_{t=1}^{T}e_t^2}{T}=MSE$$

s^2 measure is equivalent to minimizes the standard error of the regression.(k is the number of degrees of freedom)

- ✓ s2 measure intimately connected to the adjusted R2

Forecast model selection criteria

- **Akaike information criterion(AIC)**

$$AIC=e^{(\frac{2k}{T})}\frac{\sum_{t=1}^{T}e_t^2}{T}$$

- **Schwarz information criterion (SIC)**

$$SIC=T^{(\frac{k}{T})}\frac{\sum_{t=1}^{T}e_t^2}{T}$$

All of the penalty factors are functions of k/T

Most time, AIC and SIC select the same model, when they don't, use SIC.

Forecast model selection criteria

- The SIC generally ***penalizes free parameters more strongly*** than does the Akaike information criterion, though it depends on the size of T and relative magnitude of T and k.

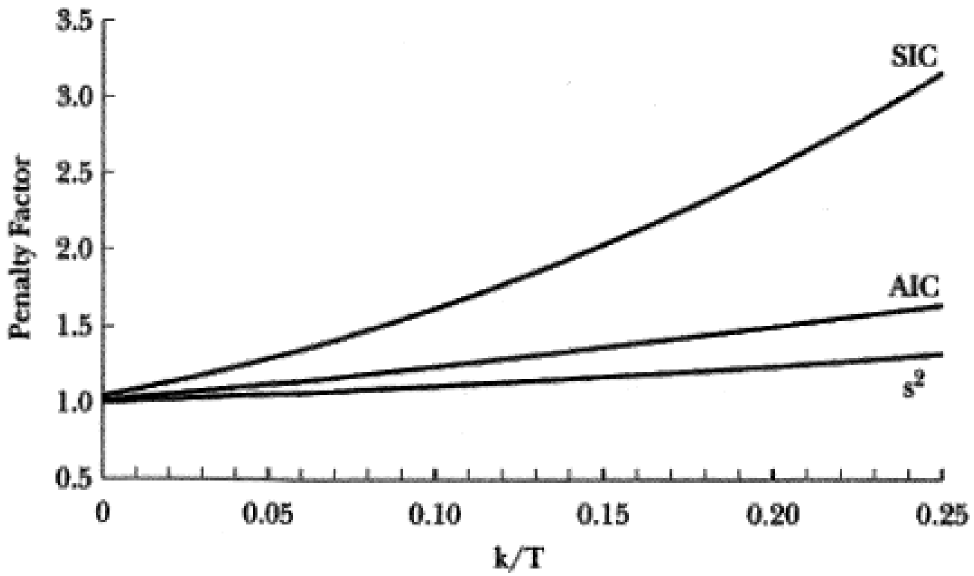


FIGURE 11-1 Degrees-of-freedom penalties, various model selection criteria.

Explain (☆☆)

Forecast model selection criteria

- How to evaluate model selection criteria: **consistency**
  - ✓probability of selecting or approximation true data-generating process(DGP) approaches 1 as sample size gets large
- AIC is **asymptotically efficient** while the SIC is not.

Criteria	Consistency
MSE	inconsistent
S^2	inconsistent
AIC	Inconsistent
SIC	Consistent

Example

Richard Frank, FRM, is running a regression model to forecast in-sample data. He is concerned about data mining and over-fitting the data. Which of the following criteria provides the highest penalty factor based on degrees of freedom?

A. Mean squared error (MSE).

B. Unbiased mean squared error (s^2).

C. Akaike information criterion (AIC).

D. Schwarz information criterion (SIC).

Example

Answer: D

The Schwarz information criterion (SIC) has the highest penalty factor. The mean squared error (MSE) does not penalize the regression model based on the increased number of parameters, k. The penalty factors for s^2, AIC, and SIC are (T / T - k), $e^{(2k/T)}$, and $T^{(k/T)}$, respectively. Thus, SIC has the greatest penalty factor.

Summary

- **Trend Model**
  - ✓Linear
  - ✓Non linear: quadratic vs log-linear
- **Model selection criteria**
  - ✓ Mean squared error (MSE)
  - ✓Mean squared error corrected for degrees of freedom (s2)
  - ✓Akaike information criterion (AIC)
  - ✓Schwarz information criterion (SIC)

CONTENTS

PART 2

Learning objectives

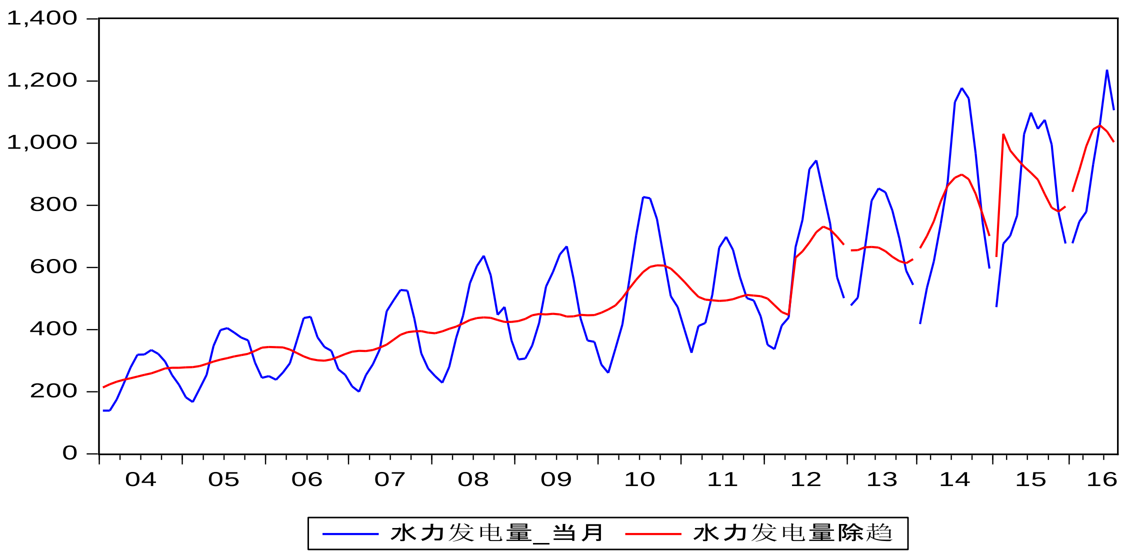
- Describe the sources of seasonality and how to deal with it in time series analysis.
- **Explain** how to use regression analysis to model seasonality.
- Explain how to construct an h-step-ahead point forecast.

Describe (☆)

Seasonality

➢ Definition

- ✓ Deterministic vs stochastic seasonality
- ✓ Sources: weather, preferences, social institutions, technologies
- ✓ pervasive in business and economics



Seasonality

- **How to deal with seasonality**
  - ✓ nonseasonal fluctuations is appropriate for macroeconomics, but not appropriate for business forecast
  - ✓ regression on seasonal dummies

$$y_t = \beta_1 TIME_t + \sum_{i=1}^{s} \gamma_i D_{it} + \epsilon_t$$

include an intercept and a full set of seasonal dummies produces perfect multicollinearity. But full set of s seasonal dummies sums to a variable whose value is always 1.

(including an intercept is equivalent to including a variable whose value is always 1)

Describe (☆)

Seasonality

- **Calendar effect**
  - ✓ **holiday variation:** some holidays' dates changes over time
  - ✓ **Trading-day variation**: different months contain different numbers of trading days or business days.

$$y_t = \beta_1 TIME_t + \sum_{i=1}^{s} \gamma_i D_{it} + \sum_{i=1}^{v1} \sigma_i^{HD} HDV_{it} + \sum_{i=1}^{v2} \sigma_i^{TD} TDV_{it} + \varepsilon_t$$

HDVs are the holiday variables, TDVs are trading-day variables, whose value each month is the number of trading days that month.

- **Forecasting**

Summary

- **Source of Seasonality**
- **How to deal with seasonality**
- **Calendar effect**

CONTENTS

PART 2

Learning objectives

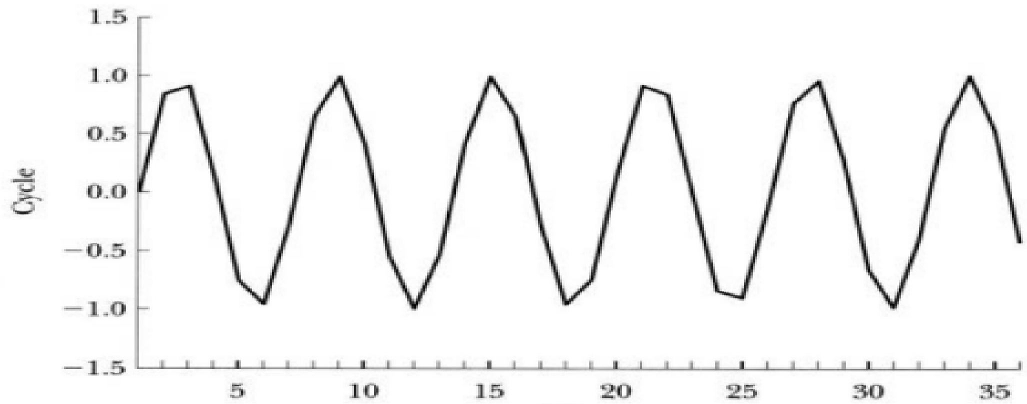
- Define covariance stationary, autocovariance function, autocorrelation function, partial autocorrelation function and autoregression
- Describe the requirements for a series to be covariance stationary
- Explain the implications of working with models that are not covariance stationary
- Define white noise, and describe independent white noise and normal (Gaussian) white noise
- Explain the characteristics of the dynamic structure of white noise
- Explain how a lag operator works.
- Describe Wold's theorem.
- Define a general linear process.
- Relate rational distributed lags to Wold's theorem
- **Calculate** the sample mean and sample autocorrelation, and describe the Box-Pierce Q- statistic and the Ljung-Box Q-statistic.
- Describe sample partial autocorrelation

Describe (☆☆)

Characterizing Cycles

- **Cycle:** any sort of dynamics not captured by trends or seasonals
- **Time Series**
  - ✓ In theory, a time series realization begins in the infinite past and continues into the infinite future

$$\{\ldots, y_{-2}, y_{-1}, y_0, y_1, y_2, \ldots\}$$



Describe (☆☆)

Characterizing Cycles

➢ **Covariance Stationary**

✓A time series is covariance stationary if it satisfies the following three conditions:

- Constant and finite expected value.
- Constant and finite variance.
- Constant and finite covariance between values at any given lag.

$$\gamma(t, \tau) = cov(y_t, y_{t-\tau}) = E(y_t - \mu)(y_{t-\tau} - \mu) = \gamma(\tau)$$

Define (☆☆)

Characterizing Cycles

➤ Autocovariance function

$$\gamma(t,\tau) = cov(y_t, y_{t-\tau}) = E(y_t - \mu)(y_{t-\tau} - \mu) = \gamma(\tau)$$

if the covariance structure is stable over time, then the $\gamma(t,\tau)$ depend only on displacement $\boldsymbol{\tau}$, not on time t:

- ✓ Symmetric: $\gamma(\tau) = \gamma(-\tau)$
- ✓ $\gamma(0) = cov(y_t, y_t) = var(y_t)$
- ✓ many economic, financial seires are not covariance stationary

  e.g.: increasing mean, seasonality corresponds to means that vary with the season.
- ✓ clearly nonstationary in levels appear covariance stationary in growth rate

Characterizing Cycles

- **Autocorrelation function (ACF):** $\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)}$, refers to the degree of correlation and interdependency between data points in a time series.
- **Partial autocorrelation function (PCF) :** p(τ) is the coefficient of $y_{t-\tau}$ in a population linear regression of y_t on $y_{t\text{-}1}$,…, $y_{t-\tau}$ measure the association between yt and $yt{-}\tau$ after controlling for the effects of y_{t-1},,…, $y_{t-\tau+1}$.
- All of the covariance stationary processes that in FRM have ACF and PCF that approaches 0.

Characterizing Cycles

- A **white noise** process is a time series process with a zero mean, constant variance, and no serial correlation.

$$\in_t \sim(0,\sigma^2)$$

- Even though a white noise process is serially uncorrelated, it may not be serially independent or normally distributed.
- If $\in_t$ is serially independent, then we say $\in_t$ is independent white noise.
- If $\in_t$ is serially uncorrelated and normally distributed, then we say $\in_t$ is normal white noise or Gaussian white noise.

$$y_t \sim iid\ N(0,\sigma^2)$$

White Noise

- Realized of white noise process: no patterns of any kind in the seires
- The **dynamic structure** of a white noise process includes the following characteristics:

✓The lack of any correlation in white noise means that all autocovariances and autocorrelations are zero beyond displacement zero.

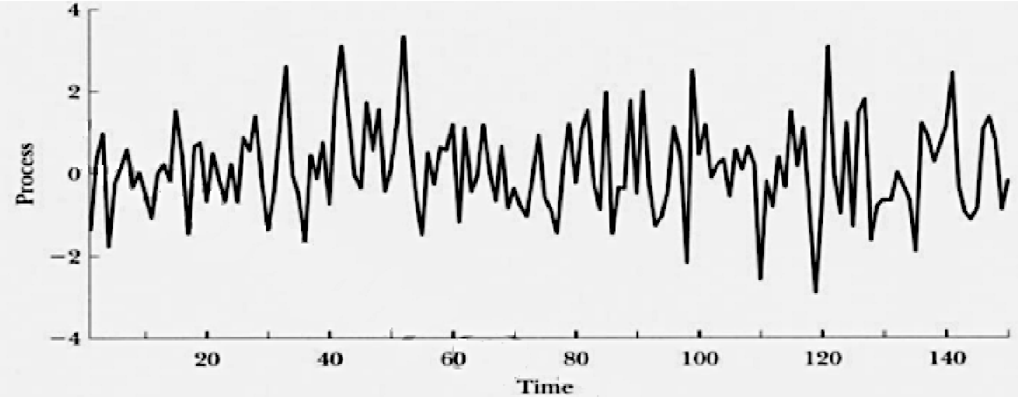


FIGURE 12-6 Realization of white noise process.

White Noise

- Why need White Noise?
  - ✓ processes with much richer dynamics are built up by taking simple transformations of white noise
  - ✓ 1-step-ahead forecast errors from good models should be white noise

- Events in a white noise process exhibit no correlation between the past and present.

Lag Operators

- A lag operator quantifies how a time series evolves by lagging a data series. For example, a lag operator, *L, operates on series, y, by lagging it as follows:*

$$\mathbf{Ly_t = y_{t-1}}$$

- first-difference operator (Δ), which applies a polynomial in the lag operator as follows:

$$\mathbf{\Delta y_t = (1\text{-}L)y_t = y_t - y_{t-1}}$$

- Polynomial in the lag operator: B(L)=$b_0 + b_1$L+… +$b_m L^m$

$$(1 + 0.9L + 0.6L^2)y_t = y_t + 0.9y_{t-1} + 0.6y_{t-2}$$

Describe (☆☆)

Wold's theorem

- **Wold's representation Theorem** is a model for the covariance stationary residual (i.e., a model that is constructed after making provisions for trends and seasonal components).
- **Wold's representation** utilizes an infinite number of distributed lags, where the one-step-ahead forecasted error terms are known as ***innovations***.
  - ✓ Innovations are not necessarily Gaussian.

Describe (☆)

Wold's theorem

- Let y_t be any zero-mean covariance-stationary process.

  Then we can write it as

$$y_t = B(L)\varepsilon_t = \sum_{i=0}^{\infty} b_t \varepsilon_{t-i}$$

$$\varepsilon_t \sim WN(0, \sigma^2)$$

Where $b_0 = 1\ and\ \sum_{i=0}^{\infty} b_i^2 < \infty$

- The **general linear process**

$$y_t = B(L)\varepsilon_t = \sum_{i=0}^{\infty} b_t \varepsilon_{t-i}$$

Describe (☆)

Rational Distributed Lags

- Infinite polynomials can be expressed as a ratio of finite-order polynomials are known as rational polynomials. The distributed lags constructed from these rational polynomials are known as rational distributed lags.

$$B(L) = \frac{\theta(L)}{\varphi(L)}$$

✓ numerator polynomial is of degree q, denominator polynomial is of degree p, so there are only p+q parameters

Sample Mean and Autocorrelation

- **Sample mean**

$$\bar{y}=\frac{1}{T}\sum_{t=1}^{T} y_t$$

- **Sample autocorrelation function**

$$\hat{\rho}(\tau)=\frac{\frac{1}{T}\sum_{t=\tau+1}^{T}[(y_t-\bar{y})(y_{t-\tau}-\bar{y})]}{\frac{1}{T}\sum_{t=1}^{T}(y_t-\bar{y})^2}$$

- **Sample Partial Autocorrelations**

$$\hat{y}=\hat{c}+\hat{\beta}_1 y_{t-1}+\ldots+\hat{\beta}_\tau y_{t-\tau}$$

Box-Pierce Q-statistic

- **Whether a series is white noise—**all its autocorrelations are jointly 0.
  - ✓ if white noise then $\hat{\rho}(\tau)$ ~N(0,1/T)
- Box-Pierce Q-statistic is approximately distributed as a random variable χ2 under the null hypothesis that y is white noise (ρ(τ)=0).

$$Q_{BP} = T\sum_{\tau=1}^{m}\hat{\rho}^2(\tau)$$

- **Q-statistics**泛指Box-Pierce Q-statistic和Ljung-Box Q-statistic

Ljung-Box Q-statistic

- **Ljung-Box Q-statistic**
  - ✓under the null hypothesis that y is white noise ($\rho(\tau)=0$), QLB is approximately distributed as a χ_m^2 random variable.

$$Q_{LP} = T(T+2)\left(\frac{1}{T-\tau}\right)\sum_{\tau=1}^{m}\hat{\rho}^2(\tau)$$

- **m selection**
  - ✓m cannot be too small
  - ✓m cannot be too large: quality of the distributional approximation deteriorates
  - ✓m in the neighborhood of $\sqrt{T}$ *is reasonable*

Example

All of the following traits characterize the covariance stationarity of a time series process, except:

A. stability of the mean.

B. stability of the covariance structure.

C. a nonconstant variance in the time series.

D. stability of the autocorrelation.

Answer: C

The time series volatility around its mean (i.e., the distribution of the individual observations around the mean) does not change over time.

Summary

- **Time series: Covariance Stationary**
- **Characterizing Cycles**
  - ✓Autoregression: autoregressive (AR) model
  - ✓Autocovariance function
  - ✓Autocorrelation function
  - ✓Partial autocorrelation function
- **Wold's representation Theorem**
- **Sample autocorrelation function**
  - ✓Box-Pierce Q-statistic
  - ✓Ljung-Box Q-statistic

CONTENTS

PART 2

Learning objectives

- Describe the properties of the first-order moving average (MA(1)) process, and distinguish between autoregressive representation and moving average representation
- Describe the properties of a general finite-order process of order q (MA(q)) process
- Describe the properties of the first-order autoregressive (AR(1)) process, and define and **explain the Yule-Walker equation**.
- Describe the properties of a general pth order autoregressive (AR(p)) process
- Define and describe the properties of the autoregressive moving average (ARMA) process.
- Describe the application of AR and ARMA processes.

Moving Average Process

- **Why MA model**
  - ✓Obvious approximation to the Wold representation
  - ✓Modeling time series as distributed lags of current and past shocks
- **MA(1) model**

$$y_t = \varepsilon_t + \theta\varepsilon_{t-1} = (1+\theta L)\varepsilon_t;$$

$$\varepsilon_t \sim WN(0,\sigma^2)$$

  - ✓covariance stationary

Moving Average Process

- **ACF(cutoff in the autocorrelation)**

$$\gamma(\tau) = E(y_t y_{t-\tau}) = E((\varepsilon_t + \theta\varepsilon_{t-1})(\varepsilon_{t-\tau} + \theta\varepsilon_{t-\tau-1}))$$
$$= \theta\sigma^2, \tau = 1$$
$$= 0, otherwise$$

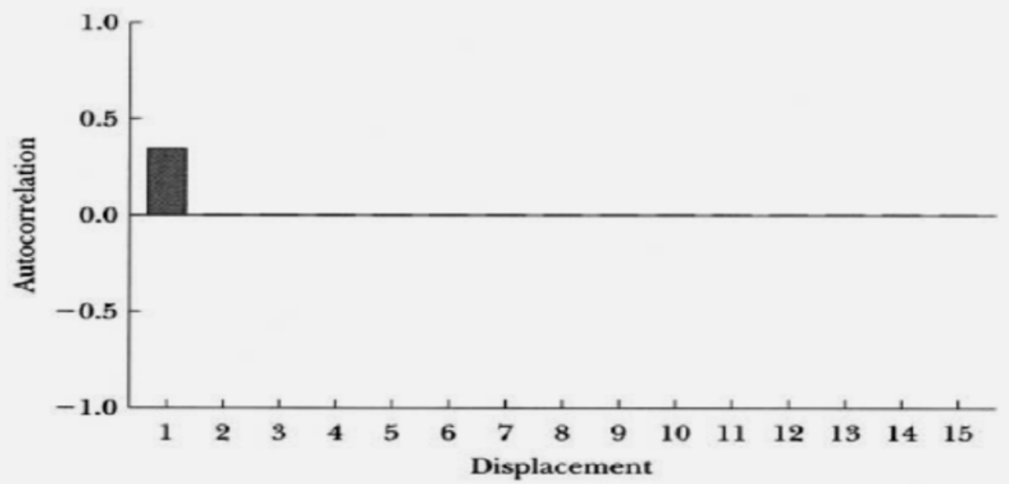


FIGURE 13-2 Population autocorrelation function, MA(1) process, θ = 0.4.

Moving Average Process

- **PCF(decay)**

✓ $\varepsilon_t = y_t - \theta\varepsilon_{t-1}$

$$y_t = \varepsilon_t + \theta y_{t-1} - \theta^2 y_{t-2} + \theta^3 y_{t-3} + \dots$$

$$\frac{1}{1+\theta L} y_t = \varepsilon_t \qquad |\theta| < 1, \text{ MA is invertible}$$

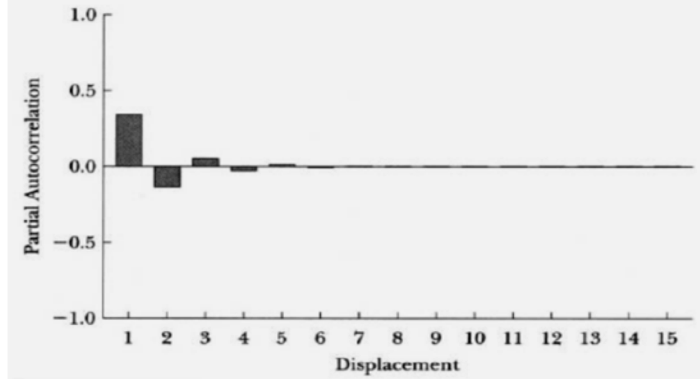


FIGURE 13-4 Population partial autocorrelation function, MA(1) process, θ = 0.4.

- $\theta > 0$, *then the patten of decay will be* one of damped oscillation; otherwise, the decay

- **MA(q)**

$$y_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \dots + \theta_q\varepsilon_{t-q} = \Theta(L)\varepsilon_t;\ \varepsilon_t : WN(0,\sigma^2)$$

$$\Theta(L) = 1 + \theta_1 L + \dots + \theta_q L^q$$

Describe (☆☆)

高顿财经 GOLDEN FINANCE

First-Order Autoregressive Process

- **Why AR model**

✓Current value of a series is linearly related to its past values, plus an additive stochastic shock.

- **AR(1) process**

$$y_t = \varphi y_{t-1} + \varepsilon_t ; \varepsilon \sim WN(0, \sigma^2)$$
$$(1 - \varphi L) y_t = \varepsilon_t$$

✓The AR(1) model is capable of capturing much more persistent dynamics than is the MA(1).

✓|Φ|<1 is the condition for covariance stationary in the AR(1).

Define (☆)

Yule-Walker Equation

- **ACF（Yule-Walker Equation）**

$$y_t = \varphi y_{t-1} + \varepsilon_t$$

Multiplying $y_{t-\tau}$ $\quad y_t y_{t-\tau} = \varphi y_{t-1} y_{t-\tau} + \varepsilon_t y_{t-\tau}$

Taking expectation $\quad \gamma(\tau) = \varphi\, \gamma(\tau - 1)$

In general $\quad \gamma(\tau) = \varphi^{\tau} \dfrac{\sigma^2}{1-\varphi^2}, \tau \quad 0,1,2,$

Note the gradual autocorrelation decay, which is typical of autoregressive processes. The autocorrelations approach 0, but only in the limit as the displacement approaches infinity.

- **PCF(cut off)**

$$y_t = \varphi y_{t-1} + \varepsilon_t$$

ACF vs PCF

	ACF	PCF
MA(1) model	cut off	Decay
AR(1) model	Decay	Cut off

Yule-Walker Equation: $\gamma(\tau) = \varphi\ \gamma(\tau - 1)$

AR(p) and ARMA(p,q)

AR(p) model

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \ldots + \varphi_p y_{t-p} + \varepsilon_t ; \varepsilon_t \sim WN(0, \sigma^2)$$

$$\Phi(L) y_t = (1 - \varphi_1 L - \varphi_2 L^2 + \ldots - \varphi_p L^p) y_t = \varepsilon_t$$

ARMA(p,q) model

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \ldots + \varphi_p y_{t-p} + \theta \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t ; \varepsilon_t \sim WN(0, \sigma^2)$$

- ✓ARMA models are often both highly accurate and highly parsimonious.
- ✓why ARMA model: 1. random shock that drives an autoregressive process; 2. sum of AR processes or sum of AR and MA process; 3. AR processes subject to measurement error

Example

Which of the following statements is a key differentiator between a moving average (MA) representation and an autoregressive (AR) process?

A. A moving average representation shows evidence of autocorrelation cutoff.

B. An autoregressive process shows evidence of autocorrelation cutoff.

C. An unadjusted moving average process shows evidence of gradual autocorrelation decay.

D. An autoregressive process is never covariance stationary.

Example

Answer: A

A key difference between a moving average (MA) representation and an autoregressive (AR) process is that the MA process shows autocorrelation cutoff while an AR process shows a gradual decay in autocorrelations.

Summary

- **MA Model**

✓MA(1) $y_t = \varepsilon_t + \theta\varepsilon_{t-1} = (1+\theta L)\varepsilon_t; \varepsilon_t : WN(0,\sigma^2)$

- **AR Model**

✓AR(1)

$$y_t = \varphi y_{t-1} + \varepsilon_t; \varepsilon : WN(0,\sigma^2)$$

$$(1-\varphi L)y_t = \varepsilon_t$$

✓Yule-Walker equation: $\gamma(\tau) = \varphi\,\gamma(\tau-1)$

- **ARMA Model**

✓ARMA(p,q)

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \ldots + \varphi_p y_{t-p} + \theta\varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t; \varepsilon_t : WN(0,\sigma^2)$$

CONTENTS

PART 2

1. Probability Theory
2. Statistics
3. Linear Regression
4. Time-Series Analysis
5. Simulation Method

4.1 Modeling and Forecasting Trend

4.2 Modeling and Forecasting Seasonality

4.3 Characterizing Cycles

4.4 Modeling Cycles: MA AR, and ARMA Models

4.5 Volatility (☆☆☆)

4.6. Correlations and Copulas

Learning objectives

- **Define and distinguish between volatility, variance rate, and implied volatility.**
- **Describe the power law.**
- **Explain** how various weighting schemes can be used in **estimating** volatility.
- Apply the exponentially weighted moving average (EWMA) model to estimate volatility.
- Describe the generalized autoregressive conditional heteroskedasticity (GARCH(p,q)) model for estimating volatility and its properties.
- **Calculate** volatility using the **GARCH(1,1) model**.
- **Explain** mean reversion and how it is captured in the GARCH(1,1) model.
- **Explain** the weights in the EWMA and GARCH(1,1) models.
- **Explain** how GARCH models perform in volatility forecasting.
- Describe the volatility term structure and the impact of volatility changes.

Explain (☆)

Definition of Volatility

- **Definition**
  - ✓defined as the standard deviation of the return provided by the variable per unit of time when the return is expressed using continuous compounding.
  - ✓Define S_i as the value of market variable at end of day I
  - ✓Define continuously compounded return as $u_i = \ln(S_i/S_{i-1})$
  - ✓Variance of the return over T days is T times the variance of the return over one day ($\sqrt{T}$ standard deviation of the return over one day)

Variance Rate and Implied Volatility

- **Variance Rate**
  - ✓Risk manage focus on the variance rather than the volatility.
  - ✓Is defined as the square of the volatility
  - ✓Business day: 252 days per year. $\sigma_{yr} = \sigma_{day}\sqrt{252}$
- **Implied Volatilities**
  - ✓the volatility that gives the market price of the option when it is substituted into the BSM option pricing model
  - ✓ VIX index (bet only on volatility)

The Power Law

Definition

✓An alternative to assuming normal distribution.

✓The law asserts that, for many variables that are encountered in practice, it is approximately true that the value of the variable, v, has the property that when x is large

$$\text{Prob}(v > x) = Kx^{-\alpha}$$

✓K and α are constants.

✓variables v as diverse as the income of an individual, the size of a city, the number of visits to a website in a day

Explain (☆) 高顿财经 GOLDEN FINANCE

Daily Volatility

- **Daily volatility**
- Define $u_i = \ln(S_i/S_{i-1})$

$$\sigma_n^2 = \frac{1}{m-1}\sum_{i=1}^{m}(u_{n-1} - \bar{u})^2 \qquad \bar{u} = \frac{1}{m}\sum_{i=1}^{m} u_{n-i}$$

- Change in a number of ways
  - ✓Define u_i as $(S_i - S_i - 1)/S_{i-1}$
  - ✓Assume that the mean value of ui is zero
  - ✓Replace m-1 by m
  - ✓This gives $\sigma_n^2 = \frac{1}{m}\sum_{i=1}^{m} u_{n-i}^2$
  - ✓Volatility σ_n is estimated at the end of day n-1

Explain (☆)

Weighting Scheme

- Instead of assigning equal weights to the observations we can set

$$\sigma_n^2 = \sum_{i=1}^{m} \alpha_i u_{n-i}^2$$

where

$$\sum_{i=1}^{m} \alpha_i = 1$$

Reason: More reasonable to give more weight to recent data, therefore : if i>j, $\alpha_i < \alpha_j$

ARCH(m) Model

- In an ARCH(m) model we also assign some weight to the long-run variance rate, V_L , (an extension to the previous weighted model)

$$\sigma_n^2 = \gamma V_L + \sum_{i=1}^{m} \alpha_i u_{n-i}^2$$

where

$$\gamma + \sum_{i=1}^{m} \alpha_i = 1$$

Setting w = γV_L, the ARCH (m) model is

EWMA Model

- In an exponentially weighted moving average model, the weights assigned to the u^2 decline exponentially as we move back through time $\alpha_i = (1-\lambda)\lambda^{i-1}$

This leads to

$$\sigma_n^2 = \lambda\sigma_{n-1}^2 \quad (1+\lambda)u_{n-1}^2$$

$$= \lambda(\lambda\sigma_{n-2}^2 \quad (1 \quad \lambda)u_{n-2}^2) \quad (1 \quad \lambda)u_{n-1}^2$$

$$= (1 \quad \lambda)(u_{n-1}^2 \quad \lambda u_{n-2}^2) \quad \lambda^2\sigma_{n-2}^2$$

$$=(1-\lambda)\sum_{i=1}^{m}\lambda^{i-1}u_{n-i}^2 + \lambda^{m}\sigma_{n-m}^2$$

Explain (☆☆☆)

高顿财经 GOLDEN FINANCE

Attractions of EWMA

- Relatively little data needs to be stored:need only remember the current estimate of the variance rate and the most recent observation on the market variable
- Tracks volatility changes
  - ✓the value of λ governs how responsive the estimate of the daily volatility is to the most recent daily percentage change
  - ✓high λ respond relatively slow to new information provided by the daily percentage changes
- RiskMetrics uses λ= 0.94 for daily volatility forecasting

GARCH (1,1)

- In GARCH (1,1) we assign some weight to the long-run average variance rate

$$\sigma_n^2 = \gamma V_L + \alpha u_{n-1}^2 + \beta \sigma_{n-1}^2$$

- Since weights must sum to 1 γ+ α + β =1
- The EWMA model is a particular case of the GARCH(1,1) model where γ =0, α =1-λ, and β = λ
- Mean reversion

GARCH (1,1)

- Setting w = γV_L, the GARCH (1,1) model is

$$\sigma_n^2 = \omega + \alpha u_{n-1}^2 + \beta \sigma_{n-1}^2$$

and

$$V_L = \frac{\omega}{1-\alpha-\beta} \qquad \alpha+\beta < 1$$

- Garch(p,q)

$$\sigma_n^2 = \omega + \sum_{i=1}^{p} \alpha_i u_{n-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{n-j}^2$$

Example

- Suppose

$$\sigma_n^2 = 0.000002 + 0.13u_{n-1}^2 + 0.86\sigma_{n-1}^2$$

✓The long-run variance rate is 0.0002 so that the long-run volatility per day is 1.4%. Suppose that the current estimate of the volatility is 1.6% per day and the most recent percentage change in the market variable is 1%. The new variance rate is

$$0.000002 + 0.13 \times 0.0001 + 0.86 \times 0.000256 = 0.00023336$$

✓The new volatility is 1.53% per day

Choose between the Models

- In practice, variance rates tend to be mean reverting
- The GARCH (1,1) model incorporates mean reversion, whereas the EWMA model does not
- Therefore GARCH (1,1) is theoretically more appealing than the EWMA model
- When $\omega = 0$, GARCH (1,1) reduces to EWMA
- When $\omega < 0$, GARCH (1,1) is not stable, should switch to EWMA

Explain (☆)

Forecast Future Volatility with GARCH (1,1)

$$\sigma_n^2 = (1-\alpha-\beta)V_L + \alpha u_{n-1}^2 + \beta\sigma_{n-1}^2$$

$$\downarrow$$

$$\sigma_n^2 - V_L = \alpha(u_{n-1}^2 - V_L) + \beta(\sigma_{n-1}^2 - V_L)$$

$$\downarrow$$

$$\sigma_{n+t}^2 - V_L = \alpha(u_{n+t-1}^2 - V_L) + \beta(\sigma_{n+t-1}^2 - V_L)$$

$$\Big\downarrow \; E(u_{n+t-1}^2) = \sigma_{n+t-1}^2$$

$$E[\sigma_{n+t}^2 - V_L] = (\alpha+\beta)E[\sigma_{n+t-1}^2 - V_L]$$

$$\Big\downarrow \; \text{Using this equation repeatedly}$$

$$E[\sigma_{n+t}^2 - V_L] = (\alpha+\beta)^t(\sigma_n^2 - V_L)$$

$$\downarrow$$

$$\boxed{E[\sigma_{n+t}^2] = V_L + (\alpha+\beta)^t(\sigma_n^2 - V_L)}$$

- The larger T is, the closer this is to V_L
- The equation can be used to estimate a volatility term structure

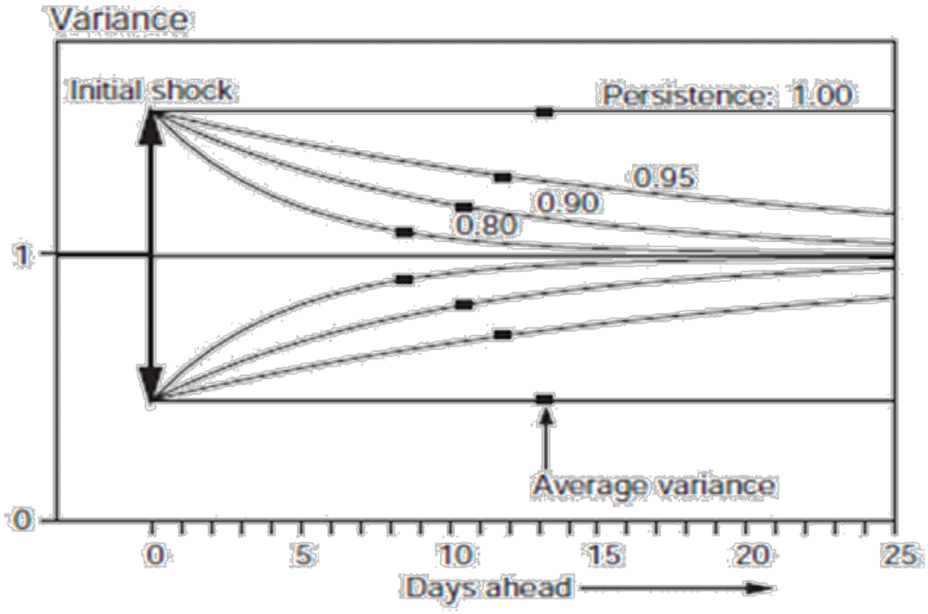
Explain (☆☆)

Forecast Future Volatility with GARCH (1,1)

➢ **GARCH(1, 1) Model**

✓Persistence: $\alpha+\beta$

✓It defines the speed at which shocks to the variance revert to their long-run values



Shocks to a GARCH Process

Volatility Term Structure

- Definition
  - ✓ The relationship between the volatilities of options and their maturities is referred to as the volatility term structure.
  - ✓ when the current volatility is above the long-term volatility, the GARCH(1,1) model estimates a downward-sloping volatility term structure, when below, an upward sloping volatility term structure

Example

The λ of an exponentially weighted moving average (EWMA) model is estimated to be 0.9. Daily standard deviation is estimated to be 1.5%, and today's stock market return is 0.8%. What is the new estimate of the standard deviation?

A. 1.68%.

B. 1.55%.

C. 1.45%.

D. 2.74%.

Answer: D

$\sigma_n^2 = 0.9 \times (0.015)^2 + (1 - 0.9) \times (0.008)^2 = 0.0002089$

Example

If the volatilities of two variables are estimated using a GARCH(1,1) model, which of the following models used to estimate covariance will generate a consistent correlation estimate between the two variables?

A. GARCH(1,1) model.

B. EWMA model.

C. Unweighted historical volatility model.

D. Geometrically weighted historical volatility model.

Example

Answer: A

To maintain consistency, covariance should be estimated using the same method as that used for the volatility estimation. Since the volatility of each variable was estimated using the GARCH(1,1) model, the covariance estimate should also use the same technique.

Summary

- **EWMA** $\sigma_n^2 = \lambda\sigma_{n-1}^2 \quad (1+\lambda)u_{n-1}^2$
- **ARCH** $\sigma_n^2 = \gamma V_L + \sum_{i=1}^{m} \alpha_i u_{n-i}^2$

  where

  $$\gamma + \sum_{i=1}^{m} \alpha_i = 1$$

- **GARCH**
  - ✓GARCH (1,1): $\sigma_n^2 = \omega + \alpha u_{n-1}^2 + \beta\sigma_{n-1}^2$
  - ✓Persistence: α+β
  - ✓Parameters: Maximum Likelihood Methods
  - ✓ $\boxed{E[\sigma_{n+t}^2] = V_L + (\alpha+\beta)^t(\sigma_n^2 - V_L)}$
- **Correlations and Covariances**

PART 2

1 . Probability Theory

2 . Statistics

3. Linear Regression

4. Time-Series Analysis

5. Simulation Method

CONTENTS

Learning objectives

- Define correlation and covariance and differentiate between correlation and dependence.
- **Calculate** covariance using the EWMA and GARCH(1,1) models.
- **Apply** the consistency condition to covariance.
- Describe the procedure of generating samples from a bivariate normal distribution.
- Describe **properties** of correlations between normally distributed variables when using a one-factor model.
- Define copula and describe the **key properties** of copulas and copula correlation.
- **Explain** tail dependence.
- Describe the Gaussian copula, Student’s t-copula, multivariate copula, and one-factor copula.

Covariance and Correlation

Covariance

$$Cov(X,Y)=E[X-E(X)][Y-E(Y)]=E(XY)-E(X)E(Y)$$

Correlation

$$\rho_{XY} = \frac{COV(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

✓A correlation of zero between two variables does not imply that there is no dependence between the two variables.

✓one variable can still have a nonlinear relationship with the other variable.

Calculate (☆☆☆)

Updating Covariances

- $cov_n = E(x_n y_n) - E(x_n)\,E(y_n) = E(x_n y_n)$
  - ✓ where x_n,y_n represents the daily return for different assets
  - ✓ Risk managers assume that expected daily return are zero.
- One alternative is to apply differential weightings to the past observations and create EWMA-type and GARCH type models to take into account the time variability of covariance
- Under EWMA: $cov_n = \lambda cov_{n-1} + (1-\lambda) x_{n-1} y_{n-1}$

Updating Covariances

- The main factor to consider when calculating covariance, however, is to maintain consistency in the calculation process.
- If one uses a particular weighting scheme (e.g., EWMA or GARCH) for variance calculations, one should use the same weighting scheme and procedure for the covariance calculations. Otherwise, the estimates are inconsistent with each other.

Apply (☆☆)

Consistency condition to covariance

- **A variance-covariance matrix** can be constructed using the calculated estimates of variance and covariance rates for a set of variables.
- A matrix is known as *positive-semidefinite if it is* ***internally consistent.***
- We would find that this variance-covariance matrix is not **internally consistent** since $\omega^T \Omega \omega \geq 0$ is not satisfied.
- The expression $\omega^T \Omega \omega \geq 0$ is the variance rate of a portfolio where an amount w is invested in market variable i.

Consistency condition to covariance

➢ **Example:** a variance-covariance matrix that is not internally consistent

$$\begin{pmatrix} 1 & 0 & 0.8 \\ 0 & 1 & 0.8 \\ 0.8 & 0.8 & 1 \end{pmatrix}$$

✓ there is no correlation between the first and second variables, but there is a strong correlation between the first and third variables as well as the second and third variables.

✓ Find $w_t = (1, 1, -1)$ that do not satisfies $\omega^T \Omega \omega \geq 0$

Generating Samples

- **Procedure of generating samples from a bivariate normal distribution**

✓Independent samples Z_1 and Z_2 from a univariate standardized normal distribution are obtained (both variables have mean zero and standard deviation one)

✓The required samples ε_1, and ε_2 are then calculated as follows:

$$\varepsilon_1 = z_1$$

$$\varepsilon_2 = \rho z_1 + z_2\sqrt{1-\rho^2}$$

Properties of correlations using one-factor model

One-factor model

✓Each U_i has a component dependent on one common factor F in addition to another component Z_i

$$U_i = a_i F + \sqrt{1 - a_i^2}\, Z_i$$

- U_i ~ N(0,1)
- a_i between -1 and 1
- F and Z_i ~ N(0,1), uncorrelated
- Every Z_i is uncorrelated with each other
- All correlations between U_i and U_j result from their dependence on a common factor, F

Properties of correlations using one-factor model

- **Advantages**
  - ✓ U_i has a mean of zero and a variance of one, and the correlation between U_i and U_j is $a_i a_j$
  - ✓ covariance matrix is always positive-semidefinite.
  - ✓ Without assuming a one-factor model, there are [N x (N - 1)] / 2 correlations to be calculated, one-factor model only estimate *N* parameters.
- **Example: CAPM model**
  - ✓ each asset return has a systematic component (beta) that is correlated with the market portfolio return.

Copula

- A **copula** creates a joint probability distribution between two or more variables while maintaining their individual marginal distributions. This is accomplished by mapping the marginal distributions to a new known distribution. It defines the correlation structure of the multivariate distribution.

$$F(\vec{x}) = C(F_1(x_1), \cdots, F_n(x_n))$$

注：$F_1(x_1)$、$F_2(X_2)$等为边际分布函数，F为联合分布函数

Copula

- **Marginal distribution** of a subset of a collection of random variables is the probability distribution of the variables contained in the subset.
- **Joint probability distribution** for *X, Y, …*is a probability distribution that gives the probability that each of *X*, *Y*, ... falls in any particular range or discrete set of values specified for that variable.
- The key property of a copula correlation model is the preservation of the original margjnal distributions while defining a correlation between them.

Copula

- Suppose we have two triangular marginal distributions for two variables *X and Y as illustrated in Figure 3.*

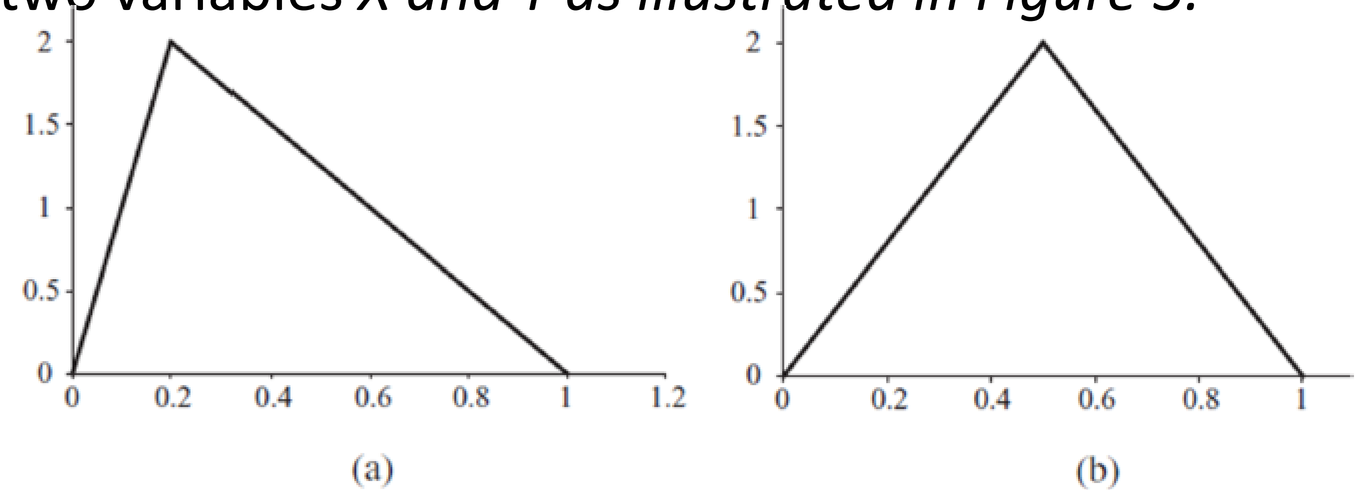


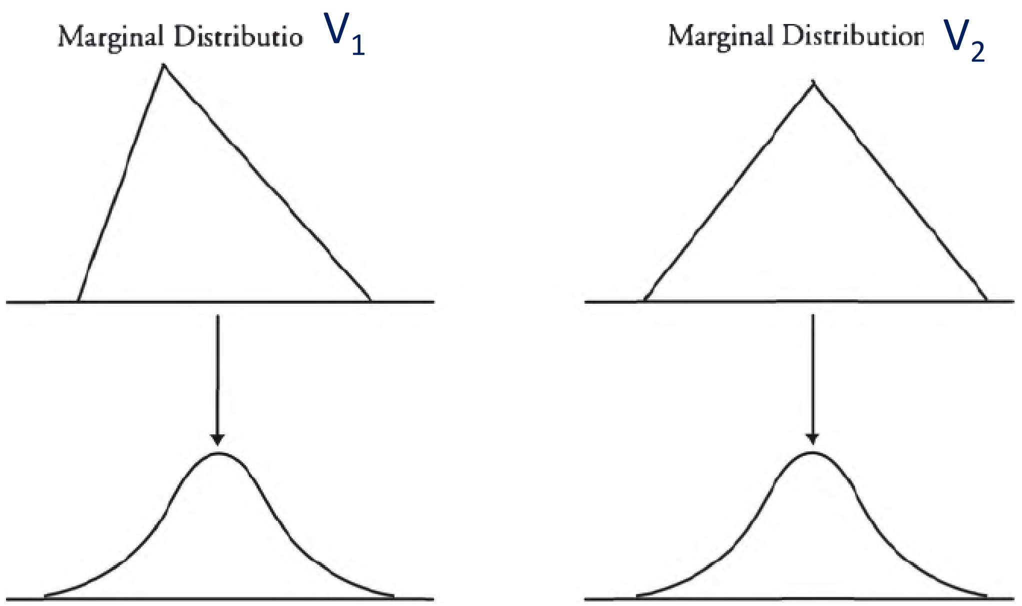
FIGURE 11.2 Triangular Distributions for V_1 and V_2

- These two triangular marginal distributions for X and Y are preserved by mapping them to a known joint distribution. Figure 4 illustrates how a **correlation copula is created.**

Describe (☆)

Copula

Figure 4: Mapping Variables to Standard Normal Distributions



- A **correlation copula** is created by converting two distributions and mapping them to known distributions with well-defined properties, such as the normal distribution.

Copula

V_1 Value	Percentile of Distribution	U_1 Value	V_2 Value	Percentile of Distribution	U_2 Value
0.1	5.00	-1.64	0.1	2.00	-2.05
0.2	20.00	-0.84	0.2	8.00	-1.41

- ✓ $P(U_1<-1.64, U_2<2.05) = P(V_1<-0.1, V_2<0.1) = 0.006$
- ✓ U_1 U_2 ~ bivariate normal distribution with correlation=0.5
- ✓ The correlation between U_1 and U_2 is referred to as the copula correlation

	V_2		
V_1	0.1	0.2	0.3
0.1	0.006	0.017	0.028
0.2	0.013	0.043	0.081

Describe (☆)

Types of Copulas

➤ Gaussian Copula

✓The copula that can be used to define a correlation structure between V_1 and V_2, which are both normal distributed variables.

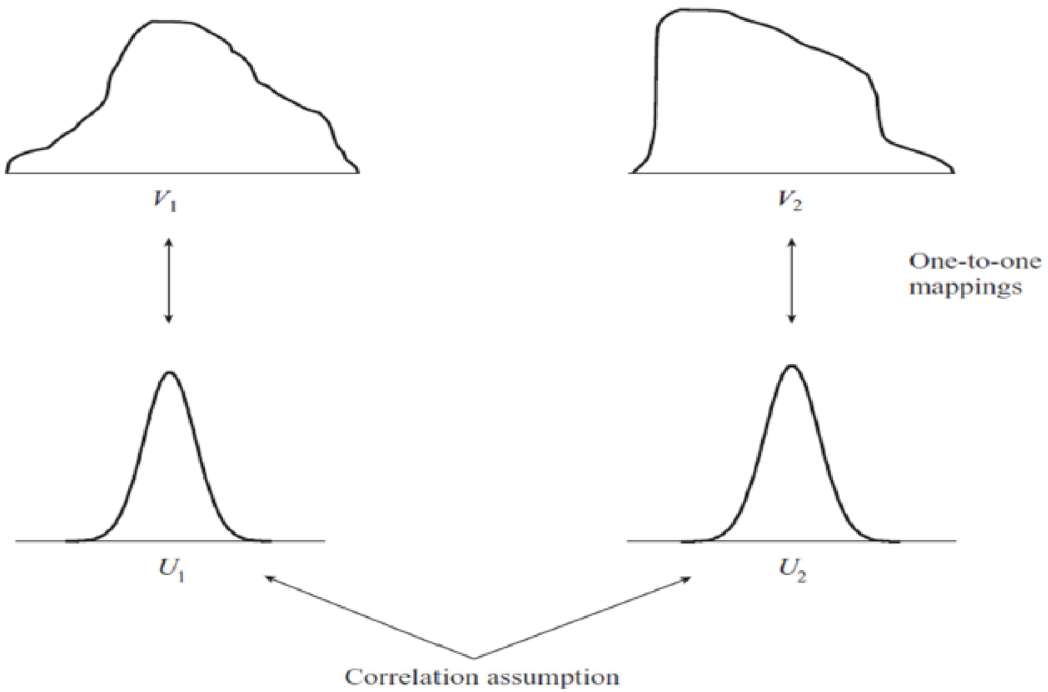


FIGURE 11.3 The Way in Which a Copula Model Defines a Joint Distribution

Other copulas

- **Student t-Copula**

  ✓The copula that can be used to define a correlation structure between U_1 and U_2, which are both following student t-distribution.

- **Multivariate Copula**

  ✓The copula that can be used to define a correlation structure between *more than two variables*.

Explain (☆☆)

Tail Dependence

- **Tail Dependence**
  - ✓it is more common for the two variables to have tail values at the same time in the bivariate Student t-distribution than in the bivariate normal distribution
  - ✓the *tail dependence* is higher in a bivariate Student t-distribution than in a bivariate normal distribution
  - ✓During a financial crisis, it is common for assets to be highly correlated and exhibit large losses at the same time.

Explain (☆☆)

高顿财经 GOLDEN FINANCE

Tail Dependence

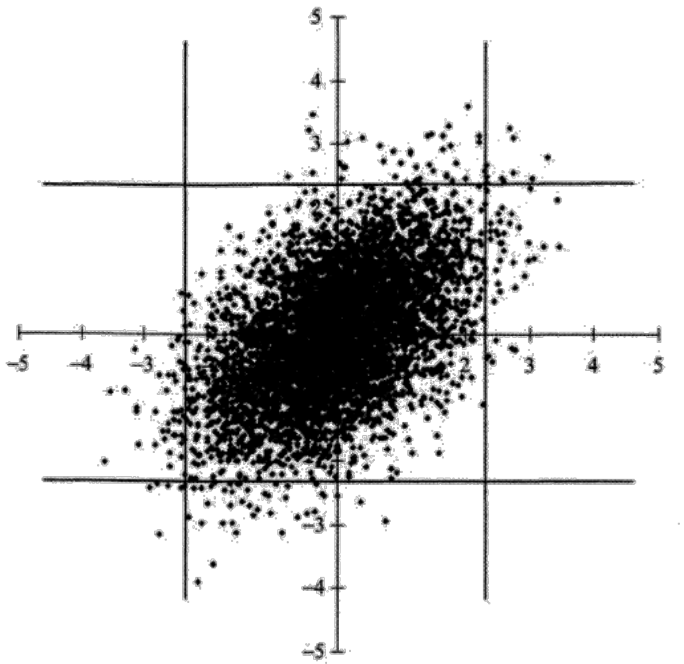


FIGURE 6-4 5,000 random samples from a bivariate normal distribution.

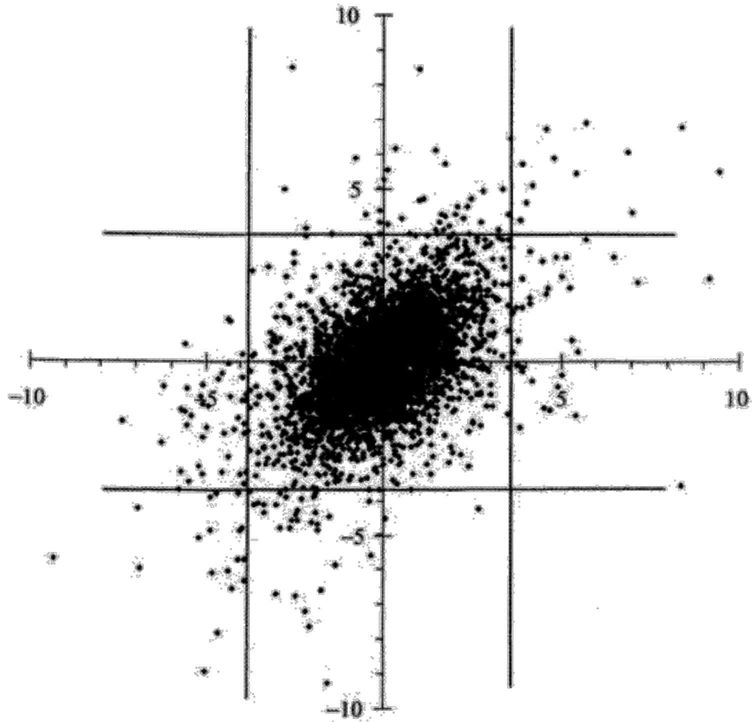


FIGURE 6-5 5,000 random samples from a bivariate Student *t*-distribution with four degrees of freedom.

Example

Suppose a risk manager wishes to create a correlation copula to estimate the risk of loan defaults during a financial crisis. Which type of copula will most accurately measure tail risk?

A. Gaussian copula.

B. Student’s r-copula.

C. Gaussian one-factor copula.

D. Standard normal copula.

Example

Answer: B

There is greater tail dependence in a bivariate Student's t-distribution than a bivariate normal distribution. This suggests that the Students t-copula is better than a Gaussian copula in describing the correlation structure of assets that historically have extreme outliers in tails at the same time.

Summary

- **Correlation**
- **Variance-covariance matrix**
  - ✓Internally consistent
- **One-factor model**

$$U_i = a_i F + \sqrt{1 - a_i^2} Z_i$$

- **Copula**
  - ✓Gaussian Copula, Student t-Copula, Multivariate

CONTENTS

PART 2

CONTENTS

PART 2

Learning objectives

- Describe the basic steps to conduct a Monte Carlo simulation.
- Describe ways to reduce Monte Carlo sampling error.
- **Explain** how to use **antithetic variate** technique to reduce Monte Carlo sampling error.
- **Explain** how to use **control variates** to reduce Monte Carlo sampling error and when it is effective.
- Describe the benefits of reusing sets of random number draws across Monte Carlo experiments and how to reuse them.
- Describe the bootstrapping method and its advantage over Monte Carlo simulation.
- Describe the pseudo-random number generation method and how a good simulation design alleviates the effects the choice of the seed has on the properties of the generated series.
- Describe situations where the bootstrapping method is ineffective.
- Describe disadvantages of the simulation approach to financial problem solving.

Monte Carlo Simulation

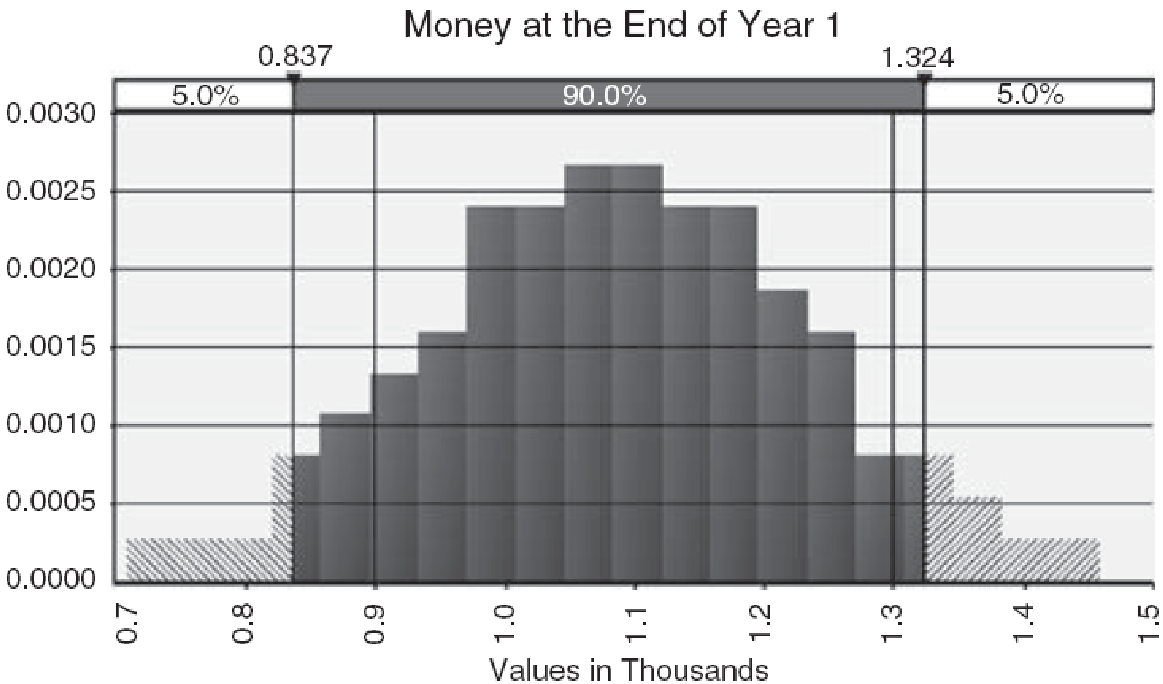
- Four basic steps required to conduct a Monte Carlo simulation.
  - ✓Step 1: Specify the data generating process (DGP).
  - ✓Step 2: Estimate an unknown variable or parameter
  - ✓Step 3: Save the estimate from step 2
  - ✓Step 4: Go back to step 1 and repeat this process N times.

Interpreting Monte Carlo Simulation Output

Example:

- C=1000(1+r)
  - ✓assume that the return on the market over the next year will follow a normal distribution
  - ✓Between 1977 and 2007, the S&P 500 returned 8.79% per annum on average, with a standard deviation of 14.65%
  - ✓draw 100 numbers from a normal distribution with mean 8.79% and standard deviation 14.65%

Interpreting Monte Carlo Simulation Output



Number of values	100
Mean	$1,087.90
Std Deviation	$ 146.15
Skewness	0.0034442
Kurtosis	2.871695
Mode	$1,118.96
5% Perc	$ 837.40
95% Perc	$1,324.00
Minimum	$ 708.81
Maximum	$1,458.52

- The minimum and the maximum are highly sensitive to the number of simulated values and whether the simulated values in the tails of the distribution provide good representation for the tails of the distribution

Monte Carlo Sampling Error

- The true expected value is computed as $s/\sqrt{N}$

  where s is the standard deviation of the output variables and N is the number of scenarios or replications in the simulation.

- Based on this equation, it intuitively follows that in order to reduce the standard error estimate by a factor of 10, the analyst must increase N by a factor of 100.

  Increasing N can be costly, so using variance reduction techniques

Variance reduction techniques

- **Antithetic variates**

✓cover different parts of the probability space

✓Take the complement of a set of random numbers and running a parallel simulation on those, covariance is negative

$$corr(u_t, -u_t) = \mathrm{cov}(u_t, -u_t) < 0$$

$$\bar{x} = (x_1 + x_2)/2$$

$$\mathrm{var}(\bar{x}) = \frac{1}{4}(\mathrm{var}(x_1) + \mathrm{var}(x_2) + 2\,\mathrm{cov}(x_1, x_2))$$

✓other method: stratified sampling, moment-matching, low-discrepancy sequencing

Variance reduction techniques

- **Control variates**

✓Employ a variable similar to that used in the simulation, but whose properties are known prior to the simulation

✓only if the control and simulation problems are very closely related. (x with unknown properties , y with known properties)

$$x^* = y + (\hat{x} - \hat{y})$$

Variance reduction techniques

- **Control variates**

$$\mathrm{var}(x^*) = \mathrm{var}(\widehat{x}) + \mathrm{var}(\hat{y}) - 2\mathrm{cov}(\widehat{x}, \widehat{y})$$

✓The control variate method will only reduce the sampling error in Monte Carlo simulations if

$$\mathrm{var}(x^*) < \mathrm{var}(\hat{x})$$

$$\mathrm{var}(\hat{y}) - 2\,\mathrm{cov}(\hat{x}, \hat{y}) < 0$$

Reusing Sets of Random Numbers

- Reusing sets of random number draws across Monte Carlo experiments reduces the estimate variability across experiments by using the same set of random numbers for each simulation.
- Accuracy of the actual estimates in each case will not be increased.
- Random number re-usage is unlikely to save computational time

Describe (☆☆)

Bootstrapping

- **Bootstrapping**
  - ✓Empirical estimators: Sampling from a given sample

$$y = y_1, y_2 \ldots y_T$$

  - ✓Re-estimate parameter $\hat{\theta}$ with new sample, get the distribution

- **Advantage**
  - ✓Without making strong distributional assumptions
  - ✓A method for detecting data snooping(e.g technical trading rule are bound, purely by chance alone)

Bootstrapping

➢ **Two situations when bootstrapping method is ineffective**

✓If outliers exist in the data, the inferences drawn from parameter estimates may not be accurate.

✓If autocorrelation exists in the original sample data, then the original historical data are not independent of one another.

Describe (☆)

Random Number Generation: Inverse Transform Method

- Inverse transform method: "inverse" of the cumulative distribution function
  - ✓A common method for converting a random number between 0 and 1 to a number from an arbitrary probability distribution

"Good" Random Number Generator

- "Truly random" number generation is time consuming and difficult
- Pseudorandom number generator: computed sequence which starts with a number (seed)
- Random number generators modulus （余数）
  - ✓$y_{i+1} = ay_i + c \quad modulo\ \ m,\ \ i = 0,1,\ldots,T$
  - ✓$R_{i+1} = y_{i+1}/m$ for i=0,1,…, T

Y_0 is the seed(initial value of y), a is a multiplier and c is an increment; modulo operator m simply functions as a clock , returning to one after reaching m.

Describe (☆☆)

Disadvantages of the simulation approach

- **For econometric or financial problem solving**
  - ✓computationally expensive
  - ✓results might not be precise
  - ✓results are often hard to replicate
  - ✓experiment-specific

Summary

- **Basic steps: Monte Carlo simulation**
- **Reduce sampling error**
  - ✓Antithetic variate
  - ✓Control variate
  - ✓Reusing random variable
- **Bootstrapping method**
- **Pseudo-random number generation method**
- **Disadvantages of the simulation**