



CredGuild

STUDY NOTES

CFA® Program Level I

Quantitative Methods

2027 curriculum · 11 readings · 48 practice questions

credguild.com

Version v2027.1 · October 2026

ABOUT THIS EDITION

CredGuild · CFA® Program Level I · Quantitative Methods
2027 curriculum · Version v2027.1 · Generated October 4, 2026

SHARING

© 2026 CredGuild. You may share this file freely, complete and unchanged. Do not sell it, and do not republish it with the CredGuild name, links or watermark removed.

USING THIS BOOK

These notes are a study aid. Use them alongside the CFA Institute curriculum. The latest version is free at credguild.com.

FOUND A MISTAKE?

Email hello@credguild.com, or use "Report an issue with this question" under any question on credguild.com.

TRADEMARKS

CFA Institute does not endorse, promote, review, or warrant the accuracy or quality of the products or services offered by CredGuild. CFA®, Chartered Financial Analyst® and CFA Institute® are trademarks owned by CFA Institute.

Contents

How to use this book	1
Quantitative Methods	2
1. Rate of Return	3
2. Types of Financial Returns	15
3. Benchmarking Returns	24
4. Time Value of Money	35
5. Skewness and Kurtosis	49
6. Lognormal Distribution	65
7. Null Hypothesis	83
8. Portfolio Standard Deviation	103
9. Simulation of Financial Asset Prices and Returns	128
10. Least Squares Regression	139
11. Machine Learning in Finance	156
Keep going at credguild.com	163

How to use this book

These are the CredGuild notes for the 2027 CFA® Program Level I curriculum, Quantitative Methods: 11 readings. They are the same notes as on credguild.com, set out for reading on paper or offline.

What each reading contains

- **What you'll learn:** the learning outcome statements the reading covers.
- **The notes, module by module:** worked examples, key concepts, common exam traps, exam shortcuts and a bottom line.
- **A quick check after each module:** one question with its answer and the full explanation.
- **Practice questions at the end, with the answers.** The full explanation for each one is free on the reading's page at credguild.com; the link is under the answers.

The coloured boxes

KEY CONCEPT

Definitions and formulas to know.

Example. A worked problem, step by step. Try it before you read the working.

COMMON EXAM TRAPS

Mistakes that cost marks.

EXAM SHORTCUTS

Faster ways through a question.

BOTTOM LINE

What to remember from the module.

Using it with the website

Paper is good for reading. Practice works better on the site: it scores each answer as you go, explains every option, keeps a list of the questions you got wrong and spreads the readings across the weeks before your exam. If you are not sure where to start, the free diagnostic test (credguild.com/cfa/level-1/diagnostic/) gives you a score in each topic.

If this book helps you, pass it on to someone else who is studying.

CFA® PROGRAM LEVEL I · TOPIC

Quantitative Methods

Exam weight 11–14% · 11 readings

How to measure and compare returns, discount cash flows, describe return data with statistics and probability distributions, test hypotheses and fit a simple regression. It ends with simulation and an introduction to financial data science.

Reading 1	Rate of Return
Reading 2	Types of Financial Returns
Reading 3	Benchmarking Returns
Reading 4	Time Value of Money
Reading 5	Skewness and Kurtosis
Reading 6	Lognormal Distribution
Reading 7	Null Hypothesis
Reading 8	Portfolio Standard Deviation
Reading 9	Simulation of Financial Asset Prices and Returns
Reading 10	Least Squares Regression
Reading 11	Machine Learning in Finance

QUANTITATIVE METHODS · READING 1

Rate of Return

What you'll learn

LOS 1.a Describe, compare and interpret returns: price, capital distribution and total return; ex ante vs. ex post and realized vs. unrealized returns; arithmetic and geometric means; annualized and continuously compounded returns.

LOS 1.b Describe, compare and interpret required returns, real and nominal risk-free rates, risk premia and inflation, and the real, excess, gross/net, after-tax and leveraged measures of return.

MODULE 1.1

Rates of Return

This reading measures returns over a single period and over several periods: the holding period return, the arithmetic and geometric means, annualized returns and continuously compounded returns. It then breaks a required return into the real risk-free rate, expected inflation and risk premia, and computes real, excess, gross, net, after-tax and leveraged returns.

LOS 1.a — Measuring and comparing returns

1. Assets, instruments and indicators

Returns can be measured for three overlapping categories of items:

Category	What it is	Typical examples	Generates cash flows itself?
Financial assets	Things of value that an investor can own	Cash, equity stakes, loans/debt (an asset for the lender), hybrid securities	Usually (dividends, interest, principal)
Financial instruments	Standardized, tradable vehicles through which financial assets change hands in markets	Equity and debt securities (shares, bonds), exchange-traded futures and other standardized contracts	Securities may pay dividends or coupons; standardized contracts need not pay anything
Financial indicators	Observable measures of value that pay nothing to anyone	Interest rates, exchange rates, market index levels	No; they are used to value other items

A machine, a building or a truck is a *real* (physical) asset, so it is neither a financial instrument nor a financial indicator.

2. Return over a single period

An asset's return has two sources: the change in its price (the **price return**, also called the **capital appreciation return**) and any cash it pays out. The **total return** for a period adds the two.

$$\text{Price return} = \frac{P_1 - P_0}{P_0} \quad \text{Capital distribution return} = \frac{D}{P_0}$$

KEY CONCEPT

$$\text{Total (holding period) return} = \frac{P_1 - P_0 + D}{P_0} = \text{price return} + \text{capital distribution return}$$

The **capital distribution return** measures a period's cash income relative to the beginning price. For a share it is the **dividend yield**; for a bond it is the **current yield**. Every component is divided by the *beginning* price P_0 . D is the total cash actually received between purchase and sale (add up every coupon or dividend paid in the window, and nothing paid outside it).

Example. A bond is bought at 97.00, pays coupons of 1.50 and 1.50 over the year and is sold at 98.20. Price return = $1.20/97 = 1.24\%$; distribution return = $3.00/97 = 3.09\%$; total return = $4.20/97 = 4.33\%$.

A **holding period return** describes the whole holding period, however long it is. The length of the period matters only when the result has to be annualized.

Other labels for returns

Pair	Meaning
Actual return (ex post return) vs. expected return (ex ante return)	An ex post return uses prices and income already observed; an ex ante return is an estimate for a future period. Decisions rest on expected returns, and the gap between expected and actual returns represents risk.
Realized returns vs. unrealized return	Realized = cash already in hand (income received, sale proceeds). Unrealized = paper gain or loss on a position still held, valued at today's market price. Taxes are usually levied on income when received and on price gains only when realized by a sale.
Real vs. realized	<i>Real</i> means adjusted for inflation (LOS 1.b). Do not confuse it with <i>realized</i> .

3. Averaging returns over several periods**KEY CONCEPT**

$$\bar{R}_{\text{arith}} = \frac{R_1 + R_2 + \dots + R_n}{n} \quad R_{\text{geo}} = [(1 + R_1)(1 + R_2) \dots (1 + R_n)]^{1/n} - 1$$

- The **geometric mean return (compound average return)** is the constant per-period rate that turns the starting value into the ending value: $(1 + R_{\text{geo}})^n$ equals 1 plus the holding period return for all n periods together. It is the measure to use for an annualized return or a **compound annual growth rate**.
- The **arithmetic mean return** adds the periodic returns and divides by their number. It ignores compounding; it implicitly assumes the amount invested stays at its starting level, although gains, losses and cash flows change it. For this reason it can overstate what the investor actually earned.

- Under the **arithmetic-geometric mean inequality principle**, $R_{\text{geo}} \leq \bar{R}_{\text{arith}}$. The gap widens as the returns become more dispersed, and the two are equal only when every period's return is identical.
- Each period enters the product as a **wealth relative**, $1 + R$. A loss keeps its sign: a -6% year enters as 0.94.

Example. Returns of +25% and -20%: arithmetic mean = 2.5%, but $\sqrt{1.25 \times 0.80} - 1 = 0\%$. €1,000 grows to €1,250 and falls back to €1,000. Wealth did not change, and only the geometric mean shows this.

The root used in a geometric mean equals the count of periods. If the periods are quarters, the result is an average *quarterly* compound return. To get an annual rate from several sub-annual returns, take the root equal to the number of *years* covered. For eight quarterly returns, the annual rate is $[\prod(1 + R_q)]^{1/2} - 1$. Calculator (TI BA II Plus) for a cube root: 1.1 [y^x] 3 [1/x] [=].

4. Annualizing a holding period return

Interest rates and market returns are normally quoted as **annualized returns**, whatever the actual length of the period. Annualizing makes investments with different horizons comparable. For a period shorter than a year, it assumes that the same return is earned again in every remaining period of the year, which actual returns rarely do.

$$\text{Annualized return} = (1 + \text{HPR})^{365/\text{days}} - 1 \quad \text{or} \quad (1 + R_{\text{period}})^c - 1$$

where c counts how many such periods make up a year (12 for monthly, 4 for quarterly, 2 for semiannual). The direction of the adjustment depends on the sign of the HPR:

KEY CONCEPT

Holding period	Positive HPR	Negative HPR (between -100% and 0)
Longer than a year (exponent < 1)	Annualized return is smaller than the HPR	Annualized return is less negative (closer to zero) than the HPR
Shorter than a year (exponent > 1)	Annualized return is larger than the HPR	Annualized return is more negative than the HPR

A zero HPR annualizes to zero. Rule of thumb: annualizing a long period pulls the return toward zero; annualizing a short period pushes it away from zero.

Example. 2% earned over 146 days: $1.02^{365/146} - 1 = 1.02^{2.5} - 1 = 5.08\%$. A 0.6% monthly return: $1.006^{12} - 1 = 7.44\%$ (not $12 \times 0.6\% = 7.2\%$). A loss of 15.36% over exactly two years: $0.8464^{1/2} - 1 = -8.00\%$ per year, which is *greater* (less negative) than -15.36%.

5. Compounding frequency and continuous compounding

For a given positive stated (quoted) annual rate, more frequent compounding means a higher **effective annual rate**, a higher future value and a lower present value. The limiting case is **continuous compounding**, whose effective annual rate, $e^r - 1$, is higher than that of any discrete frequency.

With m compounding periods a year, the rate for each period is the stated annual rate divided by m . Compounding that rate m times gives

$$\text{Effective annual rate} = \left(1 + \frac{r_s}{m}\right)^m - 1 \quad PV = \frac{FV}{(1 + r_s/m)^m} \text{ (one year ahead)}$$

The first formula is the annualizing formula of section 4 with $c = m$.

Example. The present value of 2,000 due in one year at a stated 4% is 1,923.08 with annual compounding, 1,921.96 with quarterly ($2,000/1.01^4$, an effective annual rate of $1.01^4 - 1 = 4.06\%$), 1,921.71 with monthly, 1,921.58 with daily, and $2,000e^{-0.04} = 1,921.58$ with continuous compounding (effective annual rate $e^{0.04} - 1 = 4.08\%$).

To compare a continuously compounded rate with a discrete one, convert both to effective annual rates. Going the other way, the continuously compounded stated rate that gives a chosen effective annual rate is $r_{cc} = \ln(1 + \text{EAR})$. *Example.* A deposit quoting 5% compounded quarterly has an effective annual rate of $1.0125^4 - 1 = 5.09\%$. A continuously compounded deposit matches it at a stated rate of $\ln(1.0509) = 4.97\%$.

Moving from a holding period return to a **continuously compounded return** (also called a **logarithmic return**, or log return) uses the natural logarithm:

$$r_{cc} = \ln(1 + \text{HPR}) \quad \text{and back:} \quad \text{HPR} = e^{r_{cc}} - 1$$

Because the HPR is a total return, the log return built from it includes income. With a single distribution D received at the end of the period, $r_{cc} = \ln\left(\frac{P_1 + D}{P_0}\right)$. Only when no distributions are paid does this simplify to $r_{cc} = \ln(P_1/P_0)$, which on its own is a *price* log return. *Example.* Bought at 100, sold at 100, with a 5 dividend: the price log return is $\ln(100/100) = 0$, but the total log return is $\ln(105/100) = 4.88\%$.

- For a gain, $r_{cc} < \text{HPR}$; for a loss, r_{cc} is *more* negative than the HPR (e.g., a fall from 100 to 90: $\text{HPR} = -10\%$, $\ln 0.9 = -10.54\%$).
- Log returns are additive across periods. Wealth relatives multiply, $P_2/P_0 = (P_1/P_0)(P_2/P_1)$, and the logarithm of a product is the sum of the logarithms. So with no distributions, $\ln(P_2/P_0) = \ln(P_1/P_0) + \ln(P_2/P_1)$ (in general, the sub-period values $\ln(1 + \text{HPR}_t)$ add up). Keep the sign of each sub-period (a fall contributes a negative term).

Example. A price moves from 50 to 60 and then to 54: $\ln(1.2) = 18.23\%$, $\ln(0.9) = -10.54\%$, sum = $7.70\% = \ln(54/50)$.

COMMON EXAM TRAPS

- Dividing by the *ending* price instead of the beginning price. In the bond example, dividing the 4.20 total gain by the selling price of 98.20 gives 4.28% instead of 4.33%.
- Leaving out (or double counting) a dividend or coupon; check which payments fall inside the holding period.
- Annualizing when the question asks for the holding period return (or vice versa).
- Using simple multiplication ($c \times R$) instead of compounding $(1 + R)^c - 1$; using 360 days when the question says 365; inverting the exponent (days/365). For the 2% earned over 146 days, inverting the exponent gives $1.02^{146/365} - 1 = 0.80\%$ instead of 5.08%.
- Reporting the arithmetic mean when the compound (annualized) rate is asked for, forgetting to take the root, or dropping the minus sign on a loss.

- Using $e^{HPR} - 1$ instead of $\ln(1 + HPR)$ for a continuously compounded return, or dropping a distribution by using $\ln(P_1/P_0)$ when income was paid. For the fall from 100 to 90, $e^{-0.10} - 1 = -9.52\%$ instead of $\ln 0.9 = -10.54\%$.
- Assuming annualization always shrinks a multi-year return. It does so only for a gain.

EXAM SHORTCUTS

- The direction of an annualized return is known before calculating: a positive HPR over more than a year annualizes to a smaller figure and over less than a year to a larger one, while a negative HPR (between -100% and 0) becomes less negative over more than a year and more negative over less than a year.
- The geometric mean return cannot exceed the arithmetic mean return of the same returns, so an answer with a higher geometric mean can be ruled out.
- A log return is below the HPR for a gain and more negative than the HPR for a loss, which rules out answers on the wrong side of the HPR.

BOTTOM LINE

- Total (holding period) return is $(P_1 - P_0 + D)/P_0$, the price return plus the capital distribution return, with each component divided by the beginning price P_0 .
- An ex post (actual) return uses prices and income already observed, while an ex ante (expected) return is an estimate for a future period.
- The geometric mean return $[(1 + R_1)(1 + R_2) \cdots (1 + R_n)]^{1/n} - 1$ is the measure for an annualized return or a compound annual growth rate; it cannot exceed the arithmetic mean, and the two are equal only when every period's return is identical.
- A holding period return is annualized with $(1 + HPR)^{365/\text{days}} - 1$, or $(1 + R_{\text{period}})^c - 1$ with c periods in a year; for a period shorter than a year this assumes the same return is earned in every remaining period.
- For a given positive stated annual rate, more frequent compounding gives a higher effective annual rate, $(1 + r_s/m)^m - 1$, and continuous compounding gives the highest, $e^r - 1$.
- The continuously compounded return is $r_{cc} = \ln(1 + HPR)$, which reduces to $\ln(P_1/P_0)$ only when no distributions are paid, and log returns are additive across periods.

QUICK CHECK

Question 1

A macro strategist's dashboard displays the central bank policy interest rate, the USD/JPY exchange rate and the level of the Euro Stoxx 50 index. These three items are best classified as:

- A financial assets.
- B financial instruments.
- C financial indicators.

Answer: C

Interest rates, exchange rates and market index levels are financial indicators: observable measures of value that do not produce any cash flows on their own. They matter because they are used to value other things, such as the securities and contracts that are financial instruments.

Why the other options are wrong

- **A.** Financial assets are items an investor can own that carry value, such as cash, shares or loans. Nobody owns an interest rate or an index level, and these items pay nothing to anyone.
- **B.** Financial instruments are standardized vehicles for trading financial assets in markets, such as listed shares, bonds or standardized derivative contracts. An index *level* or an exchange *rate* is a measurement and cannot itself be traded.

Key takeaway An indicator is an observed measure (rates, FX, index levels); an instrument is a standardized, tradable vehicle; an asset is something an investor owns that has value.

MODULE 1.2

Components and Measures of Return

LOS 1.b — Required returns, risk-free rates, risk premia, inflation and other return measures

1. Three ways to read an interest rate

An interest rate expresses the time value of money; differences in risk across securities explain why their equilibrium rates differ. An interest rate can be seen as:

- a **required rate of return**: the minimum return that savers must expect before they will willingly lend or invest their money (the market's equilibrium rate for that investment);
- a **discount rate**: the rate used to translate future cash flows into today's value (someone who can borrow at 7% should discount future payments at 7%);
- an **opportunity cost of current consumption**: the return given up by spending today instead of saving.

The terms interest rate and discount rate are often used interchangeably. A **risk premium** may be part of an interest rate, but "a measure of risk" is not one of these interpretations. When a question describes the *minimum rate needed to persuade investors to commit funds*, the best label is the required rate of return.

2. Building blocks of a required return

KEY CONCEPT

$$\text{Required return} = \underbrace{\text{real risk-free rate} + \text{expected inflation}}_{\text{nominal risk-free rate}} + \text{risk premium(s)}$$

Rate	Contains expected inflation?	Contains default/liquidity/other risk premia?
Real risk-free rate	No	No; there is no uncertainty about the cash flows

Rate	Contains expected inflation?	Contains default/liquidity/other risk premia?
Nominal risk-free rate (e.g., a short-term government bill)	Yes	No
Required return on a risky asset	Yes	Yes

- The **real risk-free rate** reflects only **time preference**, that is, how strongly people prefer consuming now over consuming the same amount later.
- A short-term government bill is (practically) free of default risk, but its yield carries an **inflation premium** because expected inflation is not zero, so its yield is a nominal risk-free rate rather than a real one.
- Exact link: $(1 + r_{\text{nominal}}) = (1 + r_{\text{real}})(1 + \pi^e)$. Common approximation: $r_{\text{real}} \approx r_{\text{nominal}} - \pi^e$.
- The real rate therefore moves with the nominal risk-free rate and against expected inflation. If the bill yield falls while expected inflation rises, the real rate falls.

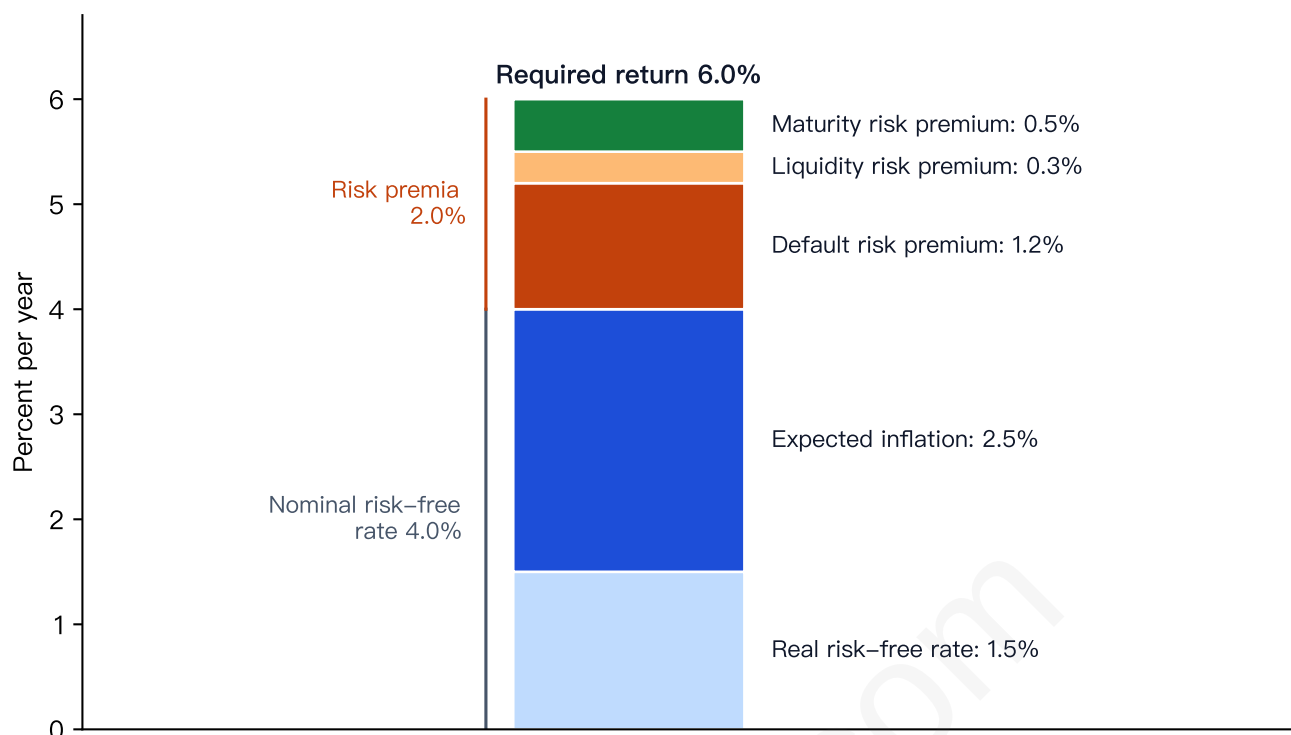
Typical risk premia

Each additional type of risk a security carries adds its own premium and raises the required return. The lists below are examples, not complete lists.

Bonds	Equities
Default risk premium: the borrower may not pay on time or in full	Size risk premium: return gap between small-cap and large-cap stocks
Liquidity risk premium: a quick sale for cash may fetch less than fair value	Value risk premium: return gap between high book-to-market (value) and low book-to-market (growth) stocks
Maturity risk premium: longer bonds' prices react more to rate changes	

The **market risk premium**, used in later readings, is the return on the market portfolio (every risky asset) minus the return on a risk-free asset.

Example. Real risk-free 1.5%, expected inflation 2.5%, default risk premium 1.2%, liquidity risk premium 0.3%, maturity risk premium 0.5%: required yield $\approx 1.5 + 2.5 + 1.2 + 0.3 + 0.5 = 6.0\%$. Equivalently, nominal risk-free 4.0% + 2.0% of premia = 6.0%.



Building a required return: nominal risk-free rate plus risk premia (example numbers)

3. Other return measures

Real return

The **real return** measures the growth in purchasing power:

$$1 + r_{\text{real}} = \frac{1 + r_{\text{nominal}}}{1 + \pi} \quad \text{so} \quad \pi = \frac{1 + r_{\text{nominal}}}{1 + r_{\text{real}}} - 1$$

With positive inflation the real return is below the **nominal return**; with deflation ($\pi < 0$) it is above it.

Example. A portfolio earns a nominal 6.0% while inflation is 2.5%. The approximate real return is $6.0\% - 2.5\% = 3.5\%$; the exact real return is $1.060/1.025 - 1 = 3.41\%$. In purchasing-power terms, €1,000 grows to €1,060 while a basket of goods rises in price from €1.00 to €1.025, so the investor can buy $1,060/1.025 = 1,034.1$ baskets instead of 1,000, that is, 3.41% more.

Excess return

The **excess return** is the return over a benchmark or reference rate: approximately $r - r_b$, exactly $(1 + r)/(1 + r_b) - 1$. For a fund return of 11% against a benchmark return of 7%, the excess return is about 4% by subtraction and $1.11/1.07 - 1 = 3.74\%$ exactly.

Gross vs. net return

KEY CONCEPT

Measure	Deduct trading commissions and other costs of generating the return?	Deduct management and administration fees?
Total return	No	No
Gross return	Yes	No
Net return	Yes	Yes

Pretax vs. after-tax nominal return

The **pretax nominal return** is earned before taxes; the **after-tax nominal return** is what remains after the tax liability. The rates that apply can depend on the tax laws of the country where the investment is held and on those of the investor's home country. Dividends, interest, short-term gains and long-term gains may each be taxed at a different rate, and many countries tax long-term gains more lightly than short-term gains. To get the after-tax nominal return with one tax rate on capital gains (t_{cg}) and another on distributions such as dividends and interest (t_d):

$$r_{\text{after-tax}} = \text{price return} \times (1 - t_{cg}) + \text{distribution return} \times (1 - t_d)$$

Example. Buy at 50, receive 1.50 of dividends, sell at 54. Price return 8%, distribution return 3%, pretax 11%. With $t_{cg} = 15\%$ and $t_d = 30\%$: $8\% \times 0.85 + 3\% \times 0.70 = 6.80\% + 2.10\% = 8.90\%$. If the whole return is price gain, only the capital gains rate applies. In practice, capital gains are usually taxed only when realized, and capital losses can often be used to offset gains.

Leveraged return

Leverage means using borrowed money or derivatives to control more of an asset than the investor's own cash would buy. The resulting **leveraged return** is the gain or loss as a percentage of the investor's cash investment. Suppose an investor puts up equity V_E , borrows V_D at rate r_D and earns the **unleveraged return** r_P on the whole amount:

KEY CONCEPT

$$r_L = \frac{r_P(V_E + V_D) - r_D V_D}{V_E} = r_P + \frac{V_D}{V_E}(r_P - r_D)$$

Example. $V_E = 600$, $V_D = 200$, $r_D = 5\%$. If $r_P = 9\%$: $(72 - 10)/600 = 10.33\%$. If $r_P = -3\%$: $(-24 - 10)/600 = -5.67\%$. Both results lie further from zero than the unleveraged 9% and -3%. The second form of the formula shows why: leverage adds $(V_D/V_E)(r_P - r_D)$ to the unleveraged return. A loss is always magnified, because the interest cost adds to it; a gain is magnified only if r_P exceeds r_D .

Derivatives such as futures and options also give leveraged returns: the cash deposited (margin or premium) is only a small fraction of the value of the underlying exposure.

COMMON EXAM TRAPS

- Adding expected inflation to a *nominal* risk-free rate (double counting), or leaving it out when starting from a *real* rate. In the required-yield example, adding the 2.5% expected inflation to the 4.0% nominal risk-free rate gives 8.5% instead of 6.0%.
- Calling a government bill yield a *real* rate. It is a nominal rate.
- Adding a maturity risk premium for a short-term instrument. Commercial paper or another short-term instrument maturing with the government bill carries no extra maturity risk, so only its default (and any liquidity) premium is added.
- Mixing up the size risk premium (small vs. large cap) and the value risk premium (high vs. low book-to-market), or building an equity premium into a bond's required yield.
- Deducting transaction costs again when moving from gross to net return.
- Swapping the two tax rates, or applying one blended rate to the total return. In the tax example, swapping them gives $8\% \times 0.70 + 3\% \times 0.85 = 8.15\%$ instead of 8.90%.
- In the leverage formula: charging interest on the whole portfolio instead of only on V_D , adding the interest instead of subtracting it, or dividing by V_D or $V_E + V_D$ instead of V_E . In the example with $r_P = 9\%$, charging 5% interest on all 800 gives 5.33% and dividing 62 by 800 gives 7.75%, instead of 10.33%.

EXAM SHORTCUTS

- The sign of $r_P - r_D$ shows whether borrowing raises or lowers a gain before any calculation: a gain is magnified only if r_P exceeds r_D , and a loss is always magnified.
- The real rate moves with the nominal risk-free rate and against expected inflation, so a falling bill yield with rising expected inflation means a lower real rate without computing it.

BOTTOM LINE

- An interest rate can be read as a required rate of return, a discount rate or an opportunity cost of current consumption; a measure of risk is not one of these interpretations.
- Required return = real risk-free rate + expected inflation + risk premia, where the real risk-free rate plus expected inflation is the nominal risk-free rate, such as a short-term government bill yield.
- The real risk-free rate reflects only time preference and is linked to the nominal rate exactly by $(1 + r_{\text{nominal}}) = (1 + r_{\text{real}})(1 + \pi^e)$, or approximately by $r_{\text{real}} \approx r_{\text{nominal}} - \pi^e$.
- Gross return deducts trading commissions and other costs of generating the return but not management and administration fees; net return deducts both.
- With one tax rate on capital gains and another on distributions, the after-tax nominal return is price return $\times (1 - t_{cg})$ + distribution return $\times (1 - t_d)$.
- The leveraged return is $r_L = r_P + (V_D/V_E)(r_P - r_D)$, a gain or loss measured against the investor's own cash V_E .

QUICK CHECK**Question 2**

Which statement about the real risk-free interest rate is *most accurate*? It:

- A** includes a premium for expected inflation but no premium for default risk.
- B** excludes both expected inflation and any uncertainty about the amount or timing of cash flows.
- C** includes a liquidity risk premium because instruments paying this rate may be hard to sell quickly.

Answer: B

The label "risk-free" rules out any doubt over the amount or timing of the cash flows (so no default risk), while "real" rules out any allowance for expected inflation. The real risk-free rate therefore contains neither, reflecting only time preference.

Why the other options are wrong

- **A.** A rate that includes an inflation premium but has zero probability of default is a *nominal* risk-free rate, not a real one.
- **C.** A liquidity risk premium is compensation for a type of risk. A risk-free rate contains no risk premiums of any kind, including a liquidity risk premium.

Key takeaway Real risk-free rate: no inflation, no risk premia. Adding expected inflation gives the nominal risk-free rate; adding risk premia to that gives the required return on a risky asset.

Practice questions**Question 3**

Olivia Brandt put €50,000 into a sector fund that rose 20% in year one and then dropped 20% in year two. To express the true average yearly change in her wealth, she should use the:

- A** arithmetic mean return.
- B** geometric mean return.
- C** capital distribution return.

Question 4

A company must pay a supplier €250,000 in exactly one year. Its treasurer discounts this payment at a stated annual rate of 6%. Under which compounding convention will the present value of the payment be *lowest*?

- A** Annual compounding.
- B** Monthly compounding.
- C** Semiannual compounding.

Question 5

The managers of the Arcadia Opportunities Fund are considering borrowing to raise the fund's investment exposure from €100 million of investor capital to €150 million. Compared with remaining unleveraged, this borrowing will *most likely*:

- A magnify gains but limit losses.
- B diminish gains and magnify losses.
- C magnify both gains and losses.

Answers: 3. B · 4. B · 5. C

Why each answer is right and why each other option is wrong: credguild.com/cfa/level-1/notes/quantitative-methods/rate-of-return/. The full bank has 74 questions on this reading.

credguild.com

QUANTITATIVE METHODS · READING 2

Types of Financial Returns

What you'll learn

LOS 2.a Calculate, compare and interpret the types of return (price, capital distribution, total, domestic-currency) earned on financial assets, instruments and indicators, and relate them to risk and asset allocation.

MODULE 2.1

Financial Assets, Instruments, and Indicators

This reading splits investment returns into price return and capital distribution return and shows which source dominates for equities, bonds, derivatives, cash and hybrid assets. It then computes the return on a T-bill, the domestic return on a foreign asset and index returns, and closes with the long-run risk-return ranking of asset classes, systematic and unsystematic risk, asset allocation and the weighted-average expected return of a portfolio.

LOS 2.a — Return components of financial assets, instruments and indicators

1. Two sources of return

Every investment return comes from one or both of two sources:

- **Price return** (also called **capital appreciation return**) is the gain or loss from a change in the market value of the investment.
- **Capital distribution return** is the income the investment pays out during the period, such as **dividends** or **coupons**. A distribution can be *fixed* (the rate is set at the outset, as with most bond coupons) or *variable* (the amount is uncertain, e.g. discretionary dividends or a coupon linked to a market reference rate).

KEY CONCEPT

$$\text{Total return} = \text{capital appreciation return} + \text{capital distribution return}$$

Capital distributions are **realized** when received. Price gains or losses are realized only if the asset is sold. If the asset is still held at the end of the period, they are **unrealized**, but they still count in total return.

2. Main types of financial instruments

KEY CONCEPT

Instrument	What it is	Dominant source of return
Equities (stocks, shares)	Ownership stake in the issuer	Price return , driven by factors that affect the company's earnings (company-specific news and broad economic factors). Dividends are variable and discretionary.

Instrument	What it is	Dominant source of return
Bonds (fixed income, debt securities)	Tradable loan; holder receives coupons and the principal (par value, face value) at maturity	Fixed capital distribution (the coupon); price return is usually the smaller part.
Derivatives	Contracts whose value derives from the performance of a specified reference (underlying) asset	Performance of the underlying asset (e.g., a contract fixing a future price for a commodity gains for the buyer if the commodity price rises).
Cash (notes, coins, non-interest-bearing deposits)	Financial asset held for liquidity	No return; its purchasing power falls as prices rise.
Hybrid assets (e.g., convertible bonds , preferred stock)	Blend of debt and equity features	A mix of debt-like and equity-like returns, e.g. a fixed coupon plus the right to convert into common shares.

3. Financial indicators

Financial indicators, such as **interest rates**, **foreign exchange rates** and **market indexes**, influence returns without producing any cash flows of their own.

Interest rates and the risk-free rate

Interest rates are paid to savers who lend funds. They are seen most directly in the rates banks quote on interest-bearing deposits and, in capital markets, in the returns offered by securities that borrowers issue in the fixed-income market.

The yield on short-term government securities of a developed economy is treated as a **risk-free rate**, because the issuer is assumed to make the promised payments (no **credit risk**). A **Treasury bill** (T-bill) is issued at a discount and pays its face value at maturity, so its holding period return and annualized rate (365-day year, with compounding) are:

$$HPR = \frac{\text{Face value} - \text{Price}}{\text{Price}}, \quad r_{\text{annual}} = (1 + HPR)^{365/t} - 1$$

Example. A 91-day bill with face value \$2,000 sells for \$1,983.00.

$HPR = 17.00/1,983.00 = 0.8573\%$; annualized $= 1.008573^{365/91} - 1 = 3.48\%$.

If the annualized rate then rises to 4.5%, the price falls to $2,000/1.045^{91/365} = \$1,978.17$, a price loss of about \$4.83. When market interest rates rise, fixed-income prices fall (a negative price return), and the effect is larger for longer-dated bonds. A rise in the issuer's perceived default risk also lowers bond values; government debt is usually regarded as less exposed to default than corporate debt.

Foreign exchange rates

A **foreign exchange rate** is the price of one country's currency stated in units of another currency (for example, 1 EUR = 1.08 USD). Exchange-rate moves change the returns of investors who hold assets denominated in a foreign currency.

The return on a foreign asset measured in its own currency is the **local currency return**; the return in the investor's home currency is the **domestic currency return**:

KEY CONCEPT

$$1 + r_{\text{domestic}} = (1 + r_{\text{local}})(1 + r_{FX})$$

where r_{FX} is the percentage change in the value of the foreign currency (negative if it depreciates). A strengthening foreign currency raises the domestic return; a weakening one lowers it. *Example.* A UK investor earns a 9% local return on US shares while the dollar appreciates 2% against the pound: $1.09 \times 1.02 - 1 = 11.18\%$. Simply adding ($9\% + 2\% = 11\%$) ignores the cross-product and is only an approximation. Exchange-rate changes affect both parts of a foreign asset's return: the price change and the income are each converted at the new rate.

Market indexes

Market indexes track the average performance of a set of investments (a market, sector, region or asset class). Because an index is more diversified than a single security, its returns tend to be less volatile. Volatility lowers compound (geometric) returns, so over many periods an index tends to have a higher geometric mean relative to its arithmetic mean than a volatile single stock does.

Example. Stock returns +30%, -20%, +10%; index returns +8%, +2%, +5%.

	Arithmetic mean	Geometric mean
Single stock	6.67%	$(1.30 \times 0.80 \times 1.10)^{1/3} - 1 = 4.59\%$
Index	5.00%	$(1.08 \times 1.02 \times 1.05)^{1/3} - 1 = 4.97\%$

The stock "wins" on the arithmetic mean but the index wins on the compound (geometric) return. The **arithmetic mean** ignores compounding, so it does not capture this **volatility drag** and overstates multiperiod performance, most of all for the more volatile asset (the individual stock).

To compare two compound returns, measure the relative outperformance on the same compound basis: $(1 + R_A)/(1 + R_B) - 1$. *Example.* Compound returns of 9.40% for one asset and 7.10% for another give $1.0940/1.0710 - 1 = 2.15\%$, slightly less than the 2.30-point simple difference.

COMMON EXAM TRAPS

- Reporting the price change alone as the total return, which leaves out the distributions.
- Treating dividends as the main source of equity return, or price change as the main source of a traditional bond's return.

- T-bill annualization: dividing by face value instead of the purchase price, or multiplying by $365/t$ instead of compounding. In the 91-day example, dividing the \$17.00 by the \$2,000 face value gives an annualized 3.45%, and multiplying 0.8573% by 365/91 gives 3.44%, instead of 3.48%.
- Domestic return: treating a depreciating foreign currency as an appreciation, or adding the two returns instead of compounding them. A depreciation enters as $(1 - \text{depreciation})$. In the UK investor example, treating the 2% dollar appreciation as a depreciation gives $1.09 \times 0.98 - 1 = 6.82\%$ instead of 11.18%.
- Saying that arithmetic means understate the multiperiod returns of volatile assets. Higher volatility lowers the geometric mean, so the arithmetic mean overstates them.

EXAM SHORTCUTS

- A strengthening foreign currency puts the domestic return above the local return and a weakening one puts it below, so answers on the wrong side of the local return can be ruled out before compounding.

BOTTOM LINE

- Total return is the capital appreciation (price) return plus the capital distribution return; distributions are realized when received, price gains only when the asset is sold, and unrealized gains still count in total return.
- Equity returns come mainly from price return, with variable and discretionary dividends, while a traditional bond's return comes mainly from its fixed coupon and price return is usually the smaller part.
- Interest rates, market indexes and foreign exchange rates are financial indicators: they influence returns but produce no cash flows of their own.
- A T-bill's holding period return is $(\text{Face value} - \text{Price})/\text{Price}$, annualized as $(1 + HPR)^{365/t} - 1$; when market interest rates rise, fixed-income prices fall, and longer-dated bonds fall more.
- The domestic currency return is $(1 + r_{\text{local}})(1 + r_{FX}) - 1$, where r_{FX} is the change in the value of the foreign currency and is negative if it depreciates.
- Volatility lowers the geometric (compound) return, so the arithmetic mean overstates multiperiod performance, most of all for the more volatile asset.

QUICK CHECK

Question 1

Tomasz Wierzbicki holds a portfolio of traditional fixed-coupon bonds. Which statement *best* describes how his total return is typically made up?

- A** Most of the return comes from the fixed coupon payments, with price return making up a smaller share.
- B** Most of the return comes from changes in the bonds' market prices, with the coupons contributing a smaller share.
- C** The fixed coupons and the price return typically contribute roughly equal shares.

Answer: A

A bond's coupon is a fixed capital distribution, and for traditional bonds it is usually the largest part of total return. The bond's market price can move while it is held, which creates some price return, but that is normally the smaller part.

Why the other options are wrong

- **B.** Price return as the dominant source describes equities, whose value is driven by changes in earnings prospects. It does not fit traditional fixed-coupon bonds.
- **C.** There is no general tendency toward an equal split. For traditional bonds the fixed coupon usually dominates.

Key takeaway Bonds: return mainly from fixed capital distributions (coupons). Equities: return mainly from price return.

MODULE 2.2

Long-Term Returns on Financial Assets

LOS 2.a — Long-term returns, risk and asset allocation

1. The risk-return tradeoff

The **risk-return tradeoff** says rational investors should demand higher long-run returns for bearing more risk. Applied to the main asset classes in a developed market, expected long-term returns rank from lowest to highest:

KEY CONCEPT

short-term government bills < government bonds < corporate bonds < large-cap equities < small-cap equities

(Government and corporate bonds are compared at similar maturities. Bills rank lowest because they are both short-term and free of credit risk.)

- *Bonds vs. equities.* Bonds are less risky (steadier, contractual distributions), so they are expected to earn less than equities.
- *Government vs. corporate bonds.* Corporate bonds carry more default risk; the extra return they offer over government bonds is the **credit spread**.
- *Short vs. long maturity.* Longer-term fixed-income securities face more inflation risk and interest rate risk, so they are expected to earn more than shorter-term ones.
- *Large-cap vs. small-cap equities.* Small-cap stocks are more volatile, so investors should demand more from them.

Relative performance depends on the period studied. The long-run record (1979–2023) falls into three phases: after 1995 equity returns and volatility rose while bond total return indexes grew steadily; in the 2008 global financial crisis all risky asset classes fell, recovering in the early 2010s; after 2015 equity returns and volatility again rose relative to bonds. Over the whole record several predictions hold. Equities outperformed bonds over the long run, longer-term bonds beat shorter-term bonds, and very short T-bills earned almost zero **real return**.

Equity **dividend yields** have fallen over time as companies return more cash through **share buybacks**. Buybacks support share prices and create capital gains, which are often taxed less than dividends. Bond distribution returns have also fallen with the multidecade fall in interest rates that lasted until the early 2020s. Two results in the same 1979–2023 data run against expectations: large-cap stocks outperformed small-cap stocks, and government bonds outperformed corporate bonds. These results reflect the period studied. Investors should still expect higher long-run returns from small caps than from large caps, and from corporate bonds than from government bonds, because those assets carry more risk.

2. Risk: systematic vs. unsystematic

Risk is the chance of earning a return different from the one expected.

KEY CONCEPT

	Systematic risk (macro risk)	Unsystematic risk (micro risk)
Affects	All investments to some degree	A particular security, issuer or industry
Reduced by diversification?	No	Yes
Examples	Inflation risk : rising prices reduce real returns. Interest rate risk : rising rates reduce the value of investments (especially long-dated bonds, and equities via higher financing costs and a lower present value of future cash flows). Market risk : a market downturn causes losses.	Issuer risk , including business risk (earnings volatility from business conditions, e.g. supply-chain problems), financial risk (use of financial leverage) and credit risk (a debt issuer's ability to meet debt payments when due). Sector risk or industry risk : a downturn specific to one industry.

Measuring risk

Risk is measured by how much returns vary around their expected value. The simplest measure is the **range**, the highest return minus the lowest return. A wider range means more variability and more risk. Standard deviation, a more common measure, is introduced in later readings. *Example*. Returns of 12%, -6%, 9% and 3% have a range of $12\% - (-6\%) = 18$ percentage points.

The extra return investors demand for bearing extra risk is a **risk premium**. Riskier assets such as equities have earned higher risk premiums over the long run than fixed income.

3. Asset allocation

A portfolio manager builds an optimal portfolio given the client's **objectives and constraints**:

Client profile	Tilt the allocation toward
Needs high returns / capital growth, higher risk tolerance	Riskier assets that deliver price return (equities, including small caps)
Needs steady periodic income, low risk tolerance (e.g., a retiree)	Assets with fixed capital distributions , i.e. fixed-income securities (bonds)
Wants both growth and income	A balanced mix of equities and fixed income

Non-interest-bearing cash cannot meet an income need.

4. Portfolio expected return

Each component of a portfolio's expected return is the **weighted average** of the asset-class components, using the portfolio weights:

KEY CONCEPT

$$E(R_p) = \sum_i w_i E(R_i),$$

total return = capital appreciation return + capital distribution return

Example. 30% equities (capital appreciation 7.0%, distribution 2.0%) and 70% bonds (capital appreciation 0.4%, distribution 4.0%):

- Capital appreciation: $0.30 \times 7.0\% + 0.70 \times 0.4\% = 2.10\% + 0.28\% = 2.38\%$
- Capital distribution: $0.30 \times 2.0\% + 0.70 \times 4.0\% = 0.60\% + 2.80\% = 3.40\%$
- Total: $2.38\% + 3.40\% = 5.78\%$

COMMON EXAM TRAPS

- Ignoring the weights. A simple average of the asset-class returns is correct only when the weights are equal, and a plain sum of the returns is not an average at all. In the 30/70 example, a simple average of the equity and bond total returns (9.0% and 4.4%) gives 6.70% instead of 5.78%.
- Answering with the wrong component: capital distribution return, capital appreciation return and total return are three different figures.
- Calling inflation or interest rate risk unsystematic because its effect differs from issuer to issuer. Both affect all investments and cannot be diversified away.
- Confusing credit risk (debt payments not made when due) with business risk (earnings volatility) or interest rate risk (rate changes moving values).

BOTTOM LINE

- Under the risk-return tradeoff, expected long-term returns in a developed market rank short-term government bills < government bonds < corporate bonds < large-cap equities < small-cap equities, with bonds compared at similar maturities.
- The credit spread is the extra return corporate bonds offer over government bonds for their greater default risk, and longer-term fixed-income securities are expected to earn more than shorter-term ones because they face more inflation and interest rate risk.
- Systematic (macro) risk, such as inflation, interest rate and market risk, affects all investments to some degree and is not reduced by diversification; unsystematic (micro) risk, such as issuer and sector risk, is.
- The range, the highest return minus the lowest return, is the simplest measure of risk, and a wider range means more variability and more risk.

- A client who needs capital growth and tolerates risk tilts toward equities, a client who needs steady income with low risk tolerance tilts toward fixed-income securities, and a client who wants both holds a balanced mix.
- Each component of a portfolio's expected return is the weighted average of the asset-class components, $E(R_p) = \sum w_i E(R_i)$, and total return is the capital appreciation return plus the capital distribution return.

QUICK CHECK

Question 2

Rafael Oduya's retirement account holds 65% in equities and 35% in government bonds. He expects a total return of 9.0% on the equities and 3.8% on the bonds. The account's expected total return is *closest* to:

- A 7.18%.
- B 6.40%.
- C 5.62%.

Answer: A

A portfolio's expected return is the weighted average of the expected returns of its components, using the allocation weights.

$$E(R_p) = 0.65 \times 9.0\% + 0.35 \times 3.8\% = 5.85\% + 1.33\% = 7.18\%$$

Adding the two returns without weights, $9.0\% + 3.8\% = 12.80\%$, counts each asset class as if it had a 100% allocation; the weights must sum to 100%, which gives the weighted 7.18%.

Why the other options are wrong

- **B.** 6.40% is the simple average $(9.0\% + 3.8\%)/2$. It would only be right with a 50/50 allocation.
- **C.** 5.62% swaps the weights, $0.35 \times 9.0\% + 0.65 \times 3.8\%$, putting 65% on the bonds. Each weight must go with its own asset class.

Key takeaway Portfolio expected return $= \sum_i w_i E(R_i)$; with non-negative weights it must lie between the lowest and highest component returns.

Practice questions

Question 3

An analyst compares the multiyear return history of Corvane Mining shares with that of a broad, diversified market index. Which conclusion is *most accurate*?

- A** Because the index is less volatile, the arithmetic mean will overstate the index's multiperiod return by more than it overstates the stock's.
- B** Because the stock is more volatile than the index, its geometric mean return tends to be pulled down relative to the index's.
- C** The arithmetic mean return will tend to understate the stock's multiperiod performance relative to the index.

Answers: 3. B

Why each answer is right and why each other option is wrong: credguild.com/cfa/level-1/notes/quantitative-methods/types-of-financial-returns/. The full bank has 19 questions on this reading.

QUANTITATIVE METHODS · READING 3

Benchmarking Returns

What you'll learn

LOS 3.a Calculate and compare money-weighted and time-weighted rates of return and judge which one fits a given performance-evaluation situation.

LOS 3.b Describe index construction choices and weighting methods (price, equal, market-capitalization, float-adjusted, fundamental), and calculate and interpret index values, divisors, weights, and price and total returns, including rebalancing and reconstitution.

MODULE 3.1

Time-Weighted and Money-Weighted Returns

This reading compares the money-weighted return, the IRR of a portfolio's cash flows, with the time-weighted return, which links subperiod returns and is unaffected by deposits and withdrawals. It then covers how a security market index is built and maintained, how price, equal, market-capitalization and fundamental weighting differ, and how to calculate an index's value, price return and total return.

LOS 3.a — Money-weighted vs. time-weighted returns

When money moves into or out of a portfolio during the evaluation period, "the portfolio's return" can mean two different things: the growth of the *investor's money*, given when it was put in, or the growth that the *manager's decisions* produced on each dollar invested. The **money-weighted rate of return** measures the first and the **time-weighted rate of return** the second.

1. Money-weighted rate of return (MWR)

The **money-weighted return** (also called the **dollar-weighted return**) is the **internal rate of return (IRR)** of the portfolio's cash flows. It is the single discount rate that gives all the flows a **net present value (NPV)** of zero:

KEY CONCEPT

$$\sum_{t=0}^T \frac{CF_t}{(1 + MWR)^t} = 0$$

Steps:

1. Choose a period length, usually one year, such that every deposit and withdrawal falls at the start or end of a period.
2. Write down the net cash flow on every date from $t = 0$ to the end, with the signs in the table below. Enter 0 for a date on which no money moves in or out. The calculator treats the entries as equally spaced, so leaving out an empty date would move every later cash flow one period earlier.
3. Solve for the IRR, for example with the calculator's cash flow worksheet (keystrokes below).

Sign convention (from the investor's point of view; the account's point of view simply reverses every sign):

KEY CONCEPT

Item	Sign
Beginning value / initial purchase	outflow (–)
Additional deposits or share purchases	outflow (–)
Dividends or income withdrawn (not reinvested), withdrawals	inflow (+)
Ending value / sale proceeds	inflow (+)

A dividend received on the same date as a new purchase is netted with it into one cash flow. For example, a dividend of 2 and a purchase of 50 give $CF = -48$. The IRR is unchanged if every sign is flipped, as long as money going in and money coming out carry opposite signs.

Calculator (TI BA II Plus): [CF] [2ND] [CLR WORK], enter CF0, C01, C02 ... (each frequency = 1), then [IRR] [CPT]. The resulting IRR is a rate per period. If the periods are years, it is already an annual rate.

2. Time-weighted rate of return (TWR)

The **time-weighted return** measures **compound growth**: the rate at which one unit of money invested at the start would have grown. Deposits and withdrawals do not affect it. Steps:

1. Value the portfolio immediately before each significant deposit or withdrawal, and split the evaluation period into **subperiods** at those dates.
2. Calculate each subperiod's **holding period return (HPR)**. When there is a cash flow, the beginning value of the next subperiod is the value *after* that deposit or withdrawal:

$$HPR_i = \frac{V_{end,i} - V_{begin,i} + D_i}{V_{begin,i}}$$

3. **Geometrically link** the subperiod returns. If the total period is longer than one year, take the geometric mean to annualize, using the number of years (not the number of subperiods):

$$TWR_{annual} = [(1 + HPR_1)(1 + HPR_2) \cdots (1 + HPR_n)]^{1/\text{years}} - 1$$

For a single share bought and later added to, the per-share HPRs can be used directly: $HPR_1 = (P_1 - P_0 + D_1)/P_0$, and so on.

3. Example

An investor opens an account with \$1,000. One year later the account is worth \$1,150 (+15%), and she deposits another \$500, bringing it to \$1,650. At the end of year 2 the account is worth \$1,782 (+8% in year 2).

- **TWR:** $[(1.15)(1.08)]^{1/2} - 1 = 11.45\%$ per year.
- **MWR:** solve $-1,000 - \frac{500}{1+r} + \frac{1,782}{(1+r)^2} = 0$, which gives $r = 10.81\%$. (CF0 = -1,000; C01 = -500; C02 = +1,782.)

The MWR is lower than the TWR because the extra \$500 was invested just before the weaker year. The MWR gives more weight to periods in which more money is invested.

4. Comparing the two measures

KEY CONCEPT

Feature	Money-weighted return	Time-weighted return
Method	IRR of all cash flows	Geometric linking of subperiod HPRs
Affected by the timing and size of external cash flows?	Yes	No
Deposit made just before strong performance	MWR higher than TWR	unaffected
Deposit made just before poor performance	MWR lower than TWR	unaffected
No external cash flows during the period (no deposits, withdrawals or paid-out income)	equal to TWR	equal to MWR
Best use	Manager (or investor) controls the cash flows	Manager does not control the cash flows

- The investment management industry prefers the time-weighted return because managers usually do not decide when clients add or withdraw money. The TWR isolates the manager's skill in selecting investments.
- The **Global Investment Performance Standards (GIPS)** are voluntary. A firm that claims compliance must present performance using **time-weighted returns**, except in specific situations where **money-weighted returns** are more appropriate. These require that the manager controls external cash flows and, in addition, that the portfolio is closed-end, has a fixed life, calls committed capital from investors, or holds mainly illiquid investments (typical of private equity).
- Withdrawals work the opposite way to deposits. Taking money out just before a strong period lowers the MWR relative to the TWR.

COMMON EXAM TRAPS

- Leaving out dividends that are paid out and not reinvested. For the MWR they are inflows on their payment date. For the TWR they belong in the numerator of that subperiod's HPR.
- Computing a TWR subperiod return without first adding the deposit just made to (or removing the withdrawal from) the beginning value, which mixes new money with investment results. In the example, measuring year 2 from \$1,150 instead of the \$1,650 after the deposit gives 54.96% instead of 8%.
- Annualizing over the number of subperiods instead of the number of years. Three quarterly subperiods inside one year need no root.
- Averaging subperiod returns arithmetically. The TWR links them geometrically. In the example, $(15\% + 8\%)/2 = 11.50\%$ instead of the TWR of 11.45%.
- The "required return" of an investor is irrelevant to both calculations. It is a hurdle rate for judging the result.
- If the subperiods are of equal length and each earns the same rate, both measures equal that rate whatever the cash flows. Differences come from cash flows combined with uneven subperiod returns.

EXAM SHORTCUTS

- With no deposits, withdrawals or paid-out income during the period, the MWR equals the TWR, so no IRR has to be computed.
- A deposit made just before the stronger subperiod puts the MWR above the TWR, and one made just before the weaker subperiod puts it below; withdrawals work the opposite way, so comparison questions need no IRR.
- If the subperiods are of equal length and each earns the same rate, the MWR and the TWR both equal that rate whatever the cash flows.

BOTTOM LINE

- The money-weighted return is the IRR of the portfolio's cash flows, the rate that gives them an NPV of zero, with money going in and money coming out entered with opposite signs.
- The time-weighted return geometrically links the holding period returns of subperiods formed at each significant deposit or withdrawal, and is annualized with a root equal to the number of years, not the number of subperiods.
- The MWR depends on the timing and size of external cash flows: a deposit just before strong performance makes it higher than the TWR and a deposit just before poor performance makes it lower, while the TWR is unaffected.
- The MWR and the TWR are equal when there are no external cash flows during the period.
- The investment management industry prefers the TWR because managers usually do not control when clients add or withdraw money; the MWR suits a manager or investor who controls the cash flows.
- Firms claiming compliance with the voluntary GIPS must present time-weighted returns, except where the manager controls external cash flows and the portfolio is also closed-end, has a fixed life, calls committed capital or holds mainly illiquid investments.

QUICK CHECK**Question 1**

Hannah Kowalczyk makes the following trades in Silverleaf Foods shares, with no taxes or transaction costs:

- Today: buys one share at \$30.00.
- One year from today: buys a second share at \$36.00.
- Two years from today: sells both shares at \$41.00 each.

Silverleaf pays a dividend of \$1.50 per share at the end of year 1 and \$2.00 per share at the end of year 2. Dividends are not reinvested. Kowalczyk's required return is 10%. Her money-weighted rate of return is *closest* to:

- A 15.9%.
- B 21.3%.
- C 22.2%.

Answer: B

The money-weighted return is the internal rate of return of the cash flows the investor actually pays and receives. At the end of year 1 the \$1.50 dividend on the first share is netted against the \$36.00 purchase. At the end of year 2 she receives the sale proceeds and the dividends on two shares. The 10% required return is not an input.

t Cash flow

0 -30.00

1 +1.50 - 36.00 = -34.50

2 $2 \times 41.00 + 2 \times 2.00 = +86.00$

$$-30.00 - \frac{34.50}{1+r} + \frac{86.00}{(1+r)^2} = 0 \Rightarrow r = 0.2131 \approx 21.3\%$$

TI BA II Plus: [CF] [2ND] [CLR WORK]; CF0 = -30; C01 = -34.5, F01 = 1; C02 = 86, F02 = 1; [IRR] [CPT] = 21.31.

Why the other options are wrong

- **A.** 15.9% is the IRR after ignoring both dividends (CF1 = -36, CF2 = +82). Dividends that are paid out are cash received by the investor.
- **C.** 22.2% is the time-weighted return: $HPR_1 = (36 - 30 + 1.5)/30 = 25.0\%$ and $HPR_2 = (41 - 36 + 2)/36 = 19.4\%$, so $\sqrt{1.250 \times 1.194} - 1 = 22.2\%$. This measure ignores the fact that more money was invested in year 2.

Key takeaway For an MWR share problem, net the dividend received against any purchase on the same date, and include dividends on all shares held at the end.

MODULE 3.2

Security Market Indexes

LOS 3.b — Security market indexes: construction, weighting and returns

1. What an index is and how its return is measured

A **security market index** represents the performance of an asset class, a security market or a segment of a market. It is built as a portfolio of **constituent securities**. Its value at any time is calculated from the market prices of those securities. The percentage change in the index value over a period is the **index return**.

KEY CONCEPT

Version	What it uses	Return it produces
Price index (price return index)	Constituent prices only	Price return

Version	What it uses	Return it produces
Return index (total return index)	Prices plus income (dividends, interest, other distributions), assumed reinvested	Total return

When constituents pay cash distributions, the total return is greater than the price return, and a total return index cannot grow more slowly (or fall faster) than the matching price index. With no distributions, the two are identical.

Returns for successive periods are **geometrically linked**:

KEY CONCEPT

$$R_{\text{period}} = (1 + R_1)(1 + R_2) \cdots (1 + R_n) - 1 \quad V_n = V_0(1 + R_1)(1 + R_2) \cdots (1 + R_n)$$

Example. Monthly index returns of +1.8%, -0.6% and +2.1% link to $(1.018)(0.994)(1.021) - 1 = 3.31\%$. Their sum, 3.3%, is only an approximation. Starting from 500, the index ends at $500 \times 1.0331 = 516.6$.

2. Building an index: five decisions, in order

1. **Target market / objective:** which market or segment the index should measure. This is always the first decision, and it drives everything else.
2. **Constituent securities:** the selection criteria and the securities that meet them.
3. **Weighting method:** price, equal, market-capitalization (possibly float-adjusted) or fundamental weighting.
4. **Calculation:** the starting index value and how the index is computed on each later date (**divisor** and **base value**).
5. **Maintenance:** rebalancing, reconstitution and periodic review of the method.

Common index types by target market (preview of Reading 41, LOS 41.d):

Type	What it covers
Broad-based equity index	A whole market: its largest listed companies, usually float-adjusted and market-cap-weighted. A composite market index measures the overall market and usually holds more than 90% of its total value.
Multi-market index	Several national markets (e.g. a regional, emerging-market or global index)
Sector index	One industry or economic sector (e.g. financials, health care), in one country or globally
Investment theme index	Stocks tied to an investor trend beyond industry or company size (e.g. an ESG-screened index)

3. Weighting methods

Price-weighted index

A **price-weighted index** is the sum of the constituent prices divided by a divisor:

$$\text{Index} = \frac{\sum P_i}{D} \quad w_i = \frac{P_i}{\sum P_j}$$

- At launch, $D = \sum P_i / \text{base value}$.
- The divisor must be changed whenever a **stock split**, **reverse stock split** or **stock dividend** occurs, or when the constituents change, so that the event alone does not move the index. A stock dividend works like a split stated as a percentage: after a 10% stock dividend, one share becomes 1.1 shares. Unlike a cash dividend, it pays out nothing and leaves each holder's ownership unchanged. For a split, recompute the sum with the post-split price and solve $D_{\text{new}} = \sum P_{\text{new}} / \text{index value}$.
- Higher-priced stocks have more weight. A 10% move in the highest-priced stock moves the index most, whatever the company's size.
- A portfolio holding an equal number of shares of each constituent replicates the index.
- It is simple to compute, but a share price says nothing about a company's value, because share counts differ widely. It is therefore a suitable benchmark mainly for a portfolio whose results depend disproportionately on the price changes of particular securities. Examples: the DJIA (30 US stocks) and the Nikkei 225.

Example. Prices of \$14, \$38 and \$73 (sum \$125) with a base value of 100 give $D = 1.25$. If the \$73 stock splits 2-for-1, the sum becomes $14 + 38 + 36.50 = 88.50$, and the new divisor is $88.50 / 100 = 0.885$.

Equal-weighted index

In an **equal-weighted index**, every constituent gets the same weight $1/N$ at each rebalancing.

- At inception, each stock is given the same **reference value**. The number of shares held is reference value $\div$ price. The **aggregate reference value** is $N \times$ reference value, and $D = \text{aggregate reference value} / \text{base value}$.
- For a period that starts at equal weights, the index return is the **arithmetic mean** of the constituent returns. The geometric mean does not apply, because the stocks are held side by side.
- Weights drift as prices move. Winners grow above $1/N$ and losers fall below it. To restore equal weights, the index must be rebalanced frequently, which means selling winners and buying losers. This is costly for a tracking portfolio.
- Small companies are overweighted relative to their size. The index is a suitable benchmark for a strategy that limits exposure to the largest companies, or for a fund that invests equal amounts in each holding. Examples: Value Line Composite Average, FT Ordinary Share Index.

Market-capitalization-weighted (value-weighted) index

In a **market-capitalization-weighted index**, each constituent's weight is its share of the total market value of the constituents:

$$w_i = \frac{P_i N_i}{\sum P_j N_j} \quad \text{Index} = \frac{\sum P_i N_i}{D}, \quad D = \frac{\text{base-period market cap}}{\text{base value}}$$

- Stock splits and stock dividends do not change a company's market value, so an ordinary stock split or stock dividend by itself leaves the stock's weight and the divisor unchanged. The divisor is adjusted when constituents are added or removed (reconstitution). The new divisor is the new total market cap divided by the unchanged index value.

- It tracks changes in aggregate investor wealth. A tracking portfolio rebalances itself automatically, and turnover is low.
- The main drawback is that a stock's weight rises as its price rises. Possibly overvalued stocks get more weight and possibly undervalued stocks get less.
- It is the usual benchmark for a passive strategy that aims to earn the market return (e.g. S&P 500).

Float adjustment

The **market float** is the value of shares actually available to the investing public. It excludes shares held by controlling (strategic) shareholders, and often those held by other corporations or governments. **Free float** is narrower still, because it also excludes shares that foreign investors cannot buy. A **float-adjusted market-capitalization-weighted index** weights each firm by float-adjusted capitalization = market cap × float %. Firms with large controlling holdings get less weight than in an unadjusted index. **Float adjustment** belongs to market-cap weighting only.

Fundamental-weighted index

In a **fundamental-weighted index**, weights come from company fundamentals like earnings, dividends, cash flow, sales or book value. These weights are not driven by the share price, although over long periods they are related to it. Such an index has a **value tilt**: compared with market-cap weighting, it overweights stocks with low P/E or high book-to-market ratios. It suits value-oriented and other fundamentals-driven active strategies. Its advantage is that it looks beyond prices and market values. Its drawback is that it may leave out firms without positive earnings or without dividends, which can bias it toward or away from some sectors. Fundamental weights also drift, so the index needs rebalancing.

KEY CONCEPT

Method	Weight driven by	Adjust divisor for splits?	Rebalancing need	Typical bias
Price	Share price	Yes	Automatic	Toward high-priced shares
Equal	$1/N$	No	High	Toward small caps
Market cap / float-adjusted	Market value	No	Automatic	Toward large and recently risen stocks
Fundamental	Earnings, book value, etc.	No	Periodic	Value tilt

4. Example: price return vs. total return

A market-cap-weighted index has a base value of 1,000 when constituent capitalization is \$500 million, so $D = 500,000$. One period later the capitalization is \$515 million, and constituents paid \$6 million of dividends.

- Price index value = $515,000,000 / 500,000 = 1,030$, so the price return is 3.0%.
- Total return index value = $(515 + 6) \text{ million} / 500,000 = 1,042$, so the total return is 4.2%.
- The capital distribution return is $6/500 = 1.2\%$.

From per-share data, compute each constituent's market value as price $\times$ shares at the start and at the end, and its dividends as dividend per share $\times$ shares. The market-cap total return is then (ending value + dividends – beginning value) $\div$ beginning value.

5. Maintenance: rebalancing vs. reconstitution

- **Rebalancing** means resetting the constituents' weights to target once price moves have shifted them, usually every quarter. It matters mainly for **equal-weighted** and **fundamental-weighted** indexes. When an equal-weighted index is rebalanced, the total value (and the index value) is left unchanged. Each position is simply reset to total $\div$ N. For example, with two holdings worth 6,300 and 5,500, each is reset to 5,900: sell 400 of the first and buy 400 of the second. Selling recent winners and buying recent losers runs against a trend-following approach.
- **Reconstitution** means changing which securities are in the index. Securities are deleted when they no longer meet the criteria (for example, they fall out of the size range), mature, are acquired, go bankrupt or are delisted. They are replaced by securities that do meet the criteria. The divisor is adjusted on that date so the index value does not jump.
- Prices react. An added security tends to rise as index-tracking managers buy it, and a deleted security tends to fall as they sell it. The weights of the remaining constituents are also adjusted to fit the weighting scheme.
- Changing the calculation formula or the rules is neither rebalancing nor reconstitution.

COMMON EXAM TRAPS

- Confusing the aggregate dollar value, the divisor and the index value.
- After a split in a price-weighted index, dividing every price by the split ratio or keeping the old divisor. In the example, keeping $D = 1.25$ after the split gives an index of $88.50/1.25 = 70.80$ instead of 100.
- Changing a stock's market-cap weight, or the divisor of a market-cap-weighted index, because of a split.
- Compounding the constituent returns of an equal-weighted index within one period, or adding index returns across periods instead of linking them.
- Leaving income out of a total return, or adding it to a price return.
- Finding a weight after price moves by growing the old weight by the stock's own return. Divide each position's new value by the new total.
- Weighting a float-adjusted index by floating share counts instead of float-adjusted market value (market cap $\times$ float %).

EXAM SHORTCUTS

- For a period that starts at equal weights, an equal-weighted index return is the arithmetic mean of the constituent returns, so no share counts or divisor are needed.
- Only a price-weighted index changes its divisor for a stock split or stock dividend; equal-weighted, market-cap-weighted and fundamental-weighted indexes leave it unchanged.

BOTTOM LINE

- A price index uses constituent prices only and gives the price return; a total return index adds income, assumed reinvested, and its total return exceeds the price return when constituents pay distributions.

- A price-weighted index is the sum of prices divided by a divisor that is changed for splits, reverse splits, stock dividends and constituent changes, and its higher-priced stocks carry more weight.
- An equal-weighted index gives each constituent $1/N$ at each rebalancing, overweights small companies relative to their size and needs frequent, costly rebalancing as weights drift.
- A market-capitalization-weighted index weights each stock by its share of total market value, tracks aggregate investor wealth and rebalances automatically, but gives more weight to stocks whose prices have risen; a float-adjusted version uses market cap $\times$ float %.
- A fundamental-weighted index weights constituents by fundamentals such as earnings, dividends, cash flow, sales or book value and has a value tilt.
- Rebalancing resets weights to target after price moves and matters mainly for fundamental-weighted and equal-weighted indexes; reconstitution changes the constituents, with the divisor adjusted on that date.

QUICK CHECK

Question 2

Which of the following statements about a total return index is *most accurate*?

- A** Its value reflects only the price changes of its constituent securities.
- B** It grows from its beginning value by compounding each period's total return.
- C** It can rise more slowly than a price return index built from the same securities with the same weights.

Answer: B

The value of a total return index grows by compounding its periodic total returns: $V_n = V_0(1 + R_1)(1 + R_2) \cdots (1 + R_n)$. Each periodic total return includes both price changes and income such as dividends and interest.

Why the other options are wrong

- **A.** An index whose value reflects only price changes is a price return index. A total return index also includes the income (dividends, interest, other distributions) paid by the constituents.
- **C.** Income paid by the constituents can never be negative. A total return index therefore always rises at least as fast as (or falls no faster than) the matching price return index.

Key takeaway Total return index $\geq$ price return index; values compound by geometric linking.

Practice questions

Question 3

The prices of the stocks in a price-weighted index add up to \$270, and the divisor is 1.80. One constituent, trading at \$90, then completes a 3-for-1 stock split. The divisor that should be used immediately after the split is *closest* to:

- A 0.60.
- B 1.40.
- C 1.80.

Question 4

An equal-weighted index contains the three stocks shown in the table below. After the period's price changes, and before any rebalancing, Kestrel Aerospace's index weight is *closest* to:

Equal-weighted index: reference values and price returns for the period

Stock	Reference value at start	Price return for the period
Kestrel Aerospace	\$25,000	+12%
Linden Packaging	\$25,000	+4%
Oakmere Retail	\$25,000	-6%

- A 33.3%.
- B 34.8%.
- C 36.1%.

Question 5

The sponsor of the Harrowgate Midcap 200 Index removes Velden Tools, which was delisted after filing for bankruptcy, and adds Corvin Optics, which now meets the index's size criteria. This change is *best* described as:

- A reconstitution.
- B rebalancing.
- C redefinition.

Answers: 3. B · 4. C · 5. A

Why each answer is right and why each other option is wrong: credguild.com/cfa/level-1/notes/quantitative-methods/benchmarking-returns/. The full bank has 61 questions on this reading.

QUANTITATIVE METHODS · READING 4

Time Value of Money

What you'll learn

LOS 4.a Calculate and interpret the present value of fixed-income instruments (zero-coupon, coupon, perpetual, amortizing) and equity instruments (preferred stock, constant-growth and multistage DDM) from their expected cash flows.

LOS 4.b Calculate and interpret the implied return of fixed-income instruments and the implied required return and growth rate of equity, given price and cash flows.

LOS 4.c Explain the cash flow additivity principle and the no-arbitrage condition, and use them to derive implied forward interest rates, forward exchange rates and option values.

MODULE 4.1

Discounted Cash Flow Valuation

This reading prices fixed-income and equity instruments as the present value of their expected cash flows and then works the same relationships backward to find implied returns and growth rates. It closes with cash flow additivity and the no-arbitrage principle, which give forward interest rates, forward exchange rates and the value of an option in a one-period binomial model. Most of the calculations use the financial calculator's TVM worksheet.

LOS 4.a — Present value of fixed-income and equity instruments

1. The time value of money

The **time value of money** means a given amount of money today is worth more than the same amount in the future. To give up money now, investors must expect a **rate of return**; that same rate, used as a **discount rate**, converts future cash flows into a **present value (PV)** and present amounts into a **future value (FV)**.

KEY CONCEPT

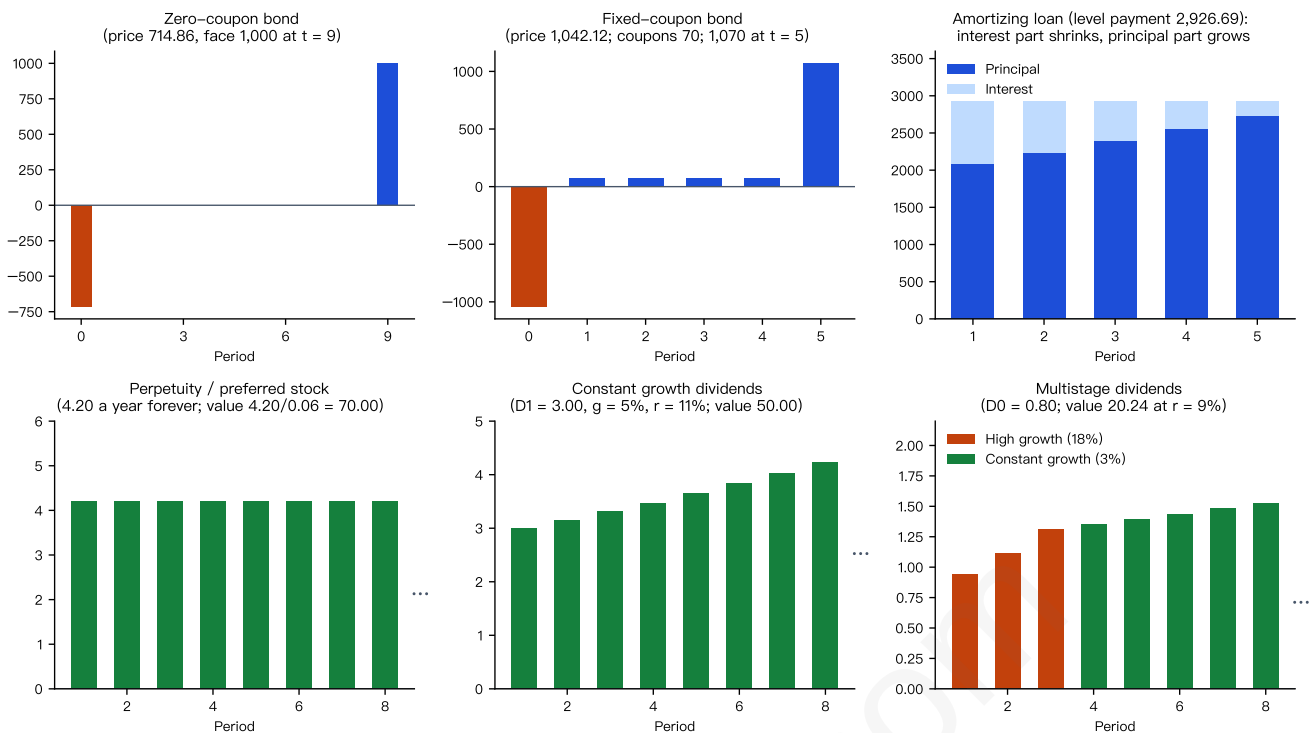
$$FV = PV(1 + r)^t \quad PV = \frac{FV}{(1 + r)^t}$$

where r is the rate per compounding period and t is the number of periods. With **continuous compounding**: $FV = PVe^{rt}$ and $PV = FVe^{-rt}$.

Financial calculator (TI BA II Plus, P/Y = C/Y = 1 and END mode unless stated) keys: N = number of periods, I/Y = rate per period (in %), PV = present value, PMT = constant periodic payment, FV = future value, CPT = compute. Enter outflows as negative and inflows as positive (e.g., price paid negative, face value received positive). Clear the TVM worksheet between problems.

With m compounding periods a year, use the rate r/m and years $\times m$ periods. Unequal cash flows are moved to the target date one at a time and added: at 6% compounded semiannually, 2,000 paid at the end of year 1 is worth $2,000 \times 1.03^4 = 2,251.02$ at the end of year 3.

The instruments below differ only in the pattern of their cash flows:



Cash flow patterns of the instruments in this module (numbers from the module's examples)

2. Fixed-income instruments

Pure discount instruments

A **pure discount instrument (zero-coupon bond)** is bought for less than its face value, and the investor receives the **face value** at maturity; the difference is the interest earned. Its price is set by two inputs: the time left to maturity and the yield to maturity, which is the rate used to discount the face value.

Example. A \$1,000 face value bond with 9 years to maturity and a yield of 3.8% is priced at $1,000/1.038^9 = \$714.86$ ($N = 9$, $I/Y = 3.8$, $PMT = 0$, $FV = 1,000$, $CPT PV = -714.86$).

If the yield is negative, the zero-coupon bond is priced at a premium (above face value): at -0.4% the same bond costs $1,000/0.996^9 = \$1,036.73$.

Fixed-coupon bonds

For a **fixed-coupon bond**, the **coupon rate** × face value sets the coupon (PMT). The **yield to maturity** (I/Y) is the rate that, used to discount the bond's cash flows, reproduces its price. The two rates generally differ. The price is the PV of the coupons and the face value, all discounted at the YTM:

$$P = \sum_{t=1}^N \frac{C}{(1 + YTM)^t} + \frac{F}{(1 + YTM)^N}$$

Example. A 5-year, \$1,000 bond with a 7% annual coupon and a YTM of 6%: $N = 5$, $I/Y = 6$, $PMT = 70$, $FV = 1,000$, $CPT PV = -1,042.12$.

KEY CONCEPT

Relationship	Price vs. par
Coupon rate > YTM	Premium (price > par)
Coupon rate = YTM	Par
Coupon rate < YTM	Discount (price < par)

A **floating-rate bond** resets its coupon to a **market reference rate** plus a fixed margin (e.g., reference 2.5% + 150 basis points = 4.0%); one **basis point** = 0.01%.

Perpetual bonds

A **perpetual bond** (a **perpetuity**) pays a fixed amount forever and has no maturity:

$$PV_{\text{perpetuity}} = \frac{PMT}{r}$$

The first payment is assumed to arrive one period from now.

Amortizing bonds and annuities

An **annuity** is a series of equal payments at equal intervals. The formulas below assume that each payment comes at the end of its period, so the first payment is one period away. An **amortizing bond** (or a fixed-rate mortgage) pays a constant amount that includes both interest and principal. As the outstanding balance falls, the interest portion of each payment declines and the principal portion increases. Whether the early payments are *mostly* interest depends on the rate and the term (with a low rate or a short term, even the first payment can be mostly principal).

$$PV_{\text{annuity}} = PMT \times \frac{1 - (1 + r)^{-N}}{r} \quad PMT = \frac{r \times PV}{1 - (1 + r)^{-N}}$$

Example. A \$12,000 loan at 7% repaid with 5 equal year-end payments: $N = 5$, $I/Y = 7$, $PV = 12,000$, $FV = 0$, $CPT PMT = -2,926.69$. For monthly payments use $N = \text{years} \times 12$ and $I/Y = \text{annual rate} \div 12$.

To split a payment into interest and principal:

1. Interest = $r \times$ the balance at the start of the period.
2. Principal repaid = payment – interest.
3. New balance = old balance – principal repaid.

For this loan, the first payment carries interest of $0.07 \times 12,000 = \$840.00$ and repays $2,926.69 - 840.00 = \$2,086.69$ of principal, leaving \$9,913.31. The second carries interest of $0.07 \times 9,913.31 = \$693.93$ and repays \$2,232.76. The interest part falls each year by exactly the amount the principal part rises. The full schedule for the loan:

Amortization schedule: \$12,000 at 7%, five year-end payments

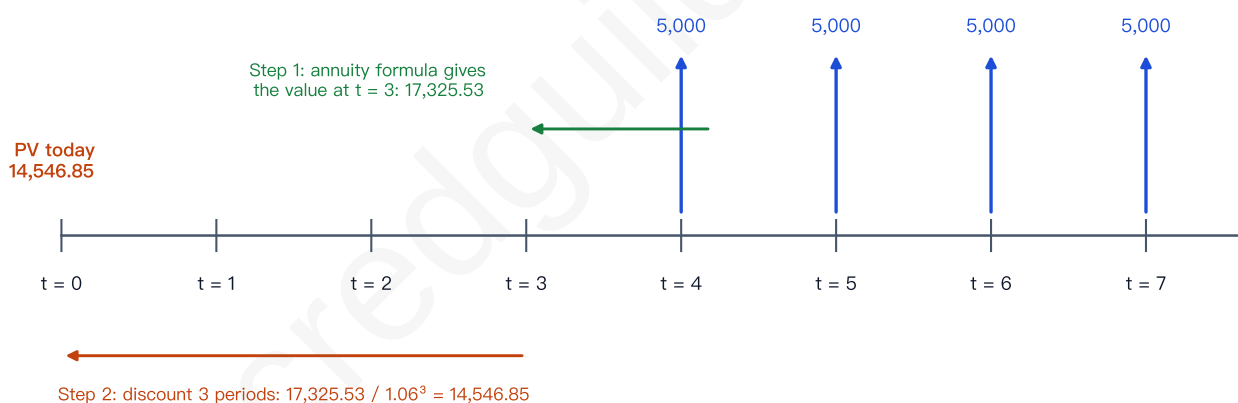
Year	Beginning balance	Payment	Interest (7%)	Principal repaid	Ending balance
1	12,000.00	2,926.69	840.00	2,086.69	9,913.31
2	9,913.31	2,926.69	693.93	2,232.76	7,680.56
3	7,680.56	2,926.69	537.64	2,389.05	5,291.51
4	5,291.51	2,926.69	370.41	2,556.28	2,735.22
5	2,735.22	2,926.69	191.47	2,735.22	0.00

Annuities that start later

The annuity formula gives a value one period before the first payment. For a later first payment:

1. Apply the annuity formula (value one period before the first payment).
2. Discount that value to today as a single sum.

Example. Four payments of 5,000 start at $t = 4$, and the rate is 6%. Step 1 gives $5,000 \times [1 - 1.06^{-4}] / 0.06 = 17,325.53$ at $t = 3$. Step 2 gives $17,325.53 / 1.06^3 = 14,546.85$ today.



Valuing an annuity whose first payment is at $t = 4$

3. Equity instruments

Equity securities are valued as the PV of expected future cash flows discounted at the investor's **required return**, the return that induces investors to hold the share. Equity usually has no maturity, and its cash flows can vary.

Preferred stock pays a fixed dividend (a stated percentage of **par value**) indefinitely, so it is valued as a perpetuity:

$$V_{\text{preferred}} = \frac{D_p}{r_p}$$

where D_p is the dividend per period and r_p is the market's required return on the preferred stock.

Example. A \$4.20 annual dividend and a 6% required return give a value of \$70.00. The stated dividend rate sets the cash flow; the required return is the discount rate.

Common stock is a **residual claim** on the company's assets after all other claims are met. Dividends are not promised; management decides whether and when to pay them. The **dividend discount models (DDMs)**:

1. **Constant dividend:** value the share as a perpetuity, as with preferred stock.
2. **Constant growth DDM (Gordon growth model):**

$$V_0 = \frac{D_1}{r_e - g_c}, \quad D_1 = D_0(1 + g_c)$$

Here V_0 is the value of all dividends from D_1 onward, r_e is the required return on common equity and g_c is the constant growth rate of dividends. The model needs the dividend expected next period (D_1), a constant growth rate, and $r_e > g_c$. *Example.* With $D_1 = \$3.00$, $r_e = 11\%$ and $g_c = 5\%$, $V_0 = 3.00/0.06 = \$50.00$.

3. **Multistage DDM:** forecast the dividends of a high-growth phase individually, then apply the Gordon model at the start of the constant-growth phase. The Gordon model gives a value one period before the dividend in its numerator: $P_n = D_{n+1}/(r_e - g_c)$. The high-growth rate may be above r_e , because those dividends are discounted one at a time (18% against 9% in the example below). Only the constant growth rate used in the Gordon step must be below r_e .

Multistage example. $D_0 = \$0.80$, growth 18% for 3 years, then 3% forever, $r_e = 9\%$.

$D_1 = 0.944$, $D_2 = 1.1139$, $D_3 = 1.3144$, $D_4 = 1.3144 \times 1.03 = 1.3539$

$P_3 = 1.3539/(0.09 - 0.03) = 22.564$

$$V_0 = \frac{0.944}{1.09} + \frac{1.1139}{1.09^2} + \frac{1.3144 + 22.564}{1.09^3} = \$20.24$$

WORKED EXAMPLE: PRICING A BOND AND SPLITTING A LOAN PAYMENT

A 5-year bond has a \$1,000 face value, a 5.25% annual coupon and a YTM of 4.35%. Separately, a borrower takes a \$250,000 loan at 4.8%, repaid with 10 equal year-end payments.

Step 1. The coupon rate (5.25%) is above the YTM (4.35%), so the bond must sell at a premium.

Step 2. Price the bond: $N = 5$; $I/Y = 4.35$; $PMT = 52.50$; $FV = 1,000$; $CPT PV = -1,039.68$. The price is \$1,039.68, above par as predicted.

Step 3. One year later, with the YTM unchanged: $N = 4$; $I/Y = 4.35$; $PMT = 52.50$; $FV = 1,000$; $CPT PV = -1,032.40$. The price is still above par, but closer to it.

Step 4. Loan payment: $N = 10$; $I/Y = 4.8$; $PV = 250,000$; $FV = 0$; $CPT PMT = -32,062.44$.

Step 5. First payment: interest $0.048 \times 250,000 = \$12,000.00$; principal $32,062.44 - 12,000.00 = \$20,062.44$; new balance $\$229,937.56$.

Step 6. Second payment: interest $0.048 \times 229,937.56 = \$11,037.00$; principal $32,062.44 - 11,037.00 = \$21,025.44$. The interest part fell by \$963.00 and the principal part rose by the same amount.

Result. The bond is priced at \$1,039.68. The loan payment is \$32,062.44, of which \$12,000.00 is interest in year 1 and \$11,037.00 in year 2.

COMMON EXAM TRAPS

- Mixing up the coupon rate (which sets PMT) and the YTM (which is I/Y), or, when the bond's own price is given, taking the yield on similar bonds as its YTM.
- Annuities and single sums: confusing a present value with a future value, adding up payments without discounting them, or discounting with simple interest, $FV/(1 + rt)$, which ignores compounding.
- Perpetuity value = payment ÷ rate, which is already today's value when the first payment is one period away. Multiplying instead, or discounting that value one more period, gives a wrong value. With the \$4.20 dividend and 6% required return above, discounting the \$70.00 value one more period gives \$66.04.
- Gordon model: using D_0 where D_1 is needed, or applying the model when growth is not constant or not below the required return.
- Multistage DDM: the terminal value P_n uses D_{n+1} and is discounted n periods, not $n + 1$. In the multistage example, discounting P_3 four periods gives \$18.80 instead of \$20.24.
- Annuity that starts later: discounting over the full number of periods to the first payment instead of one period fewer. In the example, discounting 17,325.53 four periods gives 13,723.44 instead of 14,546.85.
- Forgetting that a negative yield gives a price above face value.
- Monthly loans: converting only one of N and I/Y to months.

EXAM SHORTCUTS

- Compare the coupon rate with the YTM before calculating a bond price: a coupon above the YTM means a premium, equal means par, below means a discount, so prices on the wrong side of par can be ruled out at once.
- A zero-coupon bond, a coupon bond and a level loan are all one TVM entry; a zero is the same entry with PMT = 0.
- A level perpetuity or a preferred share needs no worksheet: value = payment ÷ rate.

BOTTOM LINE

- Every instrument here is valued as the present value of its expected cash flows, each discounted with $PV = FV/(1 + r)^t$.
- A bond sells at a premium when its coupon rate is above its YTM, at par when they are equal and at a discount when the coupon rate is below the YTM.
- A perpetuity is worth PMT/r one period before its first payment, and preferred stock is valued the same way, D_p/r_p .
- In an amortizing loan each payment's interest is the rate times the opening balance, so the interest part falls and the principal part rises over the life of the loan.
- The Gordon growth model $V_0 = D_1/(r_e - g_c)$ uses next period's dividend and requires constant growth below the required return.
- In a multistage DDM the terminal value $P_n = D_{n+1}/(r_e - g_c)$ is discounted n periods, together with the dividend paid at n .

QUICK CHECK

Question 1

Brisa Therapeutics expects to receive the licensing fees shown in the table below. Its finance team discounts them at a stated annual rate of 6%, compounded monthly. The present value of the three fees today is *closest* to:

Expected licensing fees

Received at the end of year	Fee
1	\$5,000
2	\$9,000
3	\$12,000

- A \$22,722.
- B \$22,802.
- C \$25,708.

Answer: A

The fees are unequal, so each one is discounted to today on its own and the present values are added. With monthly compounding, the rate per period is $6\%/12 = 0.5\%$ and each year has 12 periods, so the fees are discounted over 12, 24 and 36 months.

$$PV = \frac{5,000}{1.005^{12}} + \frac{9,000}{1.005^{24}} + \frac{12,000}{1.005^{36}}$$

$$= 4,709.53 + 7,984.67 + 10,027.74 = \$22,721.94$$

Calculator (P/Y = C/Y = 1, END mode), first fee: N = 12; I/Y = 0.5; PMT = 0; FV = 5,000; CPT PV = -4,709.53. Then N = 24; FV = 9,000; CPT PV = -7,984.67, and N = 36; FV = 12,000; CPT PV = -10,027.74.

Why the other options are wrong

- **B.** \$22,802 discounts at 6% with annual compounding (1.06 , 1.06^2 and 1.06^3). Monthly compounding raises the effective annual rate to $1.005^{12} - 1 = 6.17\%$, so the present value is lower.
- **C.** \$25,708 uses the monthly rate of 0.5% but discounts over only 1, 2 and 3 periods, counting years instead of months.

Key takeaway With m compounding periods a year, discount at r/m over $m \times$ years periods, and value unequal cash flows one at a time.

MODULE 4.2

Implied Returns and Cash Flow Additivity

LOS 4.b — Implied returns and implied growth

The PV relationships from Module 4.1 can be rearranged to solve for the rate of return implied by a known price and known (or assumed) cash flows.

Zero-coupon bond

$$r = \left(\frac{FV}{PV} \right)^{1/N} - 1$$

Example. A \$1,000 face value bond maturing in 6 years and priced at \$820 yields $(1,000/820)^{1/6} - 1 = 3.36\%$ ($N = 6$, $PV = -820$, $PMT = 0$, $FV = 1,000$, $CPT I/Y = 3.36$).

Coupon bond

The yield to maturity of a **coupon bond** generally has no closed-form solution, so it is found with the calculator: N , $PMT = \text{coupon}$, $FV = \text{face}$, $PV = -\text{price}$, $CPT I/Y$. *Example.* A 5-year, \$1,000 par bond with a 4% annual coupon priced at \$975.00: $N = 5$, $PMT = 40$, $FV = 1,000$, $PV = -975$, $CPT I/Y = 4.57$.

Price and yield move in opposite directions: when a bond's price rises, its YTM falls, and when the price falls, its YTM rises. If the bond above rises to \$990.00, its YTM falls to 4.23%.

Equity: implied required return

The required return on equity cannot be observed directly. Rearranging the constant growth DDM with the observed price gives:

KEY CONCEPT

$$r_e = \frac{D_1}{P_0} + g_c = \text{dividend yield} + \text{constant growth rate}$$

The **dividend yield** here is the expected dividend divided by the current price.

Equity: implied growth rate

Given a required return, the **implied growth rate** is

$$g_c = r_e - \frac{D_1}{P_0} = \text{required return} - \text{dividend yield}$$

When the required return is the unknown and the dividend given is the one just paid (D_0), first convert it: $D_1 = D_0(1 + g_c)$.

Given	Solve for	Formula
P_0, D_1, g_c	Required return	$D_1/P_0 + g_c$

Given	Solve for	Formula
P_0, D_1, r_e	Implied growth	$r_e - D_1/P_0$
r_e, g_c, P_0	Dividend yield	$r_e - g_c$

Example. A share priced at \$50.00 is expected to pay a \$2.00 dividend next year, a dividend yield of 4.0%. With 5% constant growth, the implied required return is 4.0% + 5% = 9.0%. If investors instead require 10%, the implied growth rate is 10% - 4.0% = 6.0%.

LOS 4.c — Cash flow additivity, no arbitrage and their uses

1. Cash flow additivity principle

Under the **cash flow additivity principle**, the value of a stream of cash flows equals the total obtained by discounting each cash flow separately and adding the results. Streams can be added together (combining cash flows that fall on the same date) or split into pieces in any way, and the PVs of the pieces add up to the PV of the whole.

Example. At 8%, payments of \$200, \$200, \$500 and \$200 at the ends of years 1–4 are worth the same as a 4-year annuity of \$200 (PV = 662.43) plus a single \$300 at year 3 (PV = 238.15), a total of \$900.58. This is **replication**: an annuity plus a zero-coupon bond recreates the uneven stream. Splitting off an annuity equal to the smallest (or any) level amount leaves a residual stream in the same time order. The order matters, because earlier cash flows are worth more.

To compare two projects, subtract one stream from the other date by date. If the difference is $+X$ earlier and $-X$ later, the project with the earlier cash flow has the higher PV at any positive discount rate, with no NPV calculation needed.

2. No-arbitrage principle (law of one price)

Additivity underpins the **no-arbitrage principle** (the **law of one price**): two assets whose future cash flows match in every possible state must sell for the same price now. When their prices diverge, traders buy the cheaper asset and sell the dearer one, and that trading pushes the two prices back together. Such gaps close fast, so an arbitrageur must act at once.

The principle is applied below to derive forward interest rates and forward exchange rates and to value options.

3. Forward interest rates

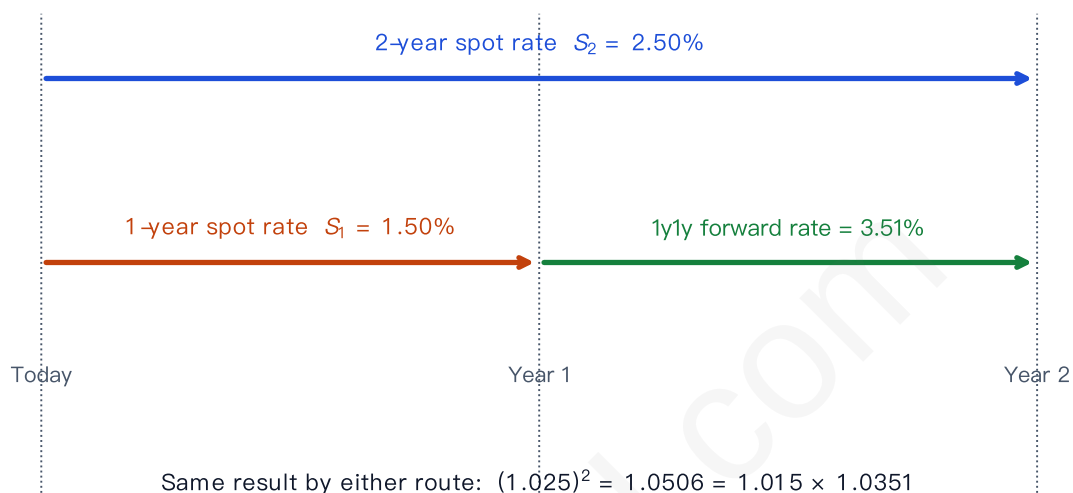
A **spot rate** S_n is the annualized rate for a loan made today for n years. A **forward rate** is for a loan made at a future date: **1y1y** is a 1-year rate one year from now, **2y1y** a 1-year rate two years from now, **3y2y** a 2-year rate three years from now. Borrowing or investing over the same horizon by any route must cost the same:

KEY CONCEPT

$$(1 + S_2)^2 = (1 + S_1)(1 + 1y1y) \Rightarrow 1y1y = \frac{(1 + S_2)^2}{1 + S_1} - 1$$

$$(1 + S_3)^3 = (1 + S_1)(1 + 1y1y)(1 + 2y1y)$$

A forward rate solved from spot rates in this way is an **implied forward rate**. *Example.* With $S_1 = 1.5\%$ and $S_2 = 2.5\%$, $1y1y = 1.025^2/1.015 - 1 = 3.51\%$. With an upward-sloping spot curve, the forward rate lies above both spot rates. An investor is indifferent between the 2-year bond and rolling 1-year bonds at S_1 then 1y1y.



Two ways to invest for two years: a 2-year spot bond, or a 1-year spot bond followed by the 1y1y forward rate

If an investor could instead lock in 4.00% for the second year, the two routes would no longer cost the same. Borrowing \$1,000,000 for two years at S_2 means repaying $1,000,000 \times 1.025^2 = \$1,050,625$. Lending the same amount for one year at 1.5% and then for one more year at the locked-in 4.00% returns $1,000,000 \times 1.015 \times 1.04 = \$1,055,600$. The \$4,975 gain at year 2 needs none of the investor's own money and carries no risk, so it is an arbitrage profit. Traders exploiting it push the forward rate back to 3.51%.

Longer forward rates follow the same rule. For the 2-year rate one year from now, $(1 + S_3)^3 = (1 + S_1)(1 + 1y2y)^2$.

4. Forward exchange rates

An **exchange rate** states how many units of one currency are needed to buy one unit of another. A rate quoted as A/B is the price of one unit of B (the **base currency**) in units of A (the **price currency**). No arbitrage links forward and spot rates to the two riskless interest rates for the forward's horizon. If the forward rate were out of line, a trader could borrow one currency, convert it at the spot rate, invest it at the other currency's riskless rate and sell the proceeds forward, locking in a riskless profit. Use one of two equivalent forms, depending on how the rates are quoted:

- With annual rates and a horizon of T years:

$$\frac{\text{Forward}}{\text{Spot}} = \frac{(1 + r_{\text{price}})^T}{(1 + r_{\text{base}})^T}$$

- With rates already stated for the forward's period (e.g., a 90-day money-market rate converted as $r \times 90/360$), no exponent is applied:

$$\frac{\text{Forward}}{\text{Spot}} = \frac{1 + r_{\text{price, period}}}{1 + r_{\text{base, period}}}$$

The percentage difference between forward and spot is approximately the interest rate differential. *Example.* For a one-year forward with spot at 1.1000 USD/EUR, a one-year USD rate of 4% and a one-year EUR rate of 2%, the forward rate is $1.1000 \times 1.04/1.02 = 1.1216$, about 1.96% above spot, close to the 2% differential. The currency with the higher interest rate trades at a forward discount.

5. Option values: one-period binomial model

An **option** gives its holder the right to buy (**call option**) or sell (**put option**) an asset at a set **exercise price**, with no obligation to do so. At expiration: call value = $\max(S_T - X, 0)$; put value = $\max(X - S_T, 0)$. At expiration, an option is **out of the money** when exercising would not pay (a call whose exercise price is above the market price, or a put whose exercise price is below it), so the holder lets it expire. An option that is **in the money** at expiration is exercised.

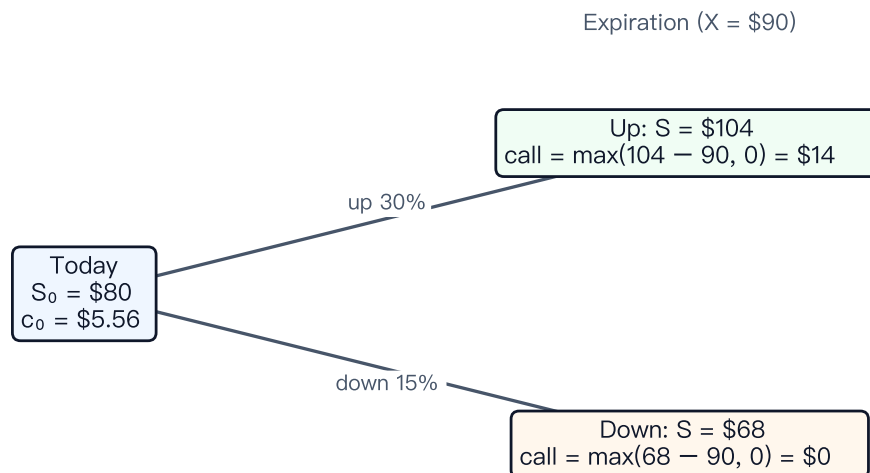
In a **binomial model**, the underlying moves to one of two prices over the period. A one-period model needs the underlying's value today, the exercise price (the option expires at the end of the period), the returns in an up move and a down move, and the risk-free rate for the period. Write one call and buy h shares so that the portfolio's value is the same after an up or down move (risk free):

$$h = \frac{c_u - c_d}{S_u - S_d}$$

This h is the **hedge ratio**: the number of shares bought per option written.

The riskless portfolio must earn the **risk-free rate**, so: value today = $\frac{hS_d - c_d}{1 + R_f}$, and the option value is $c_0 = hS_0 - \text{portfolio value today}$.

Example. $S_0 = \$80$, up 30% to \$104, down 15% to \$68, $X = \$90$, $R_f = 3.5\%$. Call payoffs: \$14 (up) and \$0 (down). $h = 14/36 = 0.3889$. Portfolio at expiration = $0.3889 \times 68 = \$26.44$ in both states. PV = $26.44/1.035 = \$25.55$. Call value = $0.3889 \times 80 - 25.55 = \5.56 .



Hedge ratio $h = (14 - 0)/(104 - 68) = 0.3889$; riskless portfolio $= hS - c = \$26.44$ in both states

Discounted at 3.5%: $\$25.55$; call value $= 0.3889 \times 80 - 25.55 = \5.56

One-period binomial tree for the call example: prices, payoffs, hedge ratio and call value

COMMON EXAM TRAPS

- Setting a price limit the wrong way round: for a zero, a return above $X\%$ requires a price below the one that gives $X\%$.
- Adding the dividend yield to the required return to get implied growth; the dividend yield is subtracted. With a 10% required return and a 4.0% dividend yield, that error gives 14.0% instead of 6.0%.
- Averaging spot rates to get a forward rate. The forward rate comes from the compounded (geometric) relationship. With $S_1 = 1.5\%$ and $S_2 = 2.5\%$, the average is 2.0%, while 1y1y is 3.51%.
- Forward FX: taking the sum or the average of the two interest rates as the forward premium or discount, or putting the base-currency rate in the numerator of the exact ratio. With spot at 1.1000 USD/EUR, putting the EUR rate on top gives $1.1000 \times 1.02/1.04 = 1.0788$ instead of 1.1216.
- Additivity: writing the residual cash flows in reverse date order, which changes their PV.
- Binomial model: weighting the payoffs 50/50, discounting the hedged portfolio at a rate other than the risk-free rate, or reporting the portfolio's PV as the option value instead of subtracting it from the share position. In the example, weighting the payoffs 50/50 gives $0.5 \times 14/1.035 = \$6.76$ and reporting the portfolio's PV gives $\$25.55$, while the call is worth $\$5.56$.

EXAM SHORTCUTS

- With an upward-sloping spot curve the implied forward rate lies above both spot rates, so a forward rate between the two spot rates can be ruled out.
- Check the direction of a forward exchange rate first: when the price currency has the higher interest rate the forward is above spot, and when it has the lower rate the forward is below spot.
- When two cash flow streams differ only by $+X$ earlier and $-X$ later, the stream with the earlier cash flow has the higher PV at any positive discount rate, with no NPV calculation needed.

BOTTOM LINE

- The implied return on a zero-coupon bond is $(FV/PV)^{1/N} - 1$; a coupon bond's YTM is found with the calculator (CPT I/Y).
- A bond's price and its YTM move in opposite directions.
- Rearranging the Gordon model gives the required return $r_e = D_1/P_0 + g_c$ and the implied growth rate $g_c = r_e - D_1/P_0$.
- Under cash flow additivity the PV of a stream equals the sum of the PVs of its pieces, so an uneven stream can be replicated with annuities and zero-coupon bonds.
- No arbitrage sets the implied forward rate $1y1y = (1 + S_2)^2/(1 + S_1) - 1$ and the forward exchange rate $\text{Forward/Spot} = (1 + r_{\text{price}})^T/(1 + r_{\text{base}})^T$.
- In a one-period binomial model the hedge ratio is $h = (c_u - c_d)/(S_u - S_d)$, the hedged portfolio earns the risk-free rate, and $c_0 = hS_0$ minus the PV of the hedged portfolio.

QUICK CHECK**Question 2**

Chiara Lombardi holds a 7-year corporate bond with a 5.3% annual coupon. Her monthly statement shows that the bond's yield to maturity rose from 5.1% last month to 5.6% this month. Over the month, the bond's price has *most likely*:

- A decreased.
- B increased.
- C remained flat.

Answer: A

Price and yield to maturity are inversely related. The yield to maturity is the discount rate that equates the bond's cash flows with its price. A higher yield means the same fixed cash flows are discounted more heavily, so the price must have fallen.

Why the other options are wrong

- **B.** A price increase would go with a lower yield to maturity, not a higher one.
- **C.** A flat price would leave the yield to maturity almost unchanged (only the passage of one month would move it slightly). It could not produce a 0.5 percentage point rise.

Key takeaway Price and YTM are inversely related: when the yield rises, the price falls.

Practice questions

Question 3

Brightwater Utilities has preferred stock outstanding that pays a fixed annual dividend of \$2.60 per share with no maturity date. An investor who requires a 7.2% rate of return should be willing to pay, per share, a maximum of *closest* to:

- A \$18.72.
- B \$33.69.
- C \$36.11.

Question 4

A small business project is expected to generate the following cash flows:

Expected year-end cash flows

End of year	Cash flow
1	\$3,000
2	\$5,000
3	\$0
4	-\$1,500

At a discount rate of 8%, the present value of this cash flow stream is *closest* to:

- A \$5,962.
- B \$7,064.
- C \$8,167.

Question 5

A stock trades at \$40. In one year it will either rise by 25% or fall by 20%. A one-year call option on the stock has an exercise price of \$44, and the risk-free rate is 4%. Using a one-period binomial model, the value of the call option today is *closest* to:

- A \$2.88.
- B \$3.08.
- C \$3.20.

Answers: 3. C · 4. A · 5. B

Why each answer is right and why each other option is wrong: credguild.com/cfa/level-1/notes/quantitative-methods/time-value-of-money/. The full bank has 44 questions on this reading.

QUANTITATIVE METHODS · READING 5

Skewness and Kurtosis

What you'll learn

LOS 5.a Calculate, interpret and evaluate measures of dispersion: range, mean absolute deviation, variance and standard deviation (population and sample).

LOS 5.b Describe, interpret and evaluate skewness and kurtosis and their effect on the mean, median, mode and tail risk of a return distribution.

LOS 5.c Calculate, interpret and evaluate covariance and correlation, including the limits of correlation (linear only, no causation, spurious correlation) and the use of scatter plots.

LOS 5.d Calculate, interpret and evaluate semivariance, target downside deviation and the coefficient of variation as measures of downside and relative risk.

MODULE 5.1

Measures of Central Tendency

This reading summarizes a return distribution with measures of central tendency and location (mean, median, mode, trimmed and winsorized means, quantiles) and of dispersion (range, mean absolute deviation, variance and standard deviation), and it interprets skewness and kurtosis. It then calculates the covariance and correlation of two variables and measures downside risk and relative dispersion with the target downside deviation and the coefficient of variation.

LOS 5.a — Central tendency and location

A **measure of central tendency** summarizes where the "center" of a dataset lies, so a single number can stand for the typical or expected value. A **measure of location** (quantiles) describes where a given observation sits within the distribution.

Population vs. sample

A **population** is every element of interest (e.g., all monthly returns a fund has ever earned). A **sample** is a subset of it (e.g., the last 36 months). Sample statistics are used to make inferences about population parameters. Of all the measures of central tendency, analysts rely on the arithmetic mean most often.

$$\text{Sample mean: } \bar{X} = \frac{\sum_{i=1}^n X_i}{n} \quad \text{Population mean: } \mu = \frac{\sum_{i=1}^N X_i}{N}$$

The three classic measures

KEY CONCEPT

Measure	Definition	Sensitive to outliers?	Notes
Arithmetic mean	Sum of the observations divided by their number (every value weighted equally)	Yes; one extreme value can move it a lot	The sample mean is an arithmetic mean

Measure	Definition	Sensitive to outliers?	Notes
Median	Middle value of the data sorted in order; with an even count, the average of the two middle values	No; it is a robust estimator (reliable with outliers or when the data come from a different distribution than assumed)	Half of the observations lie above it, half below
Mode	Most frequently occurring value	No	A dataset can have one mode, two (bimodal), several, or none

For continuous data such as returns, exact repeats are rare, so outcomes are grouped into intervals and the **modal interval**, the interval containing the most observations, is reported.

The **geometric mean** measures the compound growth rate of returns. Like the arithmetic mean, it uses every observation, so it is also pulled by extreme values.

Example. Six monthly returns: 4%, -1%, 7%, 4%, 2%, 20%.

- Mean = $36/6 = 6.0\%$.
- Sorted: -1, 2, 4, 4, 7, 20, so the median = $(4 + 4)/2 = 4.0\%$.
- Mode = 4% (appears twice).

The single 20% month pulls the mean above both the median and the mode. Replace 20% with 9% and the mean falls to 4.17%, while the median and mode stay at 4%.

Dealing with outliers

Outliers are unusually large or small observations. They can come from measurement, data entry or sampling errors, or simply from the real variability of the data. If they are errors, they distort the mean. If the data are measured correctly, keeping every value may be best, because deleting genuine observations throws away information about the distribution.

- **Trimmed mean:** drop a set percentage of the observations at the extremes, split equally between the two ends, then average what remains. Trimming identifies the outliers objectively but may remove useful information. A 4% trimmed mean drops the lowest 2% and the highest 2%. The denominator is the number of observations *remaining*.
- **Winsorized mean:** keep every observation but *replace* the extremes with percentile values. For a 96% winsorized mean, values below the 2nd percentile are set equal to the 2nd percentile and values above the 98th percentile are set equal to the 98th percentile; then take the mean of all n values.

KEY CONCEPT

	Trimmed mean	Winsorized mean
Extreme values	Removed	Replaced by percentile values
Number of values averaged	Fewer than n	All n
Information kept	Less	More

With only a small number of outliers, the winsorized mean is a better measure of the center than either the arithmetic mean or a trimmed mean. Both adjustments move the estimate *away* from the side where the extremes lie: a few huge positive values make the trimmed/winsorized mean lower than the arithmetic mean; a few huge negative values make it higher.

Quantiles

A **quantile** is a cut-off value: a given share of the observations lies at or below it. The median is the 50th percentile, which is also the second quartile and the fifth decile.

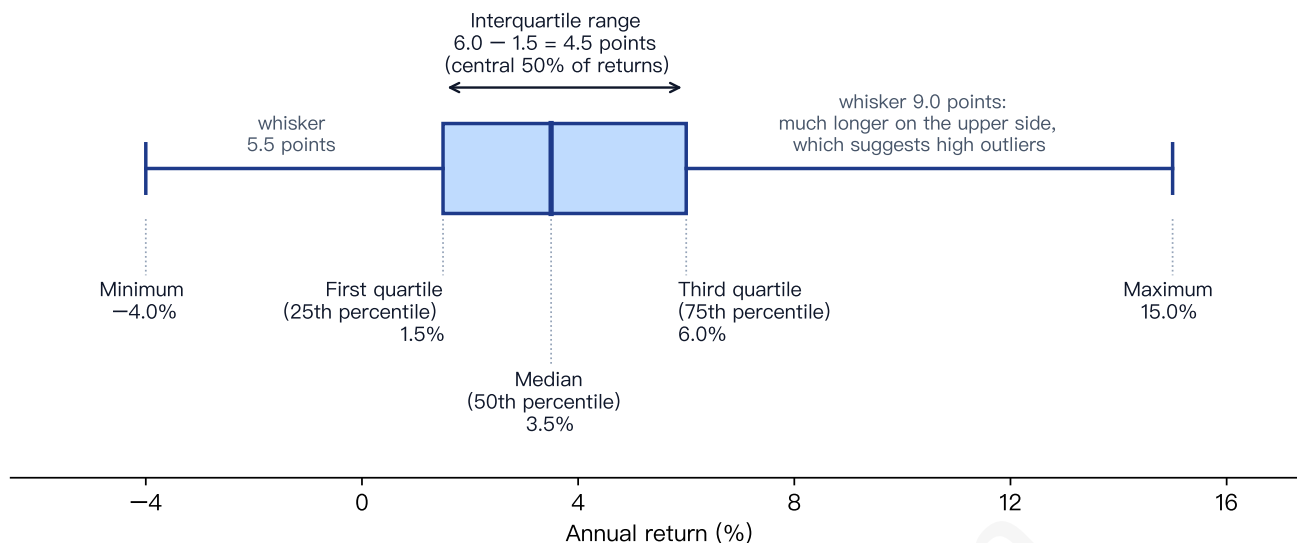
KEY CONCEPT

Quantile	Divides the data into	Example
Quartile	4 parts	3rd quartile = 75th percentile
Quintile	5 parts	1st quintile = 20th percentile
Decile	10 parts	9th decile = 90th percentile
Percentile	100 parts	An observation at the 70th percentile has 70% of the observations below it

Because each quantile is named by the share of observations at or below it, the positions always run in the same order: first decile (10th percentile), first quintile (20th), first quartile (25th), median (50th). A lower position can equal but never exceed a higher one.

The **interquartile range** (IQR) runs from the first quartile (25th percentile) to the third quartile (75th percentile), so its width is $Q_3 - Q_1$. It contains the central 50% of the data.

A **box and whisker plot** draws a box that spans the interquartile range, a line inside the box marks the median, and the whiskers extend to the smallest and largest observations (the full range). A whisker much longer on one side hints at outliers on that side.



Reading a box and whisker plot: annual returns with a long upper whisker

Stationarity

Statistical inference works best when data are **stationary**: the mean and variance of the underlying distribution do not change over time. If they do change, the series is **nonstationary** and forecasts or tests based on it are less reliable. Price or index *levels* that trend upward are typically nonstationary; their periodic percentage *changes* are much more likely to be stationary.

COMMON EXAM TRAPS

- Calling the middle value of ordered data the mean. It is the median; with an even count, average the two middle values.
- Assuming that a value equal to the median also equals the mean. The two may be close without being equal, so compute the mean before choosing.
- Expecting the median to move when only the largest (or smallest) observation changes. As long as that value keeps its rank, only the mean moves.
- Removing $X\%$ from each end for an $X\%$ trimmed mean. The correct cut is $X/2\%$ from *each* end.
- Treating a winsorized mean as if the extremes were dropped (the count stays at n), or reading a percentile as the share of observations *above* the value.
- Linking stationarity to the mean equalling the median or the geometric mean. Stationarity concerns a stable mean and variance over time.

EXAM SHORTCUTS

- Changing only the largest or the smallest observation moves the mean but not the median, as long as that value keeps its rank.
- Quantile positions run in a fixed order (first decile, first quintile, first quartile, median), so an answer that puts a lower quantile above a higher one can be ruled out.
- A few huge positive values make the trimmed and winsorized means lower than the arithmetic mean, and a few huge negative values make them higher.

BOTTOM LINE

- The arithmetic mean weights every observation equally and is sensitive to outliers, while the median, the middle value of sorted data (the average of the two middle values with an even count), is a robust estimator.
- The mode is the most frequent value; a dataset can have one mode, two (bimodal), several or none, and for continuous data the modal interval is reported.
- A trimmed mean drops a set percentage of the extreme observations, split equally between the two ends, and averages the rest; a winsorized mean replaces the extremes with percentile values and averages all n observations.
- When there are only a few outliers, a winsorized mean measures the center better than the arithmetic mean or a trimmed mean does.
- A quantile is a cut-off value with a given share of the observations at or below it, and the interquartile range $Q_3 - Q_1$ contains the central 50% of the data.
- Stationary data have a mean and variance that do not change over time; trending price levels are typically nonstationary, while their periodic percentage changes are much more likely to be stationary.

QUICK CHECK**Question 1**

The table below shows one-year returns for seven equity funds.

One-year returns of seven equity funds

Fund	Return (%)
Harbor Growth	5
Juniper Value	11
Lakeview Income	10
Meridian Core	11
Northgate Select	15
Oakmont Small Cap	3
Pinecrest Global	17

Which of the following statements about these returns is *most accurate*?

- A The mean is equal to the median.
- B The mean is equal to the mode.
- C The median is equal to the mode.

Answer: C

Sorted, the returns are 3, 5, 10, 11, 11, 15, 17. The middle (fourth) value is 11%, and 11% is also the only value that appears twice, so the median and the mode are both 11%. The mean is lower, at about 10.29%.

Sorted: 3, 5, 10, 11, 11, 15, 17. The fourth value is the median, 11%.

Mode = 11% (appears twice).

$$\bar{X} = \frac{5 + 11 + 10 + 11 + 15 + 3 + 17}{7} = \frac{72}{7} = 10.29\%$$

Why the other options are wrong

- **A.** The mean (10.29%) is below the median (11%) because the low returns of 3% and 5% pull the average down.
- **B.** The mean (10.29%) is not equal to the mode (11%).

Key takeaway Compute all three measures explicitly; the mean is rarely exactly equal to an observed value.

MODULE 5.2

Dispersion, Skewness, and Kurtosis**LOS 5.a — Measures of dispersion**

Dispersion is the variability of observations around their central tendency. In investments, central tendency measures reward and dispersion measures risk.

Range and mean absolute deviation

$$\text{range} = \text{maximum value} - \text{minimum value}$$

The **range** is simple, but because it uses only the two most extreme observations it says nothing about how the values between them are spread.

The **mean absolute deviation** (MAD) averages how far each observation lies from the arithmetic mean, ignoring the sign:

$$\text{MAD} = \frac{\sum_{i=1}^n |X_i - \bar{X}|}{n}$$

Absolute values are needed because the plain deviations from the mean always sum to zero. A MAD of 5% means that, on average, an observation lies 5 percentage points from the mean. The **median absolute deviation**, $\text{Med}(|X_i - \text{Med}(X)|)$, is a more robust alternative.

Variance and standard deviation

Variance averages the *squared* deviations from the mean; squaring removes the signs but magnifies outliers, so variance is sensitive to them. For asset returns, a higher variance indicates greater risk. Taking the positive square root of the variance gives the **standard deviation**, which is expressed in the same units as the data

(variance is in squared units, e.g. "percent squared").

KEY CONCEPT

	Population	Sample
Variance	$\sigma^2 = \frac{\sum_{i=1}^N (X_i - \mu)^2}{N}$	$s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}$
Standard deviation	$\sigma = \sqrt{\sigma^2}$	$s = \sqrt{s^2}$

Using $n - 1$ rather than n in the **sample variance** is called **Bessel's correction**. Dividing by n would systematically underestimate the population variance (especially in small samples), making it a **biased estimator**. Dividing by $n - 1$ makes s^2 an unbiased estimator of σ^2 . Exam convention: the sample standard deviation s is also described as an unbiased estimator of σ . Current practice: s is the standard estimate of σ , but the square root of an unbiased variance estimate is slightly biased downward; the effect is small except in very small samples.

Example. A sample of four annual returns: 10%, 4%, -2%, 12%.

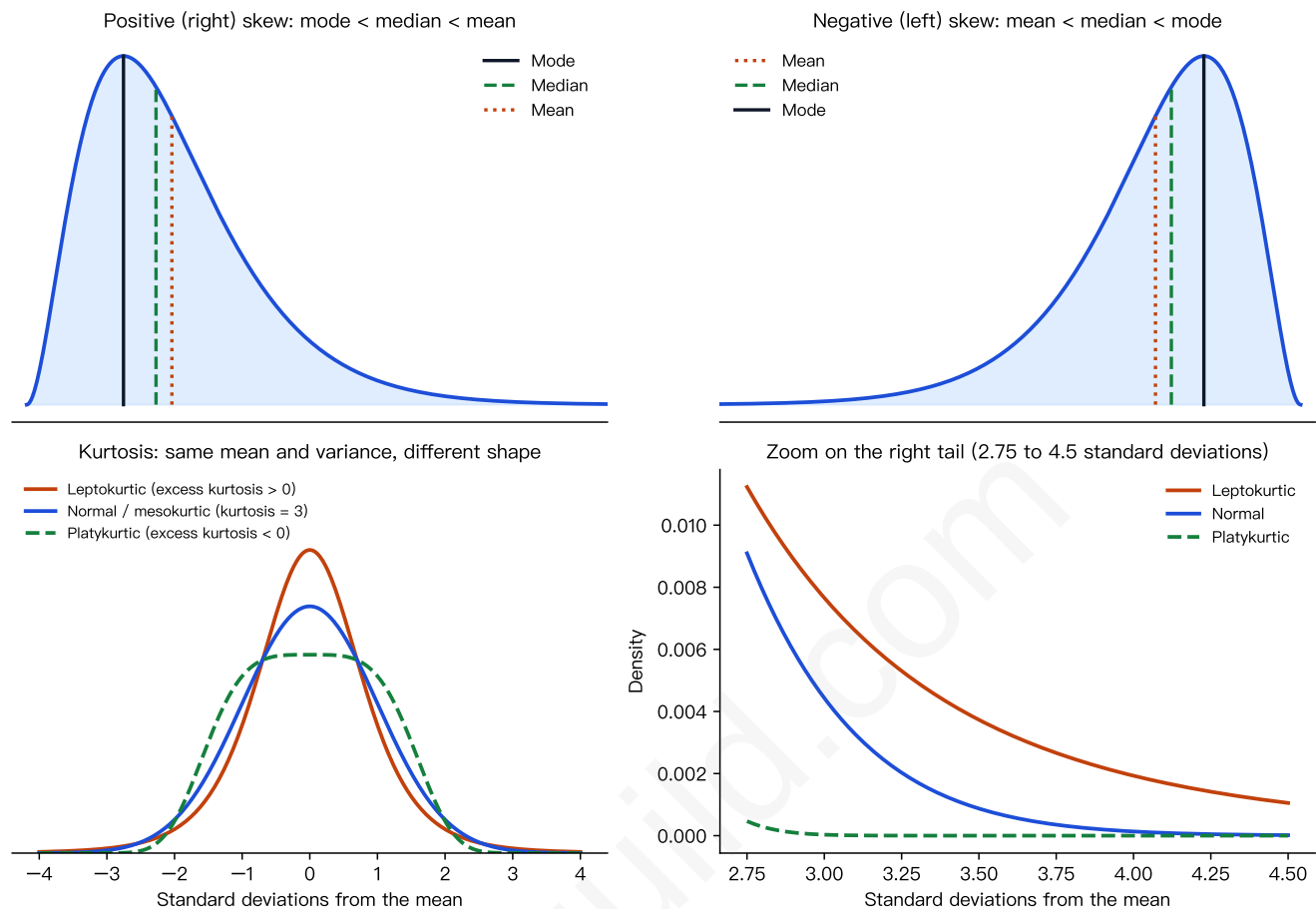
- Mean = $24/4 = 6\%$. Deviations: 4, -2, -8, 6 (sum = 0).
- MAD = $(4 + 2 + 8 + 6)/4 = 5.0\%$.
- Sum of squared deviations = $16 + 4 + 64 + 36 = 120$.
- Sample variance $s^2 = 120/3 = 40$ ($\%^2$); sample standard deviation $s = \sqrt{40} = 6.32\%$.
- Had these been the whole population: $\sigma^2 = 120/4 = 30$, $\sigma = 5.48\%$.
- Range = $12 - (-2) = 14$ percentage points.

Calculator (TI BA II Plus): [2ND][DATA], enter X01 = 10, X02 = 4, X03 = -2, X04 = 12; [2ND][STAT] (1-V) and scroll: $\bar{x} = 6$, $S_x = 6.32$ (sample), $\sigma_x = 5.48$ (population).

Scaling with time

Under the **root of time law**, variance grows linearly with the length of the period when returns are independent: if daily variance is σ^2 , the variance over T days is $T\sigma^2$, so the standard deviation is $\sigma\sqrt{T}$. In practice, long-horizon variance is often lower than this rule predicts.

LOS 5.b — Skewness and kurtosis



Skewness and kurtosis: how shape shifts the mean, median and mode, and how tails compare with a normal curve

Skewness

A **symmetrical** distribution looks the same on both sides of its mean; its mean, median and mode are equal. With symmetry, a loss interval (e.g., -6% to -4% around a zero mean) occurs as often as the matching gain interval (+4% to +6%). **Skewness** measures the lack of symmetry.

KEY CONCEPT

Shape	Long tail	Order of central measures	Sample skewness
Symmetrical	none	mean = median = mode	0
Positively skewed (skewed right)	right (large positive outliers)	mode < median < mean	> 0
Negatively skewed (skewed left)	left (large negative outliers)	mean < median < mode	< 0

The mean is pulled toward the long tail because it is the measure most affected by outliers. In a skewed unimodal distribution, the median lies between the mode and the mean. For returns, positive skew means that more than half of the periods earn less than the mean (the median is below it), while a few periods bring very large gains. Negative skew is the mirror image: most periods earn more than the mean, and occasional large losses pull the mean down. Sample skewness averages the cubed standardized deviations, where s is the sample standard deviation (the LOS asks for interpretation, not calculation):

$$\text{skewness} = \frac{1}{n} \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{s} \right)^3$$

Right-skewed data have positive sample skewness, since upside deviations from the mean are bigger on average. A quick way to judge the direction of skew is to compare the distance from the mode (peak) to each end of the range: the longer side is the tail.

Kurtosis

Kurtosis measures how peaked a distribution is and how thick its tails are, relative to a normal distribution (kurtosis = 3). **Excess kurtosis** = kurtosis − 3. Kurtosis uses the fourth power of the standardized deviations:

$$\text{kurtosis} = \frac{1}{n} \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{s} \right)^4 \quad \text{excess kurtosis} = \text{kurtosis} - 3$$

KEY CONCEPT

Term	Kurtosis	Excess kurtosis	Shape vs. normal
Leptokurtic	> 3	> 0 (positive)	More peaked, fatter tails: more small deviations <i>and</i> more very large deviations from the mean
Mesokurtic	= 3	= 0	Same as normal
Platykurtic	< 3	< 0 (negative)	Flatter (less peaked), thinner tails

Fat tails matter for risk: a leptokurtic return distribution produces extreme outcomes (both large gains and large losses) more often than a normal model predicts. Actual asset returns tend to show both skewness and excess kurtosis, so risk models that assume normality understate the chance of extreme losses. In general, more negative skew and greater excess kurtosis both mean greater risk.

COMMON EXAM TRAPS

- Dividing by n when the data are a sample, or by $n - 1$ when they are the whole population. For the four sample returns, dividing by 4 gives a standard deviation of 5.48% instead of 6.32%.
- Dividing the MAD by $n - 1$, or averaging the signed deviations, which always sum to zero. In the example, $20/3 = 6.67\%$ instead of 5.0%.
- Stopping at the variance when the standard deviation (its square root) is asked for. In the example, that answer is 40 (%²) instead of 6.32%.

- Comparing *kurtosis* with 0 instead of 3 (only *excess kurtosis* is compared with 0). A kurtosis of 2.5 is platykurtic even though the number is positive.
- Putting the mean on the wrong side: it is the *largest* of the three measures under positive skew and the *smallest* under negative skew.
- Reading one very low outlier (e.g. a single very poor score) as positive skew. It creates negative skew and pulls the mean below the median.

EXAM SHORTCUTS

- In a skewed unimodal distribution the median lies between the mode and the mean, and the mean lies on the side of the long tail, so the direction of skew fixes the order of all three measures.

BOTTOM LINE

- The range, maximum minus minimum, uses only the two extreme observations, while the mean absolute deviation $\sum |X_i - \bar{X}|/n$ averages the unsigned deviations from the mean.
- Sample variance divides the sum of squared deviations from the mean by $n - 1$ (Bessel's correction), which makes s^2 an unbiased estimator of σ^2 ; population variance divides by N .
- Taking the positive square root of the variance gives the standard deviation, which is in the same units as the data, while the variance is in squared units.
- Under the root of time law, with independent returns the variance over T periods is $T\sigma^2$ and the standard deviation is $\sigma\sqrt{T}$, although long-horizon variance is often lower in practice.
- A positively skewed distribution has a long right tail with mode < median < mean, a negatively skewed one has a long left tail with mean < median < mode, and a symmetrical one has all three equal.
- A normal distribution has kurtosis 3; excess kurtosis is kurtosis – 3, a leptokurtic distribution (above 3) has fatter tails, and more negative skew and greater excess kurtosis both mean greater risk.

QUICK CHECK

Question 2

A histogram of a biotech fund's daily returns shows a long tail stretching to the right, created by a few very large gains. For this unimodal distribution, which relationship among the measures of central tendency is *most likely*?

- A mean < median < mode
- B mean = median = mode
- C mean > median > mode

Answer: C

A long right tail means the distribution is positively skewed. The large positive outliers pull the mean upward the most, the median less, and the mode (the peak) not at all, so mean > median > mode.

Why the other options are wrong

- **A.** This ordering describes a negatively skewed distribution with a long left tail.
- **B.** Equality of the three measures describes a symmetrical distribution, which has no long tail.

Key takeaway Positive (right) skew: mode < median < mean.

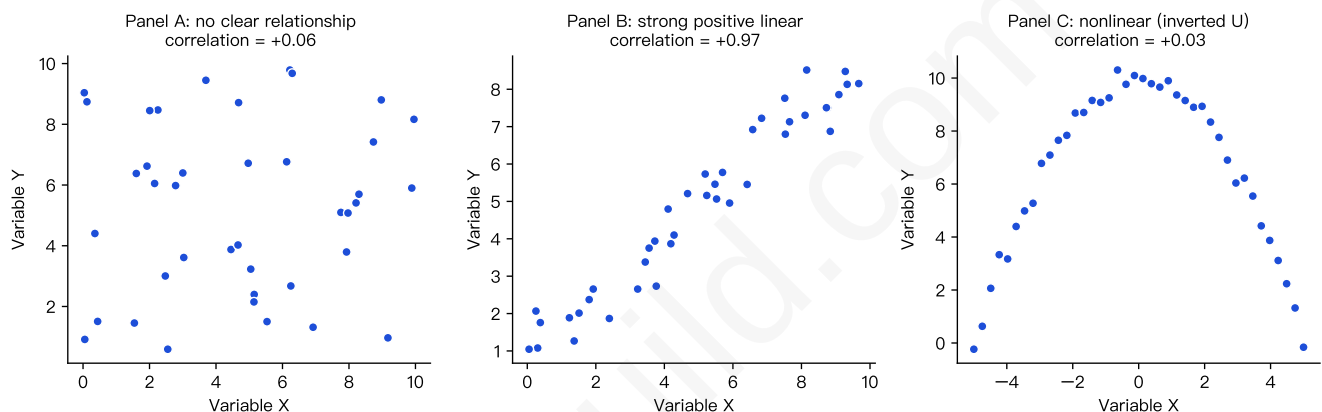
MODULE 5.3

Covariance, Correlation, and Alternative Measures of Dispersion

LOS 5.c — Covariance and correlation

Scatter plots

A **scatter plot** graphs paired observations of two variables, one on each axis. It shows at a glance whether the variables move together, whether the relationship is roughly a straight line, and whether there is a **nonlinear** relationship that a correlation coefficient would miss.



Scatter plots: no relationship, a strong linear relationship, and a nonlinear relationship with near-zero correlation

Covariance

Covariance measures how two variables move together:

$$\text{Cov}_{XY} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n - 1} \quad (\text{sample})$$

(The population version uses the population means and divides by N .) A positive covariance means the variables tend to be above (or below) their means at the same time; a negative covariance means they tend to move in opposite directions.

Example. Four annual returns on fund X are 6%, -2%, 10% and 2%, and on fund Y 5%, 1%, 9% and 1%. Both means are 4%.

1. Deviations from the means (percentage points): fund X gives 2, -6, 6, -2; fund Y gives 1, -3, 5, -3.
2. Multiply each pair and add: $2 + 18 + 30 + 6 = 56$.
3. Divide by $n - 1 = 3$: $\text{Cov}_{XY} = 56/3 = 18.67$ (%²).

Every cross-product is positive, so the two funds were above or below their means in the same years.

Covariance is hard to interpret: its size depends on the units of the data, and its units are the product of the two variables' units (e.g. percent squared). A covariance of 30 shows only the *direction* of co-movement; it says nothing about how strong it is. Covariance has no upper or lower bound. The covariance of a variable with itself is its variance: $\text{Cov}_{XX} = s_X^2$.

Correlation

The **correlation coefficient** (correlation) standardizes covariance by the two standard deviations:

KEY CONCEPT

$$\rho_{XY} = \frac{\text{Cov}_{XY}}{s_X s_Y} \iff \text{Cov}_{XY} = \rho_{XY} s_X s_Y$$

Here s_X and s_Y are the sample standard deviations. The population version uses σ_X and σ_Y in the same way.

KEY CONCEPT

Value	Interpretation
$\rho = +1$	Perfect positive correlation: deviations from the mean move in exact proportion, same direction; all points lie on one upward-sloping straight line, however steep or flat
$0 < \rho < 1$	Positive linear association; strength grows as ρ approaches 1
$\rho = 0$	No linear relationship
$-1 < \rho < 0$	Negative linear association
$\rho = -1$	Perfect negative correlation: all points lie on one downward-sloping straight line

Correlation has no units and always lies between -1 and $+1$. It is *not* a probability: $\rho = -0.6$ does not mean "a 60% chance the variables move in opposite directions." Because the covariance is divided by the two standard deviations, the correlation moves further from zero when the standard deviations shrink and the covariance stays the same.

Example. Returns on two funds have variances of 0.0036 and 0.0100 and a covariance of 0.0021. Standard deviations are $\sqrt{0.0036} = 0.06$ and $\sqrt{0.0100} = 0.10$, so

$$\rho = \frac{0.0021}{0.06 \times 0.10} = 0.35$$

This is a positive but modest linear association. The variances must be converted to standard deviations first.

For the four years of fund returns in the covariance example, $s_X = \sqrt{80/3} = 5.16\%$ and $s_Y = \sqrt{44/3} = 3.83\%$, so $\rho = 18.67/(5.16 \times 3.83) = 0.94$, a strong positive linear association.

Limits of correlation

- **Correlation is not causation.** A strong correlation does not show that one variable drives the other, or which way any causality runs. Say the variables show a positive (or negative) **association** and investigate cause separately.
- **Outliers** can create or destroy a correlation; check whether removing them changes the result materially and whether they are information or noise.
- **Spurious correlation** is correlation that arises by chance, or because both variables are related to a third variable (e.g., two series that both trend upward with inflation or population), with no causal link between them.

LOS 5.d — Downside risk and relative dispersion

Semivariance and target downside deviation

Variance and standard deviation treat deviations above and below the mean alike. In some situations it is more useful to count only outcomes below the mean or below a chosen target, which measures **downside risk**.

- **Semivariance:** a variance built only from the outcomes that fall *below the mean*. The squared deviations of the below-mean observations are summed and divided by the sample size:

$$\sigma_{\text{semi}}^2 = \frac{\sum_{X_i < \bar{X}} (X_i - \bar{X})^2}{n}$$

Its square root is the **semideviation**, a downside counterpart of the standard deviation.

- **Target downside deviation** (also **target semideviation**): like a standard deviation, but only outcomes *below a chosen target B* contribute; the denominator is still the full sample size minus one:

$$s_{\text{target}} = \sqrt{\frac{\sum_{X_i < B} (X_i - B)^2}{n - 1}}$$

Example. Five annual returns: 12%, -4%, 7%, 3%, 10%; target 6%. Only -4% and 3% are below target: squared shortfalls = 100 + 9 = 109. $s_{\text{target}} = \sqrt{109/4} = \sqrt{27.25} = 5.22\%$. Raising the target to 8% brings the 7% return into the count and makes the other two shortfalls larger: $\sqrt{(144 + 25 + 1)/4} = \sqrt{42.5} = 6.52\%$. A higher target never lowers the target downside deviation.

Coefficient of variation

Comparing standard deviations of assets with very different mean returns can mislead. A **relative dispersion** measure scales risk by the mean. The **coefficient of variation** (CV) is:

KEY CONCEPT

$$CV = \frac{s}{\bar{X}} \quad \left(\text{or } \frac{\sigma}{\mu} \right)$$

It measures **risk per unit of mean return**, so a lower CV is better (less risk per unit of return). Because it is a ratio of two numbers in the same units, it is unit-free and can be quoted as a decimal or a percentage. The Sharpe ratio (Reading 8) turns the CV around: it divides the excess return over the risk-free rate by the standard deviation. With a zero risk-free rate, the CV is simply the inverse of the Sharpe ratio. When a Sharpe ratio and a risk-free rate are given, first recover $s = (\bar{X} - R_f)/\text{Sharpe ratio}$, then divide by the mean.

Example. Fund M has a mean of 5% and a standard deviation of 4%, so $CV = 0.80$. Fund N has a mean of 11% and a standard deviation of 7.7%, so $CV = 0.70$. Fund N has the higher standard deviation but the better (lower) CV.

Measure	Uses	Scale
Standard deviation	All deviations from the mean	Absolute (units of the data)
Target downside deviation	Only shortfalls below a target	Absolute
Coefficient of variation	Standard deviation $\div$ mean	Relative (per unit of return)

COMMON EXAM TRAPS

- Inverting the CV or using the variance. $CV = \text{standard deviation} \div \text{mean}$; when a variance is given, take its square root first. For fund M, inverting gives $5/4 = 1.25$ and using the variance of 16 ($\%^2$) gives 3.20, instead of 0.80.
- Subtracting a risk-free rate given in the question. The CV uses the mean itself; excess return over standard deviation is a Sharpe-type ratio, a different measure.
- Picking the asset with the lowest standard deviation as the best risk-return tradeoff, or the one with the highest standard deviation as the highest relative risk. Both are judged by the CV.
- Dividing the covariance by one variance (a slope-type ratio) or by the product of the variances, instead of by the product of the standard deviations. In the two-fund example, $0.0021/(0.0036 \times 0.0100) = 58.33$ and $0.0021/0.0036 = 0.58$, instead of 0.35.
- Reading a correlation near zero as proof that the variables are unrelated. It rules out only a linear relationship.
- Dividing the target downside deviation by the number of below-target outcomes (or that number minus one) instead of the full sample size minus one. In the 6% target example, $\sqrt{109/2} = 7.38\%$ instead of $\sqrt{109/4} = 5.22\%$.

EXAM SHORTCUTS

- With a zero risk-free rate the CV is the inverse of the Sharpe ratio, so one can be read straight from the other.
- A higher target never lowers the target downside deviation, which rules out an answer in which raising the target reduces it.

BOTTOM LINE

- Sample covariance is $\sum(X_i - \bar{X})(Y_i - \bar{Y})/(n - 1)$; its sign shows the direction of co-movement, but its size depends on the units and it has no upper or lower bound.

- Correlation $\rho_{XY} = \text{Cov}_{XY} / (s_X s_Y)$ has no units, lies between -1 and $+1$, and $\rho = 0$ means no linear relationship.
- Correlation is not causation, outliers can create or destroy a correlation, and spurious correlation arises by chance or because both variables are related to a third variable.
- A scatter plot can reveal a nonlinear relationship that a correlation coefficient would miss.
- Semivariance uses only the outcomes below the mean, and the target downside deviation uses only the outcomes below a target B but keeps the full sample size minus one in the denominator.
- The coefficient of variation $CV = s / \bar{X}$ measures risk per unit of mean return, is unit-free, and a lower CV is better.

QUICK CHECK

Question 3

The covariance between the returns on Norland Utilities and Quarry Point Materials is 21.6 (percent squared). The standard deviation of Norland's returns is 8% and that of Quarry Point's returns is 6%. The correlation coefficient of the two return series is *closest* to:

- A +0.34.
- B +0.45.
- C +0.60.

Answer: B

Correlation standardizes covariance by dividing it by the product of the two standard deviations.

$$\rho = \frac{\text{Cov}}{\sigma_N \sigma_Q} = \frac{21.6}{8 \times 6} = \frac{21.6}{48} = 0.45$$

Why the other options are wrong

- A. +0.34 divides the covariance by Norland's variance only ($21.6/64$).
- C. +0.60 divides the covariance by Quarry Point's variance only ($21.6/36$).

Key takeaway $\rho = \text{Cov} / (\sigma_1 \sigma_2)$; dividing by a single variance gives a slope-type ratio.

Practice questions

Question 4

Priya Raman plans to model the monthly percentage changes in a copper price index. Her statistical tests will be most reliable if the series is stationary, which means that:

- A its geometric mean and arithmetic mean are the same.
- B its mean and median both lie within its interquartile range.
- C the mean and variance of its distribution do not change over time.

Question 5

A multi-asset portfolio produced the returns shown below over the last four years.

Annual returns of a multi-asset portfolio (sample)

Year	Return
1	19.0%
2	7.5%
3	-9.6%
4	4.9%

Treating these returns as a sample, their coefficient of variation is *closest* to:

- A 0.46.
- B 1.87.
- C 2.16.

Answers: 4. C · 5. C

Why each answer is right and why each other option is wrong: credguild.com/cfa/level-1/notes/quantitative-methods/skewness-and-kurtosis/. The full bank has 60 questions on this reading.

QUANTITATIVE METHODS · READING 6

Lognormal Distribution

What you'll learn

LOS 6.a Compute and interpret the unconditional mean, variance, standard deviation and covariance of random variables, including from scenario and joint probability tables.

LOS 6.b Describe the shape, principal moments and typical financial uses of the discrete uniform, binomial, Poisson, continuous uniform, normal, lognormal and logistic distributions.

LOS 6.c Compute and interpret conditional expectations, variances and covariances, including from joint probability tables and probability trees.

LOS 6.d Use Bayes' formula to update a prior probability into a posterior probability when new information arrives.

MODULE 6.1

Probability Distributions and Expected Values

This reading describes random variables through their probability distributions and summarizes them with the expected value, variance and covariance. It then matches the common discrete and continuous distributions to the financial quantities they model, and shows how conditional expectations and Bayes' formula update a forecast when new information arrives.

LOS 6.a — Unconditional mean, variance and covariance of a random variable

Random variables: discrete vs. continuous

A **random variable** is a quantity whose value is not known until an outcome is observed. It is also called a *stochastic variable*.

Feature	Discrete random variable	Continuous random variable
Possible values	A countable set (e.g., the number of issuers in a 50-bond portfolio downgraded this year)	Infinitely many values inside a range (e.g., tomorrow's percentage change in the gold price)
Probability function	Probability mass function (PMF): $p(x) = P(X = x)$ can be positive	Probability density function (PDF): $P(X = x) = 0$ for every single point
Probabilities are assigned to	Individual outcomes	Intervals, e.g. $P(a \leq X \leq b)$
Can outcomes be negative?	Yes, if the countable set includes negatives	Yes, or the range may be restricted to positive values

A **probability distribution** lists the probabilities of all possible outcomes of a random variable, and these probabilities sum to 1. A **probability function** $p(x) = P(X = x)$ gives the probability that X equals a particular value. For a discrete variable, $p(x) = 0$ for a value that cannot occur and $p(x) > 0$ for one that can.

Whether a variable is discrete does not depend on its outcomes being equally likely (the uniform case) or positive. Prices quoted in cents are technically discrete. Because there are so many possible values, analysts usually treat them as continuous and speak of the probability of a range of price changes.

What makes a valid probability function

A function $p(x)$ qualifies as a probability function only if both conditions hold:

KEY CONCEPT

1. $0 \leq p(x) \leq 1$ for every possible outcome x ;
2. $\sum_x p(x) = 1$ over all possible outcomes.

A list such as 0.50, 0.40, 0.20, -0.10 sums to one but fails condition 1. Five values of 0.25 each pass condition 1 but sum to 1.25, so they fail condition 2.

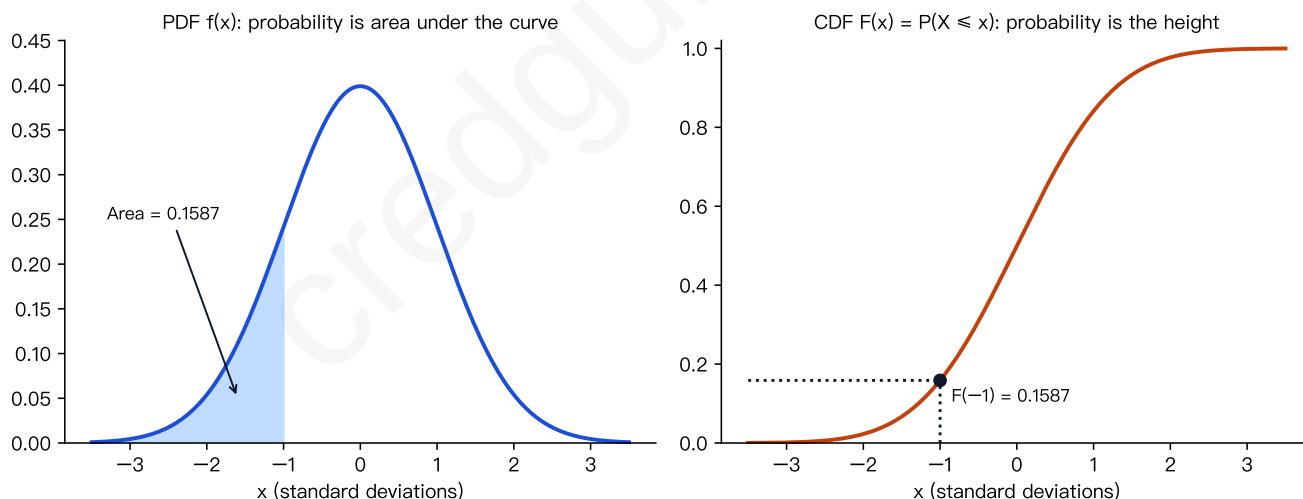
For a continuous variable, a single point carries no probability, so including or excluding the endpoints does not change the probability of an interval: $P(a \leq X \leq b) = P(a < X < b)$. The height of the PDF at a point is a density. A probability is the area under the PDF over an interval.

Cumulative distribution function (CDF)

The **cumulative distribution function (CDF)**, also called the **distribution function**, gives the probability that X takes a value less than or equal to x :

$$F(x) = P(X \leq x) \quad P(a < X \leq b) = F(b) - F(a)$$

For a discrete variable, $F(x)$ is the sum of the probabilities of all outcomes up to and including x . For the standard normal distribution (Module 6.2), $F(-1) \approx 0.1587$: the area under the PDF to the left of -1 equals the height of the CDF at -1 .



Standard normal PDF and CDF: the area left of -1 under the PDF equals the height of the CDF at -1

Example. A fund's quarterly return is continuous on $[-5\%, +15\%]$ with $F(x) = (x + 5)/20$. Then $P(0\% < R \leq 10\%) = F(10) - F(0) = 15/20 - 5/20 = 0.50$.

Unconditional expected value (mean)

The **unconditional mean** (also the **unconditional expectation** or **expected value**) weights each outcome by its probability:

$$E(X) = \sum_{i=1}^n P(x_i) x_i$$

The weights are the probabilities, and they must sum to 1, so a probability that is not stated can be found as 1 minus the sum of the others. A simple average of the outcomes weights each by $1/n$, so it is the expected value when every outcome is equally likely; with unequal probabilities it generally gives a different answer. The sample mean of Reading 5 summarizes returns already observed (an ex post measure). An expected value weights possible future outcomes by their probabilities, so it is a forecast (an ex ante measure).

Unconditional variance and standard deviation

The **unconditional variance** is the probability-weighted squared spread around the expected value, and the **unconditional standard deviation** is its square root:

KEY CONCEPT

$$\sigma^2(X) = \sum_{i=1}^n P(x_i) [x_i - E(X)]^2 \quad \sigma(X) = \sqrt{\sigma^2(X)}$$

Variance is in *squared* units (e.g., %²), so the standard deviation (back in % units) is the figure compared with returns. An equivalent shortcut is $\sigma^2 = E(X^2) - [E(X)]^2$.

Unconditional covariance

Unconditional covariance shows the direction of the linear relationship between two random variables, that is, whether they tend to be above or below their own means at the same time. Its strength is read from the correlation (last point below). The covariance is computed as:

$$\text{Cov}(X, Y) = \sum_{i=1}^n P(x_i, y_i) [x_i - E(X)] [y_i - E(Y)]$$

The shortcut form is $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$.

- Positive covariance: the variables tend to move in the same direction; negative: opposite directions.
- For scenario (state) tables, $P(x_i, y_i)$ is simply the probability of the state. In a joint probability table, cells with zero probability contribute nothing.
- Units are the product of the two variables' units (e.g., %² or decimal²), so the size is hard to interpret. Dividing by the two standard deviations gives the unit-free **correlation**: $\rho = \text{Cov}(X, Y) / (\sigma_X \sigma_Y)$, which always lies between -1 and +1.

Measure	Formula core	Units	Range
Variance	$\sum P(x - \mu)^2$	squared units	≥ 0
Covariance	$\sum P(x - \mu_X)(y - \mu_Y)$	product of units	any sign, unbounded
Correlation	$\text{Cov} / (\sigma_X \sigma_Y)$	none	-1 to +1

WORKED EXAMPLE

State	Probability	Return on P	Return on Q
Expansion	0.20	15%	4%
Normal	0.50	8%	7%
Downturn	0.30	-2%	2%

- $E(R_P) = 0.2(15) + 0.5(8) + 0.3(-2) = 6.4\%$; $E(R_Q) = 0.2(4) + 0.5(7) + 0.3(2) = 4.9\%$.
- $\sigma_P^2 = 0.2(8.6)^2 + 0.5(1.6)^2 + 0.3(-8.4)^2 = 37.24 (\%^2)$, so $\sigma_P = 6.10\%$.
- $\sigma_Q^2 = 0.2(-0.9)^2 + 0.5(2.1)^2 + 0.3(-2.9)^2 = 4.89 (\%^2)$, so $\sigma_Q = 2.21\%$.
- $\text{Cov}(R_P, R_Q) = 0.2(8.6)(-0.9) + 0.5(1.6)(2.1) + 0.3(-8.4)(-2.9) = -1.548 + 1.68 + 7.308 = 7.44 (\%^2)$.
- $\rho = 7.44 / (6.10 \times 2.21) \approx 0.55$.

Calculator. The BA II Plus DATA/STAT worksheet computes a probability-weighted mean when each outcome is entered with a frequency proportional to its probability (e.g., 20, 50 and 30). The mean is $\bar{x}$ and the population standard deviation is σ_x .

COMMON EXAM TRAPS

- Averaging outcomes equally instead of weighting by probabilities. For portfolio P in the worked example, the simple average is $(15 + 8 - 2)/3 = 7.0\%$, against an expected return of 6.4%.
- Forgetting the square root: reporting the variance when the standard deviation is asked, or mixing decimals (0.0035) with percent-squared (35.04). For portfolio P in the worked example, the variance is 37.24 ($\%^2$) and the standard deviation is 6.10%.
- Using absolute deviations instead of squared deviations, which gives a different measure from the variance and standard deviation.
- Computing covariance without the probabilities, or dividing an unweighted sum of cross-products by the number of states.
- Confusing covariance with correlation; the answer to "covariance" carries squared units and is not bounded by ± 1 .
- Believing a continuous variable has a positive probability at a single point, or that the PDF height is a probability.
- Reading a CDF value as the probability of that single outcome, or leaving the outcome itself out of $F(x) = P(X \leq x)$.

EXAM SHORTCUTS

- When the outcomes listed are mutually exclusive and exhaustive and one probability is missing, it equals 1 minus the sum of the others.
- Variance can be found as $\sigma^2 = E(X^2) - [E(X)]^2$, which avoids computing each deviation.
- For a continuous variable, including or excluding the endpoints does not change the probability of an interval.

BOTTOM LINE

- A discrete random variable has a probability mass function with positive probabilities at individual outcomes; a continuous variable has a density, zero probability at any single point and probabilities assigned to intervals.
- A probability function is valid only if every $p(x)$ lies between 0 and 1 and the probabilities sum to 1.
- The CDF is $F(x) = P(X \leq x)$, and $P(a < X \leq b) = F(b) - F(a)$.
- Expected value is $E(X) = \sum P(x_i)x_i$; variance is $\sum P(x_i)[x_i - E(X)]^2$ in squared units, and the standard deviation is its square root.
- Covariance shows only the direction of co-movement; correlation $\rho = \text{Cov}(X, Y)/(\sigma_X \sigma_Y)$ is unit-free and lies between -1 and +1.

QUICK CHECK**Question 1**

A random variable X can take only the values 0, 1, 2 and 3. Which of the following proposed functions is a valid probability function for X ?

- A** $p(0) = 0.40, p(1) = 0.35, p(2) = 0.30, p(3) = -0.05$
- B** $p(x) = 0.30$ for each of the four values
- C** $p(x) = (x + 1)/10$

Answer: C

A probability function must satisfy two conditions: every probability lies between 0 and 1 inclusive, and the probabilities over all possible outcomes add up to exactly 1. Only $p(x) = (x + 1)/10$ meets both: it gives 0.1, 0.2, 0.3 and 0.4, each within $[0, 1]$, summing to 1.0.

- Set with $p(3) = -0.05$: sum = $0.40 + 0.35 + 0.30 - 0.05 = 1.00$, but $p(3) = -0.05 < 0$, so it fails condition 1.
- Constant $p(x) = 0.30$: each value is within $[0, 1]$, but the sum is $4 \times 0.30 = 1.20 \neq 1$, so it fails condition 2.
- $p(x) = (x + 1)/10$: $p = 0.1, 0.2, 0.3, 0.4$; all in $[0, 1]$ and $\sum p = 1.0$, so it is valid.

Why the other options are wrong

- **A.** The probabilities sum to one, but a negative probability is impossible; satisfying only the 'sum to one' condition is not enough.
- **B.** Each probability is between 0 and 1, but together they sum to 1.20. A probability function must exhaust exactly 100% of the probability; an equal-probability function over four outcomes would need 0.25 each.

Key takeaway Check both conditions every time: $0 \leq p(x) \leq 1$ for each x , and $\sum p(x) = 1$.

MODULE 6.2

Discrete and Continuous Probability Distributions

LOS 6.b — Moments and uses of the key distributions in finance

The four principal moments

Moment	What it describes	Benchmark (normal distribution)
1. Mean	Central location	—
2. Variance	Dispersion around the mean	—
3. Skewness	Asymmetry; positive = long right tail	0 (symmetric)
4. Kurtosis	Weight of the tails relative to the center	3 (excess kurtosis = kurtosis – 3 = 0)

The most useful features of each distribution are what it models and its shape: bounded or unbounded, symmetric or skewed, thin or fat tails. The moment formulas below are for reference.

Distribution	Mean	Variance	Skewness	Excess kurtosis
Discrete uniform (n outcomes, a to b)	$(a + b)/2$	$[(b - a + 1)^2 - 1]/12$	0	$-\frac{6(n^2 + 1)}{5(n^2 - 1)}$
Binomial	np	$np(1 - p)$	$\frac{1 - 2p}{\sqrt{np(1 - p)}}$	$\frac{1 - 6p(1 - p)}{np(1 - p)}$
Poisson	λ	λ	$1/\sqrt{\lambda}$	$1/\lambda$
Continuous uniform	$(a + b)/2$	$(b - a)^2/12$	0	-1.2
Normal	μ	σ^2	0	0
Lognormal	$e^{\mu + \sigma^2/2}$	$(e^{\sigma^2} - 1)e^{2\mu + \sigma^2}$	> 0	> 0
Logistic	μ	$s^2\pi^2/3$	0	1.2

Discrete distributions

Discrete uniform. A finite set of n outcomes, each with probability $1/n$ (e.g., the integers 1 to 8, each with probability 0.125); the probability of k outcomes in a range is k/n , and the CDF at the k th outcome is k/n . For a discrete uniform variable on the consecutive integers $a, a + 1, \dots, b$: mean = $(a + b)/2$ and variance = $[(b - a + 1)^2 - 1]/12$. For a general equally likely set (e.g., $\{1, 3, 5\}$), compute the arithmetic mean and the population variance directly: here the mean is 3 and the variance is $(4 + 0 + 4)/3 = 8/3$. The consecutive-integer formula would wrongly give 2.

Binomial. Counts the number of successes in a fixed number n of independent trials, each with the same success probability p and only two possible results (success/failure). A binomial variable with a single trial is a **Bernoulli distribution**.

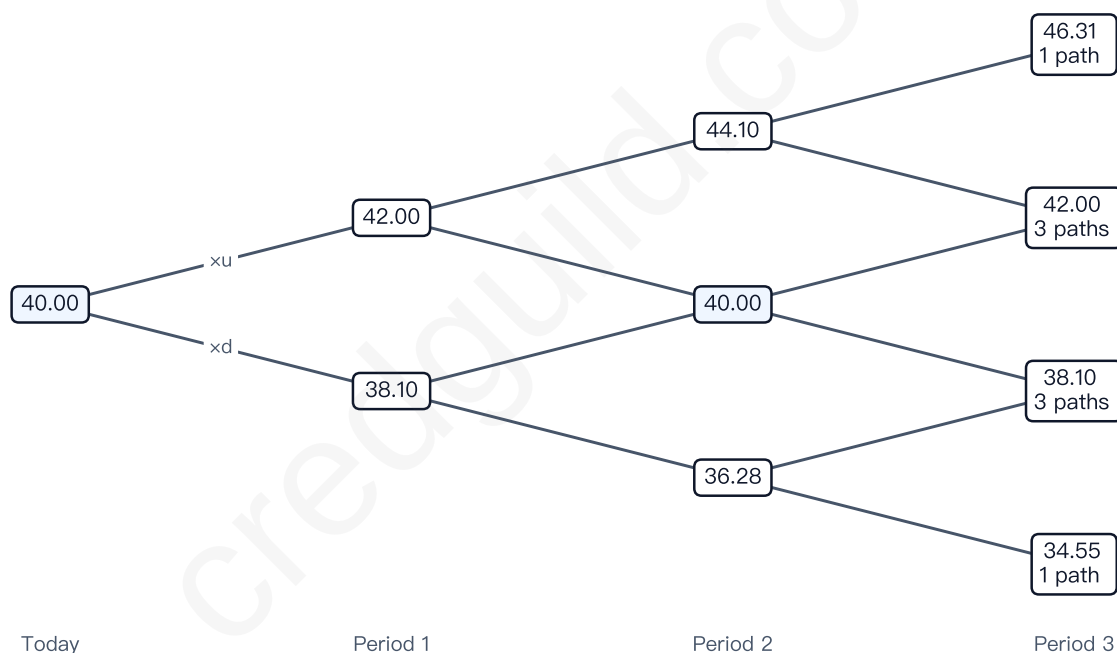
KEY CONCEPT

$$E(X) = np \quad \sigma^2(X) = np(1 - p)$$

Example. 16 independent new-issue auctions, each with a 0.75 chance of being fully covered: expected fully covered auctions = $16 \times 0.75 = 12$; variance = $16 \times 0.75 \times 0.25 = 3$. The expected number of successes need not be a possible outcome (10 trials with $p = 0.43$ give 4.3); it is the average number of successes over many repetitions.

Binomial models are also used for asset prices: at each step the price is multiplied by an **up factor** u or a **down factor** d , so after i up moves and j down moves the price is $S_0 u^i d^j$. The possible paths are drawn as a **probability tree** (or **decision tree**). With $u = 1.05$ and $d = 1/1.05$, a price of 40 after one up and one down move returns to 40, whatever the order.

$u = 1.05, d = 1/1.05$: an up move followed by a down move returns the price to 40.00



Three-period binomial price tree with $u = 1.05$ and $d = 1/1.05$: the tree recombines

Poisson. Counts how many times an event occurs in a fixed interval when events arrive independently at a known, constant average rate λ . It is the usual model for rare events such as defaults per year, trading halts per month or operational losses per quarter.

$$P(k \text{ events in interval } t) = \frac{(\lambda t)^k e^{-\lambda t}}{k!} \quad \text{mean} = \text{variance} = \lambda \text{ (for one unit of time, } t = 1)$$

Example. With $\lambda = 2$ outages per year, $P(\text{no outage}) = e^{-2} = 0.135$, so the probability of at least one outage is $1 - e^{-2} = 0.865$. For a pool of n independent credits each with default probability p , $\lambda = np$.

The Poisson is the limit of a binomial as $n \rightarrow \infty$ and $p \rightarrow 0$ with np fixed.

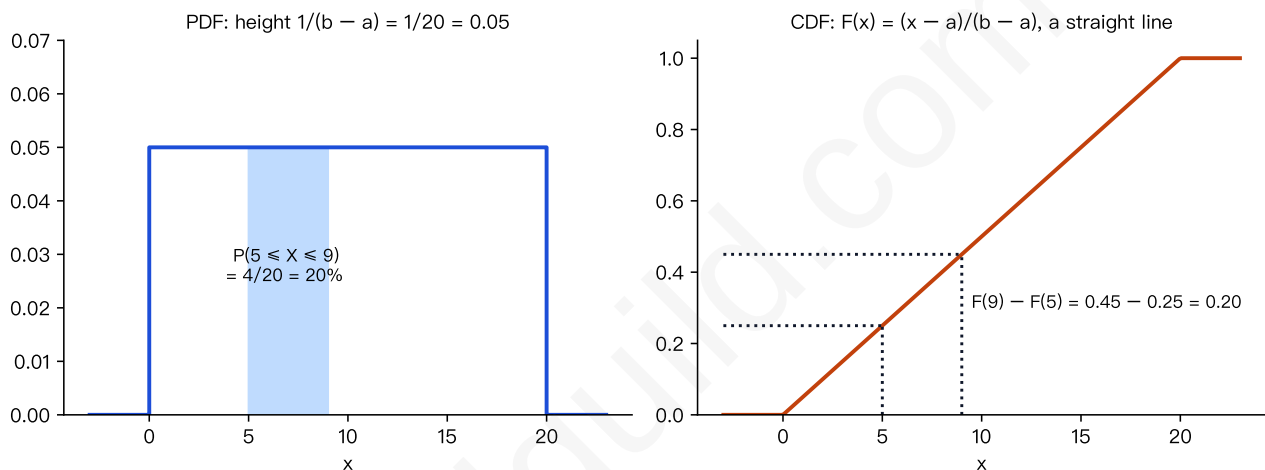
Continuous distributions

Continuous uniform on $[a, b]$: every equal-width sub-interval is equally likely, so the CDF rises in a straight line, $F(x) = (x - a)/(b - a)$ for $a \leq x \leq b$.

KEY CONCEPT

$$P(x_1 \leq X \leq x_2) = \frac{x_2 - x_1}{b - a} \quad \text{mean} = \frac{a + b}{2} \quad \sigma^2 = \frac{(b - a)^2}{12}$$

Skew 0, kurtosis 1.8 (excess -1.2). Used to generate random draws for Monte Carlo simulation. *Example.* Uniform on $[0, 20]$: $P(5 \leq X \leq 9) = 4/20 = 20\%$.



Continuous uniform distribution on $[0, 20]$: probability as an area, and the straight-line CDF

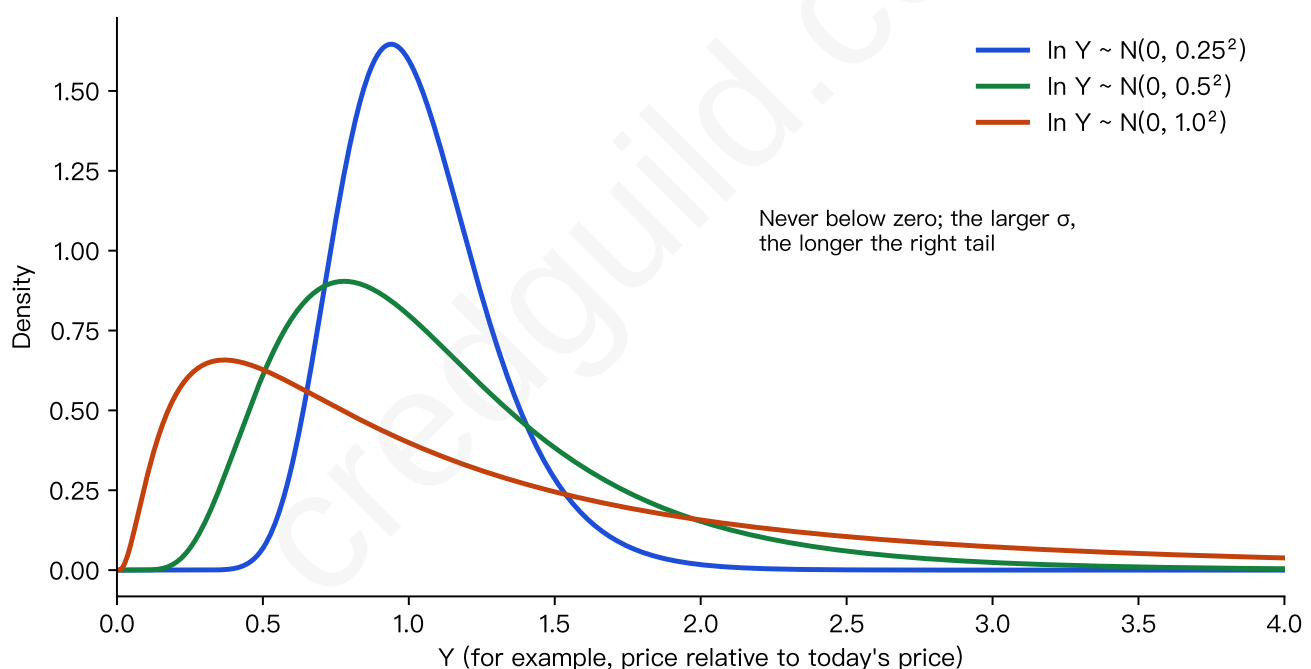
Normal (also called the **Gaussian distribution**), $N(\mu, \sigma^2)$: symmetric bell curve, fully described by its mean and variance, or equivalently its mean and standard deviation (skew 0, excess kurtosis 0). About 68% of outcomes lie within $\pm 1\sigma$ and about 95% within $\pm 2\sigma$. Sums of many independent variables tend toward normality, and a binomial converges to a normal as n grows. It is unbounded in both directions, so it can produce negative values. The normal with mean 0 and variance 1, $N(0, 1)$, is the **standard normal distribution**; its horizontal axis measures standard deviations from the mean.

Example. Annual returns are normal with a mean of 7% and a standard deviation of 15%. About 68% of years should fall between $7 - 15 = -8\%$ and $7 + 15 = 22\%$, and about 95% between -23% and 37% . A -23% return lies $(-23 - 7)/15 = -2$ standard deviations from the mean. The curve is symmetric, so the 5% of years outside $\pm 2\sigma$ split evenly: about 2.5% of years should lose more than 23%. Subtracting the two bands gives the share between them: about $(95\% - 68\%)/2 = 13.5\%$ of outcomes lie between one and two standard deviations below the mean, and the same share lies between one and two standard deviations above it.

Lognormal. Y is lognormal when $\ln Y$ is normal; equivalently $Y = e^X$ with X normal.

- Because $e^X > 0$ for any X , the lognormal is bounded below by zero and never takes a negative value.

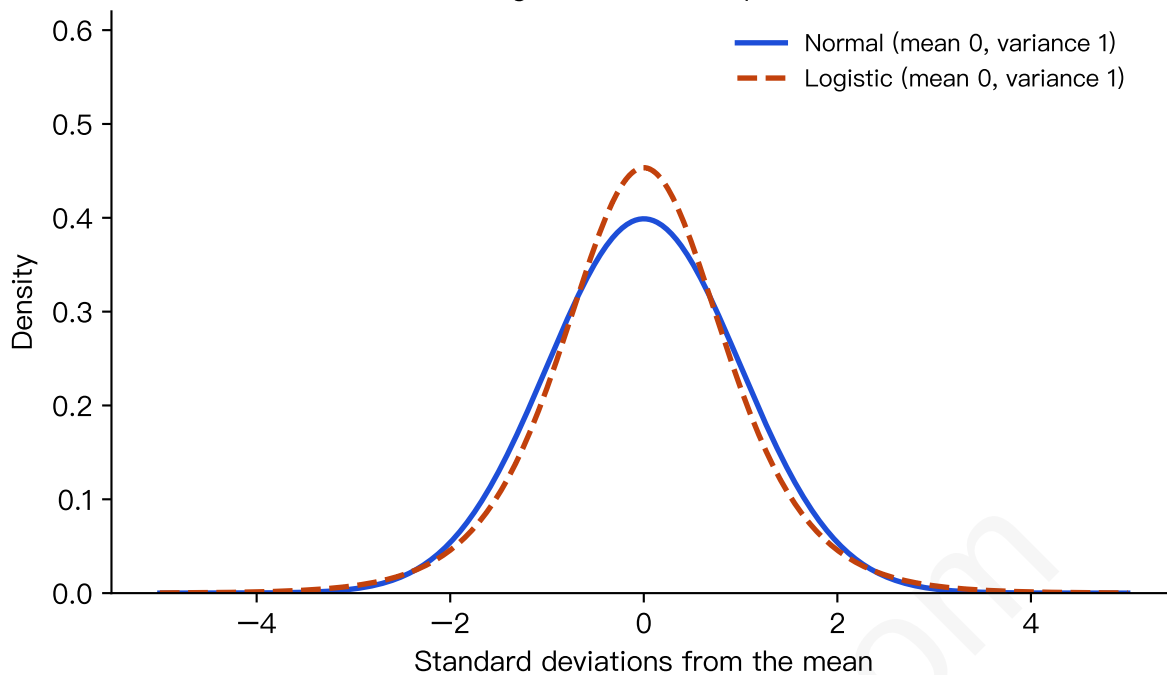
- It is **positively (right) skewed**, with a long right tail.
- That makes it well suited to modeling asset prices. A price following a lognormal cannot fall below zero, so the implied return can never be worse than -100% . A normal price model, by contrast, would allow negative prices (returns below -100%), which is impossible for a limited-liability asset.
- If continuously compounded price returns $\ln(P_1/P_0)$ (no distributions paid) are normally distributed, then prices are lognormal, because $P_T = P_0 e^r$ with r normal. Over several periods, r is the sum of the one-period log returns (Reading 1), and sums of many independent returns tend toward a normal distribution. A lognormal price model therefore fits well even when single-period returns are not exactly normal.
- Its two parameters are the mean and SD of $\ln Y$, called the **location parameter** and the **scale parameter**. They are not the mean and SD of Y itself: the mean of Y is $e^{\mu + \sigma^2/2}$.
- Prices can be projected with a geometric Brownian motion model. Here μ is the expected (arithmetic) growth rate of the price, not the location parameter of $\ln Y$ used above, and $T(\mu - \sigma^2/2)$ is the **drift term**: $\ln S_T = \ln S_0 + T(\mu - \sigma^2/2)$. *Example.* $S_0 = 25$, $\mu = 8\%$, $\sigma = 20\%$, $T = 2$: $\ln S_T = \ln 25 + 2(0.08 - 0.02)$, so $S_T = 25e^{0.12} = 28.19$.
- Exam convention: the price from the drift-term formula is the expected price in T years. Current practice: with μ as the expected growth rate, that figure is the median of the lognormal price distribution; the mean price is $S_0 e^{\mu T}$ (here $25e^{0.16} = 29.34$), which is higher because of the right skew.



Lognormal densities for three standard deviations of $\ln Y$: all start at zero and are skewed to the right (illustrative parameters)

Logistic. Bell-shaped and symmetric like the normal (skew 0), but with **fatter tails**. Its excess kurtosis is $+1.2$, so extreme outcomes are more likely than under a normal with the same variance. The variance is $\pi^2 s^2/3$ for scale parameter s . The logistic distribution is used in *logistic regression*, where the dependent variable is binary (0/1), and in machine learning.

Same variance: logistic has a taller peak and fatter tails



Normal vs. logistic densities with equal variance.

Matching the distribution to the problem

Situation	Distribution
Number of successes in a fixed number of independent yes/no trials	Binomial
Count of rare events in an interval at a constant rate	Poisson
Equally likely finite outcomes	Discrete uniform
Random draws for simulation over a range	Continuous uniform
Returns, test statistics, sums of many variables	Normal
Asset prices (non-negative, right-skewed)	Lognormal
Binary-outcome regression; symmetric but fat-tailed data	Logistic

COMMON EXAM TRAPS

- Calling the Poisson a "timing" or "waiting-time" distribution: it models the count of events in an interval.
- Thinking a lognormal is negatively skewed or symmetric; it is positively skewed and bounded at zero.
- Reversing the log relation: if Y is lognormal then $\ln Y$ is normal (not lognormal); if X is normal then e^X is lognormal.
- Saying a normal is appropriate for prices "because it only lets prices fall to zero". A normal is unbounded and allows negative prices.
- Attributing skewness to the logistic distribution: its skew is zero, and it differs from the normal in kurtosis (fatter tails).

- Using the binomial when the number of trials is not fixed or the success probability changes from trial to trial.

EXAM SHORTCUTS

- Identify the distribution from the wording: successes in a fixed number of trials is binomial, a count of events in an interval is Poisson, random draws for simulation are continuous uniform, and asset prices are lognormal.
- Normal bands: about 68% of outcomes lie within $\pm 1\sigma$ and 95% within $\pm 2\sigma$, so about 2.5% lie beyond 2σ on each side and about 13.5% lie between 1σ and 2σ on each side.
- For a Poisson variable, the probability of at least one event is $1 - e^{-\lambda}$.

BOTTOM LINE

- A binomial variable counts successes in n independent trials with a constant p : $E(X) = np$ and $\sigma^2 = np(1 - p)$; a single trial is a Bernoulli distribution.
- A Poisson variable counts events in an interval at a constant rate λ , and its mean and variance both equal λ .
- For a continuous uniform variable on $[a, b]$, $P(x_1 \leq X \leq x_2) = (x_2 - x_1)/(b - a)$.
- The normal distribution is symmetric, fully described by its mean and variance, and unbounded in both directions.
- If $\ln Y$ is normal, Y is lognormal: bounded below by zero and positively skewed, which suits asset prices.
- The logistic distribution is symmetric like the normal but has fatter tails (excess kurtosis +1.2).

QUICK CHECK

Question 2

Each week, the fixed-income desk at Corrib Securities submits competitive bids in 25 separate government bond auctions. The desk's chance of winning an allocation is 0.36 in every auction, and the result of one auction does not affect any other. For the number of auctions the desk wins in a week, the expected value and the variance are, respectively:

- A 16.0 and 5.76.
- B 9.0 and 2.40.
- C 9.0 and 5.76.

Answer: C

The number of wins is a binomial random variable: there is a fixed number of independent trials (25 auctions), each with only two outcomes (win or lose) and the same probability of success ($p = 0.36$). Its mean is np and its variance is $np(1 - p)$.

$$E(X) = np = 25 \times 0.36 = 9.0$$

$$\sigma^2(X) = np(1 - p) = 25 \times 0.36 \times 0.64 = 5.76$$

(The standard deviation would be $\sqrt{5.76} = 2.40$.) As with any discrete count, the mean of 9.0 is the long-run average number of wins per week, not a guaranteed result.

Why the other options are wrong

- **A.** 16.0 uses the probability of *losing* an auction, 25×0.64 , so it is the expected number of losses. The variance of 5.76 is right, because $np(1 - p)$ is the same whichever outcome is called a success.
- **B.** The mean of 9.0 is right, but 2.40 is the *standard deviation*, $\sqrt{5.76}$, which is the square root of the variance.

Key takeaway Binomial: $E(X) = np$ and $\sigma^2 = np(1 - p)$. Define "success" as the outcome being counted, and do not stop at the standard deviation when the variance is asked.

MODULE 6.3

Conditional Expectations and Bayesian Updating

LOS 6.c — Joint distributions, conditional expectations, variances and covariances

Joint and marginal probabilities

A **joint probability distribution** gives the probability of every combination of outcomes of two or more variables; for two variables it is *bivariate*. For discrete variables it is often shown as a **contingency table**. Row and column totals are the **marginal distributions**.

For two variables in general, the **joint cumulative distribution function** gives the probability that *both* variables are at or below given values: $F_{12}(x_1, x_2) = P(X_1 \leq x_1, X_2 \leq x_2)$. A bivariate distribution is described by each variable's mean and variance plus how they move together (their covariance or correlation).

400 listed firms	Technology	Non-technology	Total
Return > 10%	64	56	120
Return ≤ 10%	96	184	280
Total	160	240	400

- Marginal: $P(\text{tech}) = 160/400 = 0.40$; $P(R > 10\%) = 120/400 = 0.30$.
- Joint: $P(\text{tech and } R > 10\%) = 64/400 = 0.16$.
- Conditional: $P(R > 10\% \mid \text{tech}) = 64/160 = 0.40$, using only the technology column.

Conditional mean, variance and covariance

A **conditional expectation** $E(X \mid S)$ is the expected value of X given that condition S holds, read "the expected value of X given S ". A forecast revised for new information is a conditional expectation; for example, "if the merger is approved, my EPS estimate rises to \$4.20".

$$E(X \mid S) = \sum_i P(x_i \mid S) x_i$$

The **conditional variance** measures dispersion around the **conditional mean** $E(X | S)$ rather than the **unconditional mean** $E(X)$:

$$\text{Var}(X | S) = \sum_i P(x_i | S) [x_i - E(X | S)]^2 = E(X^2 | S) - [E(X | S)]^2$$

The **conditional covariance** is defined the same way for two variables, using conditional probabilities and conditional means. It is a different concept from the conditional variance of one variable.

A **conditional distribution** describes one variable after the other variables in a joint distribution are restricted to certain values. Conditioning changes the relevant probability distribution, so it can change both the expected value and the variance.

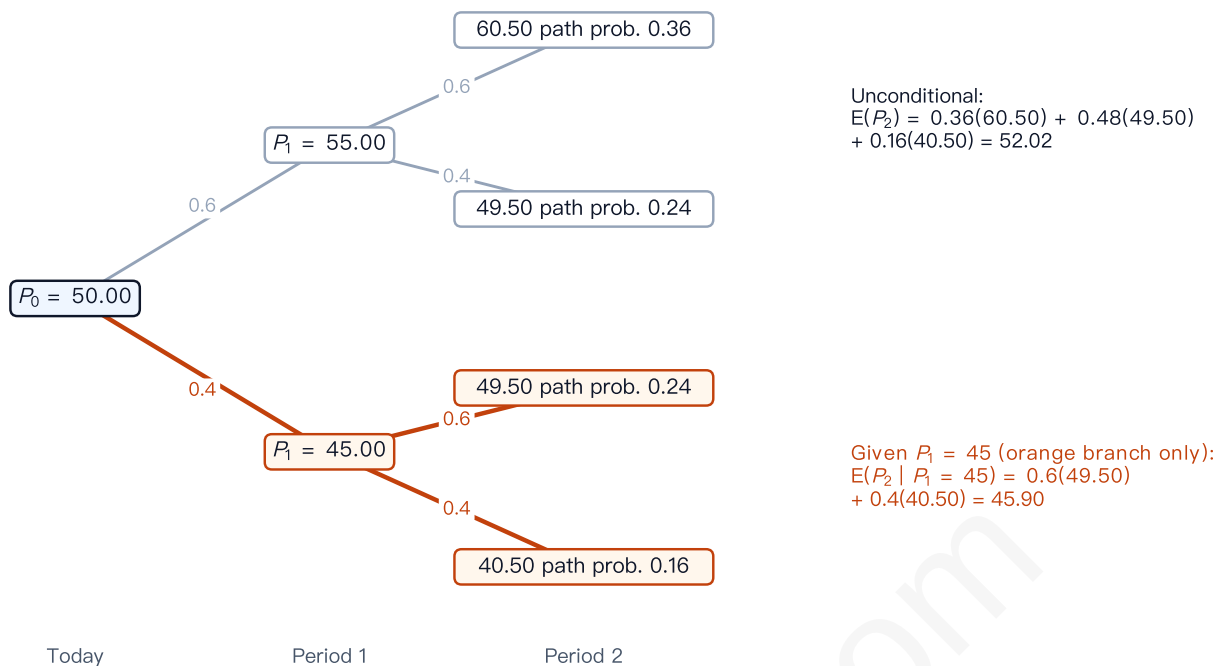
Concept	Probabilities used	Deviations measured from
Unconditional mean/variance	$P(x_i)$ over all outcomes	$E(X)$
Conditional mean/variance	$P(x_i S)$ over outcomes consistent with S	$E(X S)$
Joint probability	$P(A \text{ and } B)$	— (not an expectation)

Probability trees

A tree shows sequential outcomes; the probability of a full path is the product of the branch probabilities along it. The unconditional expected value weights every end node by its path probability, the product of the branch probabilities along that path. Imposing a condition (for instance, "the price fell in period 1") prunes the tree. Only branches that start from the observed node remain, and their branch probabilities become the conditional probabilities.

Example. A share at 50 moves up 10% with probability 0.6 or down 10% with probability 0.4 in each of two periods (independent moves).

- End prices: 60.50 (prob. 0.36), 49.50 (prob. 0.48 via two paths), 40.50 (prob. 0.16). Unconditional $E(P_2) = 0.36(60.50) + 0.48(49.50) + 0.16(40.50) = 52.02$.
- Given $P_1 = 45$ (a fall in period 1): $E(P_2 | P_1 = 45) = 0.6(49.50) + 0.4(40.50) = 45.90$; $\text{Var}(P_2 | P_1 = 45) = 0.6(3.6)^2 + 0.4(-5.4)^2 = 19.44$, so the conditional SD is 4.41.
- Given $P_1 = 55$: $E(P_2 | P_1 = 55) = 0.6(60.50) + 0.4(49.50) = 56.10$. Weighting the two conditional expectations by their branch probabilities gives back the unconditional value, $0.6(56.10) + 0.4(45.90) = 52.02$.



Two-period stock price tree and the branch that remains once the period-1 price is known to be 45

A common error is to weight the remaining outcomes by their *joint* (path) probabilities, e.g. $0.4 \times 0.6 \times 49.50 + 0.4 \times 0.4 \times 40.50$, which gives 18.36. That figure is $P(S) E(P_2 | S)$ rather than the conditional expectation.

LOS 6.d — Updating probabilities with Bayes' formula

Bayes' formula revises a **prior probability** (also called an **a priori probability**) of an event into a **posterior probability** after new information I is observed. It follows from the **multiplication rule**, $P(I | E)P(E) = P(I \text{ and } E)$: writing the joint probability two ways gives $P(E | I)P(I) = P(I | E)P(E)$, and dividing by $P(I)$ gives

KEY CONCEPT

$$P(E | I) = \frac{P(I | E)}{P(I)} \times P(E) = \frac{P(I | E) P(E)}{P(I | E)P(E) + P(I | E^c)P(E^c)}$$

The denominator is the **unconditional probability of the information**: the sum of its joint probabilities with each state.

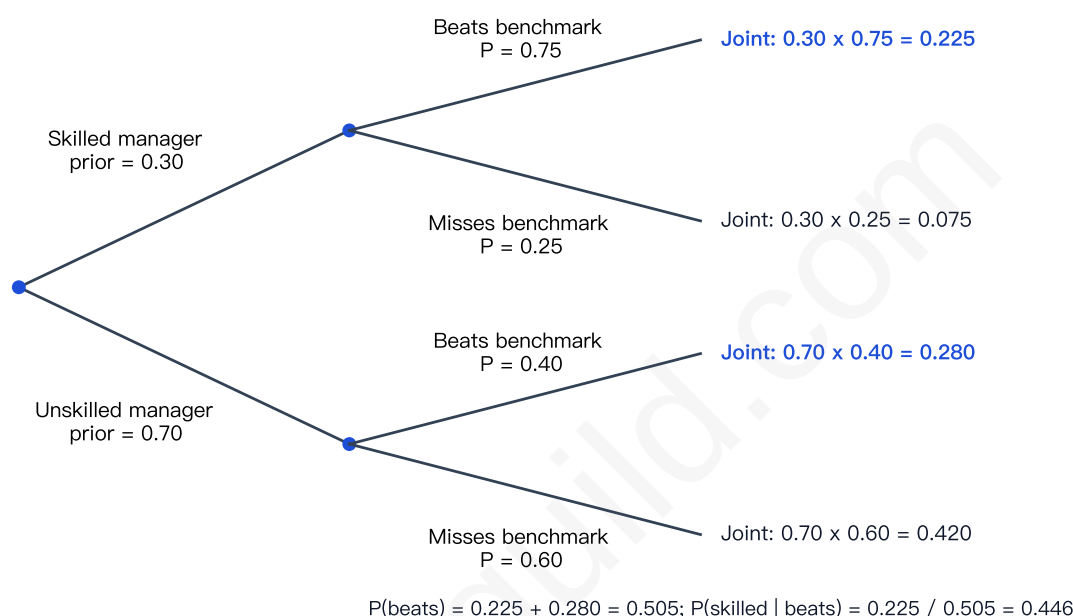
The formula can be checked against the 400-firm table in LOS 6.c. Counting directly, 64 of the 120 firms that returned more than 10% are technology firms, so $P(\text{tech} | R > 10\%) = 64/120 = 0.533$. Bayes' formula reaches the same figure from the reverse conditional probability: $P(R > 10\% | \text{tech}) P(\text{tech}) / P(R > 10\%) = 0.40 \times 0.40 / 0.30 = 0.533$.

Applying Bayes' formula:

1. Write down the prior $P(E)$ and its complement $P(E^c) = 1 - P(E)$.

2. Write down the likelihoods $P(I | E)$ and $P(I | E^c)$; take complements if the information is a "does not happen" event.
3. Multiply to get the joint probabilities of I with each state.
4. Add the joint probabilities to get $P(I)$.
5. Posterior = the joint probability for the state of interest $\div P(I)$.

Example. Of the fund managers in a database, 30% are skilled. A skilled manager beats the benchmark in a given year with probability 0.75; an unskilled one with probability 0.40. A manager beat the benchmark. What is the probability that she is skilled?



Bayes' formula as a probability tree: skilled vs. unskilled fund managers

$$P(\text{skilled} | \text{beat}) = \frac{0.30 \times 0.75}{0.30 \times 0.75 + 0.70 \times 0.40} = \frac{0.225}{0.505} = 0.446$$

The evidence raises the probability from the 30% prior to about 45%, but because unskilled managers are the majority, a single good year is far from proof of skill.

COMMON EXAM TRAPS (LOS 6.C AND 6.D)

- Reporting the joint probability $P(I \text{ and } E)$ (step 3) as the answer instead of dividing by $P(I)$. In the fund manager example this gives 0.225 instead of 0.446.
- Reporting $P(I)$ (step 4), the unconditional probability of the information, instead of the posterior. In the same example this gives 0.505.
- Confusing $P(I | E)$ with $P(E | I)$; they are generally very different.
- Comparing likelihoods only, $P(I | E) / [P(I | E) + P(I | E^c)]$, which ignores the priors. There it gives $0.75 / 1.15 = 0.652$.
- Forgetting to use complements when the information is that something did not happen.
- In trees, using joint path probabilities for a conditional expectation, or averaging all end nodes equally.

- Measuring conditional variance around the unconditional mean.

EXAM SHORTCUTS

- For Bayes' formula, write one row per state with prior $\times$ likelihood, add the rows to get $P(I)$, and divide the row for the state of interest by that total.
- Check a conditional expectation answer: the conditional expectations weighted by their branch probabilities must give back the unconditional expectation.

BOTTOM LINE

- A joint probability is $P(A \text{ and } B)$, marginal probabilities are the row and column totals, and a conditional probability uses only the outcomes consistent with the condition.
- The conditional expectation is $E(X | S) = \sum P(x_i | S)x_i$, and the conditional variance is measured around $E(X | S)$.
- In a probability tree a path probability is the product of its branch probabilities.
- Bayes' formula is $P(E | I) = P(I | E)P(E)/P(I)$, where $P(I)$ adds the joint probabilities of the information with every state.
- $P(I | E)$ and $P(E | I)$ are generally different, and the posterior still depends on the prior.

QUICK CHECK

Question 3

A high-yield portfolio holds 40% BB-rated bonds and 60% B-rated bonds. Historically, 10% of BB-rated bonds and 30% of B-rated bonds default within five years. If a bond chosen at random from the portfolio defaults within five years, the probability that it was BB-rated is *closest* to:

- A 0.100.
- B 0.182.
- C 0.333.

Answer: B

This is a Bayesian update: the prior probability that a bond is BB-rated (0.40) is revised using the information that the bond defaulted. The posterior is the joint probability of 'BB and default' divided by the unconditional probability of default.

$$P(\text{BB and D}) = P(D | \text{BB})P(\text{BB}) = 0.10 \times 0.40 = 0.04$$

$$P(\text{B and D}) = P(D | B)P(B) = 0.30 \times 0.60 = 0.18$$

$$P(D) = 0.04 + 0.18 = 0.22$$

$$P(BB | D) = \frac{0.04}{0.22} = 0.182$$

Why the other options are wrong

- **A.** 0.100 is $P(\text{default} | BB)$, the likelihood. The question asks for the reverse conditional probability, the posterior $P(BB | \text{default})$.
- **C.** 0.333 is the ratio of the two default rates ($0.10/0.30$); it ignores the portfolio weights (priors) and the unconditional probability of default.

Key takeaway $P(A | B) \neq P(B | A)$. Build the joint probabilities, add them for the denominator, then divide.

Practice questions

Question 4

A consultant at Northgate Partners has estimated the scenario returns shown in the table below. The standard deviation of returns on Fund Alder is *closest* to:

Scenario forecasts for Fund Alder and Fund Birch

Scenario	Probability	Return on Fund Alder	Return on Fund Birch
Boom	20%	19%	12%
Expansion	25%	15%	11%
Stagnation	35%	9%	10%
Contraction	20%	1%	8%

- A 1.34%.
- B 6.18%.
- C 6.78%.

Question 5

A commodity desk simulates thousands of possible copper price paths and wants to be certain that no simulated price is ever negative. Which distribution for the price level is *best* suited to this goal?

- A Normal.
- B Logistic.
- C Lognormal.

Answers: 4. B · 5. C

Why each answer is right and why each other option is wrong: credguild.com/cfa/level-1/notes/quantitative-methods/lognormal-distribution/. The full bank has 46 questions on this reading.

credguild.com

QUANTITATIVE METHODS · READING 7

Null Hypothesis

What you'll learn

LOS 7.a Explain the central limit theorem and use it to build confidence intervals, measure parametric VaR and choose among sampling methods.

LOS 7.b Construct and interpret parametric tests of differences between means, mean differences, a single variance, two variances and a correlation coefficient; recognize p-values and p-hacking.

LOS 7.c Compare parametric and nonparametric tests, identify when a nonparametric test is appropriate, and construct and interpret the Spearman rank correlation test and the chi-square test of independence.

MODULE 7.1

The Central Limit Theorem, Confidence Intervals, and Sampling

This reading moves from a sample to the population: how sample means behave under the central limit theorem, how confidence intervals are built and how samples are drawn. It then sets out the steps of a hypothesis test, the test statistic for each parametric test, the two types of error, and the nonparametric tests used for ranks and categories.

LOS 7.a — The central limit theorem, confidence intervals and sampling

The central limit theorem (CLT)

A **population parameter** such as μ or σ is a fixed number, usually unknown. A **sample statistic** such as $\bar{X}$ or s is computed from the sample actually drawn, so it would take a different value in another sample: it is itself a random variable. The distribution of its values over all possible samples of the same size is the statistic's sampling distribution.

Take many **simple random samples** of size n from a population with mean μ and a *finite* variance σ^2 . The **central limit theorem** says that the **sampling distribution of the sample mean** $\bar{X}$ becomes approximately normal as n grows, whatever the shape of the population itself (skewed, flat, lumpy):

KEY CONCEPT

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad \text{for large } n \text{ (rule of thumb: } n \geq 30)$$

Three consequences follow:

- The mean of all possible sample means equals the population mean: $E(\bar{X}) = \mu$.
- The variance of the sample mean is the population variance divided by the sample size, σ^2/n , which shrinks as n rises.
- The CLT describes the distribution of the sample mean. Individual observations keep the population's shape, and the CLT says nothing about the distribution of the sample standard deviation.

Standard error and sampling error

Sampling error is the difference between a sample statistic and the population parameter it estimates (e.g., $\bar{X} - \mu$). Its typical size is the **standard error of the sample mean**:

KEY CONCEPT

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \text{ (}\sigma \text{ known)} \quad s_{\bar{X}} = \frac{s}{\sqrt{n}} \text{ (}\sigma \text{ unknown, the usual case)}$$

When σ is unknown, s replaces it, and the result is an estimate of the standard error. The standard deviation describes how widely individual observations scatter around the mean, and a larger sample does not make it smaller. The standard error describes how precisely $\bar{X}$ estimates μ .

Calculator. From raw data, enter the observations in the DATA worksheet, read the sample standard deviation S_x in the STAT worksheet (1-V), and divide it by $\sqrt{n}$.

Example. A population of quarterly returns has $\sigma = 18\%$. With $n = 36$, $\sigma_{\bar{X}} = 18/6 = 3\%$; quadrupling the sample to 144 halves the standard error to 1.5%.

Desirable properties of an estimator

Property	Meaning
Unbiased	Expected value of the estimator equals the parameter (the sample mean is unbiased for μ).
Efficient	Smallest sampling variance among unbiased estimators. An estimator that is both unbiased and efficient is the minimum variance unbiased estimator (MVUE) ; the unbiased estimator with the lowest variance among those formed as a linear combination of the sample data is the best linear unbiased estimator (BLUE) .
Consistent	Accuracy improves as n increases (standard error falls toward zero).
Robust	Not sensitive to outliers or non-normal data (the sample mean is <i>not</i> robust).

Confidence intervals

A **point estimate** is a single value (e.g., $\bar{X}$). A **confidence interval** is a range built around it that contains the true parameter with a stated **confidence level** $1 - \alpha$, where α is the **significance level**:

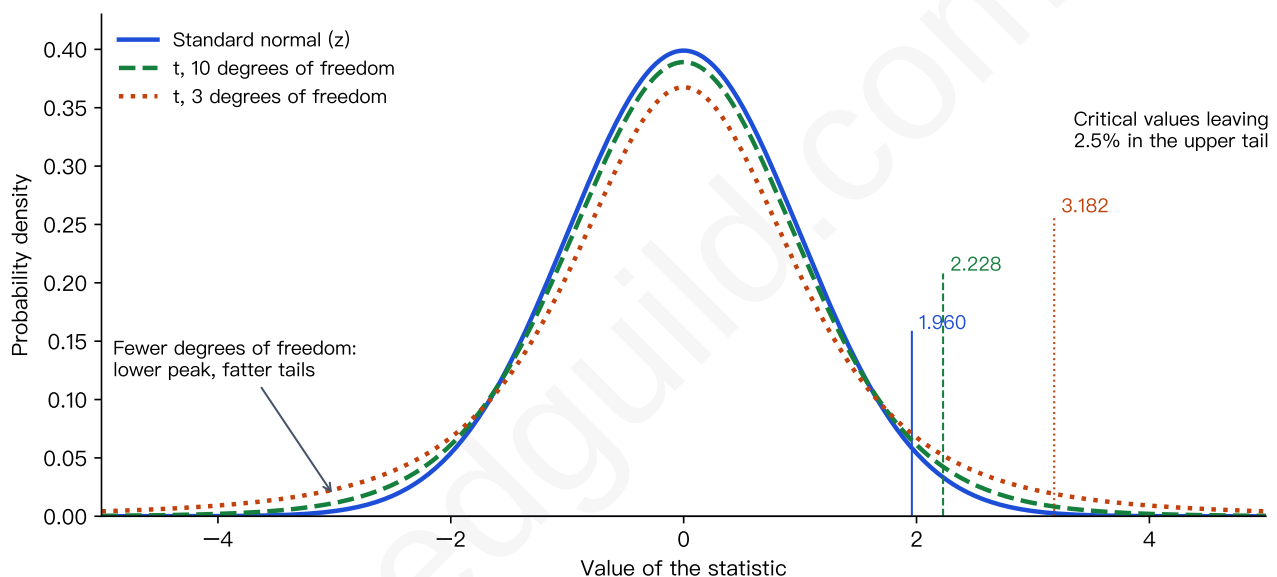
$$\text{point estimate} \pm \underbrace{(\text{reliability factor} \times \text{standard error})}_{\text{margin of error}}$$

Situation	Reliability factor	Interval for μ
Normal population, σ known	$z_{\alpha/2}$	$\bar{X} \pm z_{\alpha/2} \sigma / \sqrt{n}$
Normal (or approximately normal) population, σ unknown (use s); required when n is small	$t_{\alpha/2}$ with $n - 1$ df	$\bar{X} \pm t_{\alpha/2} s / \sqrt{n}$

Situation	Reliability factor	Interval for μ
σ unknown, large sample ($n \geq 30$), any population shape (CLT)	$t_{\alpha/2}$ preferred; $z_{\alpha/2}$ acceptable as an approximation	$\bar{X} \pm t_{\alpha/2} s / \sqrt{n}$, or $\bar{X} \pm z_{\alpha/2} s / \sqrt{n}$ as a large-sample approximation

Two-tailed z reliability factors $z_{\alpha/2}$ are 1.65 for 90% (1.645 before rounding), 1.96 for 95% and 2.58 for 99% (2.576 before rounding). Three things set the width of an interval. A higher confidence level (a larger reliability factor) widens it, a more dispersed population widens it, and a larger sample (a smaller standard error) narrows it.

The **Student's t-distribution** is bell-shaped and symmetric but has fatter tails than the normal; its critical values are larger than z-values (wider intervals) and fall toward the z-value as degrees of freedom rise. For a 95% interval, the t reliability factor is 2.228 with 10 df, 2.086 with 20 df and 2.042 with 30 df, against 1.96 for z. Select a t-value from a table using the area in one tail ($\alpha/2$ for a two-sided interval) and $n - 1$ degrees of freedom.



Student's t-distribution with 3 and 10 degrees of freedom compared with the standard normal distribution

Example (σ known). With 49 observations, $\bar{X} = 6.2$ and $\sigma = 14$, the standard error $= 14/7 = 2$; 95% interval $= 6.2 \pm 1.96 \times 2 = 2.28$ to 10.12 . *Example (σ unknown, small sample).* Ten observations from a population assumed normal give s . The 95% reliability factor is $t_{9, 0.025} = 2.262$ rather than 1.96. If a small sample comes from a clearly non-normal population, neither a t- nor a z-interval is reliable.

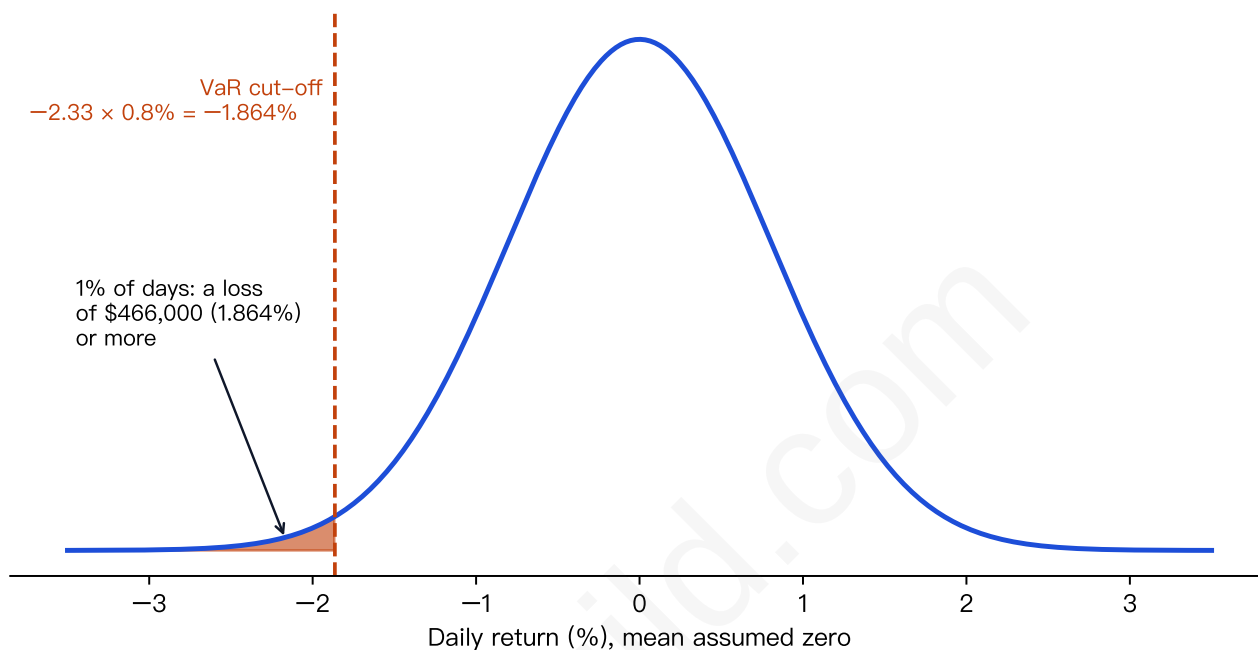
Value at risk (VaR) — a one-tailed interval

Value at risk is the loss that will be equaled or exceeded with a stated (small) probability over a given period. "The daily 1% VaR is \$0.47 million" means a 1% chance on any day of losing \$0.47 million or more (equivalently, 99% confidence that the loss will be smaller). VaR is the minimum loss at the stated probability; larger losses remain possible, so it is not the worst possible loss.

Under the **parametric method** (normal returns, mean return assumed to be zero):

$$\text{VaR}_\alpha = z_\alpha \times \sigma \times \text{portfolio value}$$

with the whole α in one tail: $z_{10\%} = 1.28$, $z_{5\%} = 1.65$ (1.645 before rounding), $z_{1\%} = 2.33$. With σ in percent and the portfolio value in dollars, VaR is a dollar loss; dropping the portfolio value gives VaR as a percentage of the portfolio. *Example.* For a \$25 million portfolio with daily $\sigma = 0.8\%$, the daily 1% VaR = $2.33 \times 0.008 \times \$25\text{m} = \$466,000$, or 1.864% of the portfolio.



Daily 1% VaR as the left-tail cut-off of the return distribution (\$25 million portfolio, daily $\sigma = 0.8\%$)

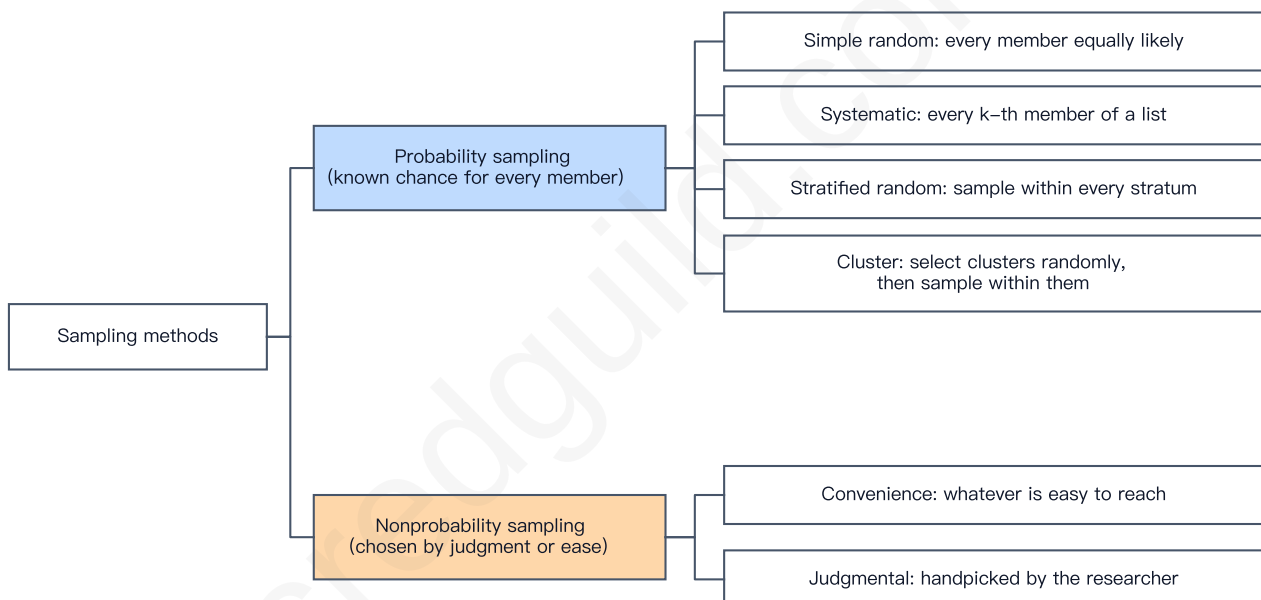
Drawbacks: returns are often skewed/fat-tailed, parameters change over time, history may not reflect new conditions, and it handles nonlinear payoffs (options) poorly.

Sampling methods

Family	Method	How the sample is drawn
Probability sampling (every member has a known, nonzero chance)	Simple random sampling	Every member has the <i>same</i> chance of selection, through random draws from the full list (so every possible sample of size n is equally likely). Equal individual chances alone are not enough; a systematic sample with a random start also gives every member the same chance.
	Systematic sampling	Select every k -th member of the population list.
	Stratified random sampling	Split the population into strata by a characteristic (sector, size, rating); draw a random sample <i>within each stratum</i> , usually in proportion to its weight in the population; pool the results.

Family	Method	How the sample is drawn
Nonprobability sampling	Cluster sampling	Split into clusters that each resemble the population; randomly select clusters and sample within them. The CLT applies if the clusters are large enough.
	Convenience sampling	Use whatever is easy to reach (e.g., colleagues at one office); cheap but likely unrepresentative.
	Judgmental sampling	The researcher handpicks elements using experience and judgment; may import the researcher's bias, but expert judgment can help on complex research questions.

Probability sampling is generally more accurate; nonprobability sampling is cheaper. Stratified and cluster sampling both split the population into groups. Stratified sampling draws from every group, so the sample keeps the population's mix across groups; cluster sampling uses only some groups, so differences between groups can be lost.



Sampling methods

The CLT does not rescue a non-representative sample.

COMMON EXAM TRAPS

- The CLT requires a finite variance and a large n , where n is the size of each sample; drawing more samples of the same small size does not help. It does not require a normal population, and it describes the sample mean.
- The standard error divides by $\sqrt{n}$ rather than by n . With $\sigma = 18\%$ and $n = 36$, dividing by 36 gives 0.5% instead of 3%.
- Two-sided 95% uses 1.96; one-tailed 5% (VaR) uses 1.645. Always check how many tails. In the VaR example, the two-tailed 1% value 2.58 would give \$516,000 instead of \$466,000.

- "Random draws, then keep every second one drawn" is still simple random sampling, because keeping fixed positions of an already random sequence leaves the selection purely random. Systematic sampling means every k -th item of a fixed, ordered *population list*.
- Grouping first and sampling within each group is stratified sampling; handpicking is judgmental sampling; choosing what is easiest to reach is convenience sampling.

EXAM SHORTCUTS

- Quadrupling the sample size halves the standard error.
- Reliability factors: two-tailed 90%, 95% and 99% use 1.65, 1.96 and 2.58; one-tailed 10%, 5% and 1% (VaR) use 1.28, 1.65 and 2.33.
- A higher confidence level widens an interval and a larger sample narrows it, which rules out answer choices that move the wrong way.

BOTTOM LINE

- By the central limit theorem, the mean of samples of size $n \geq 30$ from any population with a finite variance is approximately normal with mean μ and variance σ^2/n .
- The sample mean's standard error is $\sigma/\sqrt{n}$, or $s/\sqrt{n}$ when σ is unknown.
- A confidence interval is the point estimate $\pm$ reliability factor $\times$ standard error, with a t reliability factor and $n - 1$ df when σ is unknown.
- A desirable estimator is unbiased, efficient and consistent.
- Parametric VaR is $z_\alpha \times \sigma \times$ portfolio value with the whole α in one tail; it is the minimum loss at the stated probability, not the worst possible loss.
- Simple random, systematic, stratified and cluster sampling are probability methods; convenience and judgmental sampling are nonprobability methods.

QUICK CHECK

Question 1

A fixed-income analyst splits the universe of investment-grade corporate bonds into five credit-rating buckets. From each bucket she randomly draws a number of bonds proportional to that bucket's share of the universe and then combines the draws into one sample of 400 bonds. Her approach is best described as:

- A simple random sampling.
- B systematic sampling.
- C stratified random sampling.

Answer: C

The analyst first divides the population into subgroups (strata) using a classification scheme (here, credit rating) and then draws a random sample within each stratum, pooling the results. That is stratified random sampling. Sampling each stratum in proportion to its weight ensures that the rating mix of the sample mirrors the population.

Why the other options are wrong

- **A.** In simple random sampling every bond in the universe would have the same chance of selection from a single, undivided list; here the population is first split into rating buckets and each bucket is sampled separately.
- **B.** Systematic sampling selects every k-th member of an ordered population list. The analyst makes random draws inside each rating bucket instead.

Key takeaway Dividing the population into groups first and then sampling randomly within every group is stratified random sampling.

MODULE 7.2

Hypothesis Tests of the Population Mean

LOS 7.b — Hypothesis tests of a population mean

What a hypothesis test does

A **hypothesis** is a claim about what value a population parameter takes (e.g., "the mean daily return of this strategy is zero"). A sample is used to decide whether the data give enough evidence to reject that statement.

- **Parametric tests** make assumptions about the population's distribution and test its parameters (mean, variance).
- **Nonparametric tests** make few distributional assumptions or test things that are not parameters (ranks, category counts); Module 7.4 covers them.

The four steps

1. State the **null hypothesis** H_0 and the **alternative hypothesis** H_a .
2. Identify the **test statistic** and its distribution.
3. Specify the **significance level** α .
4. Establish the **decision rule** (the **critical value(s)**) and compare.

Hypotheses. The null is usually the claim the researcher hopes to reject, and it always contains the equality ($=$, $\leq$ or $\geq$). The alternative is what the researcher wants to find evidence for. The two are mutually exclusive and exhaustive.

Evidence sought	H_0	H_a	Test	Rejection region
Mean differs from μ_0 (either direction)	$\mu = \mu_0$	$\mu \neq \mu_0$	two-tailed test	both tails, $\alpha/2$ each
Mean is greater than μ_0	$\mu \leq \mu_0$	$\mu > \mu_0$	one-tailed test	right tail, all of α
Mean is less than μ_0	$\mu \geq \mu_0$	$\mu < \mu_0$	one-tailed test	left tail, all of α

Test statistic for a mean

The test statistic measures how many standard errors the sample mean lies from the hypothesized value:

KEY CONCEPT

$$\text{test statistic} = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \quad (\text{use } \sigma/\sqrt{n} \text{ if } \sigma \text{ is known})$$

When s is used and the population is normal (or approximately normal), the statistic follows a **t-distribution** with $n - 1$ degrees of freedom; with a large sample ($n \geq 30$) the CLT makes the t-test approximately valid even for a non-normal population, and the **z-statistic** (standard normal) is an acceptable approximation. A small sample from a clearly non-normal population supports neither test. When the population is normal and its variance is known, the statistic uses σ and follows the standard normal distribution at any sample size; z critical values have no degrees of freedom. The sign matters: a negative statistic means the sample mean is *below* μ_0 .

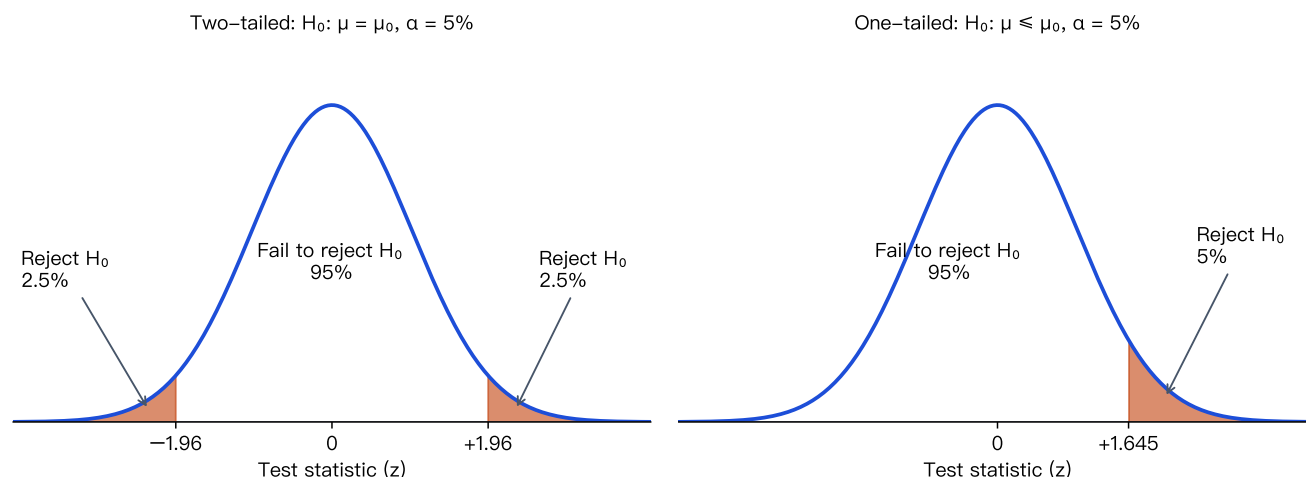
Critical z-values (exam values; three-decimal values in parentheses):

KEY CONCEPT

Significance level	Two-tailed test	One-tailed test
10%	± 1.65 (1.645)	1.28 (1.282)
5%	± 1.96	1.65 (1.645)
1%	± 2.58 (2.576)	2.33 (2.326)

Decision rule. Reject H_0 if the test statistic falls in the rejection region (beyond the critical value); otherwise **fail to reject** H_0 . The null is never "accepted"; failing to reject means only that the data were not strong enough to reject it. The **confidence level** is $1 - \alpha$. For a two-tailed test the rule is equivalent to a confidence interval: H_0 is rejected when the sample mean falls outside the interval $\mu_0 \pm (\text{critical value} \times \text{standard error})$ built around the hypothesized value.

The critical value depends on three things: the distribution of the test statistic (z, or t with its degrees of freedom), the significance level, and whether the test is one-tailed or two-tailed. At $\alpha = 5\%$, for example, a statistic of 1.75 exceeds the critical value of a right-tailed test (1.65) but not that of a two-tailed test (1.96), so the two tests reach different decisions.



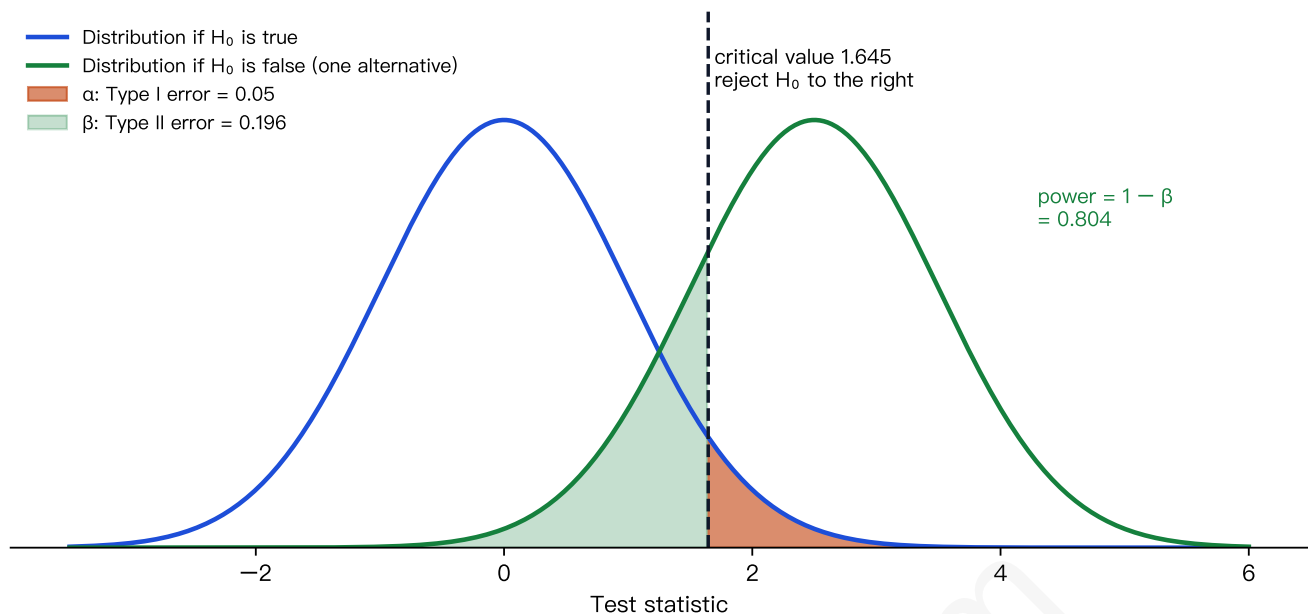
Rejection regions for z-tests at a 5% significance level: two-tailed (left) and one-tailed, right tail (right).

Example. A sample of 81 daily returns has $\bar{X} = 0.42\%$ and $s = 1.8\%$; the test is $H_0: \mu = 0$, two-tailed, at $\alpha = 5\%$. The standard error is $1.8/9 = 0.2\%$ and the statistic is $0.42/0.2 = 2.10$. Because 2.10 exceeds 1.96, H_0 is rejected. At $\alpha = 1\%$ (critical value ± 2.58), H_0 would not be rejected.

What moves the statistic. For a sample mean above μ_0 , the statistic (and the chance of rejection) *rises* when the sample mean rises, the standard deviation falls or the sample size rises (smaller standard error).

Type I and Type II errors, power

	H_0 is true	H_0 is false
Reject H_0	Type I error (false positive); probability = α	Correct decision; probability = power of the test = $1 - \beta$
Fail to reject H_0	Correct decision; probability = $1 - \alpha$ (confidence level)	Type II error (false negative); probability = β



Type I and Type II errors for a one-tailed test at $\alpha = 5\%$ (illustrative alternative)

- The significance level is the probability of a Type I error, that is, of rejecting a true null by chance.
- The **power of a test** is the probability of correctly rejecting a false null: $\text{power} = 1 - P(\text{Type II error})$.
- Only one error type can occur in a given test: a Type I error needs a true null, a Type II error needs a false null.
- Trade-off: with the sample size and the test statistic unchanged, lowering α (e.g., from 5% to 1%) *raises* the probability of a Type II error and *lowers* power; raising power in that setting means accepting a higher α . A lower α also means a larger critical value, so the matching confidence interval is wider.
- Exam convention: for a given α , power can be raised only by increasing the sample size, which lowers the probability of a Type II error while the probability of a Type I error stays at α . Current practice: at a given α , less noisy data (a lower standard deviation) or a more powerful test statistic also raise power.
- The chance of a Type II error depends jointly on α and the sample size, but the relation is not simple and that probability is hard to compute in practice. When several test statistics could be used for the same hypothesis, the most powerful one is normally preferred.

Example. A spam filter tests H_0 : "this message is legitimate." Sending a legitimate message to the junk folder rejects a true null (Type I); letting spam through fails to reject a false null (Type II).

WORKED EXAMPLE: A CONFIDENCE INTERVAL AND A ONE-TAILED TEST

A sample of 64 monthly returns has a mean of 0.85% and a standard deviation of 2.4%. The population standard deviation is unknown. Build a 95% confidence interval for the mean, then test at the 5% level whether the mean return is above 0.5%.

Step 1. Standard error $= 2.4/\sqrt{64} = 0.30\%$.

Step 2. 95% confidence interval. With $n = 64$, the z reliability factor is an acceptable approximation: $0.85 \pm 1.96 \times 0.30$, from 0.262% to 1.438%. The t reliability factor with 63 df is 1.998, which gives a slightly wider interval, from 0.251% to 1.449%.

Step 3. Test whether the mean return is above 0.5% at the 5% level: $H_0 : \mu \leq 0.5\%$ and $H_a : \mu > 0.5\%$. The test is one-tailed, so all of the 5% is in the right tail and the critical value is 1.65 (1.669 for t with 63 df).

Step 4. Test statistic = $(0.85 - 0.5)/0.30 = 1.17$.

Step 5. 1.17 is below the critical value, so H_0 is not rejected: the sample does not show a mean return above 0.5%.

Step 6. If the true mean is above 0.5%, this decision is a Type II error. A Type I error cannot have occurred, because H_0 was not rejected.

COMMON EXAM TRAPS

- A 5% significance level corresponds to a 95% confidence level; "95% significance" misreads the two terms.
- The probability of a Type I error in a two-tailed 5% test is 5%; 2.5% is the area in each tail.
- Power is $1 - \beta$; $1 - \alpha$ is the confidence level.
- H_a for a two-tailed test uses " $\neq$ "; "<" or ">" signals a one-tailed test.
- Check which tail: for $H_0 : \mu \geq \mu_0$, the null is rejected only when the statistic is *below* the negative critical value.
- A two-tailed 10% test and a one-tailed 5% test share the critical value 1.65, because both place 5% in the tail beyond it.

EXAM SHORTCUTS

- The hypothesis that contains the equality sign is H_0 .
- A two-tailed 10% test and a one-tailed 5% test share the critical value 1.65.
- For a two-tailed test, H_0 is rejected when the sample mean falls outside $\mu_0 \pm \text{critical value} \times \text{standard error}$, so a confidence interval answers the test directly.

BOTTOM LINE

- H_0 always contains the equality and is usually the claim the researcher hopes to reject; H_a is the claim the researcher seeks evidence for.
- The test statistic is $(\text{sample mean} - \mu_0) \div \text{standard error}$, following a t-distribution with $n - 1$ df when s is used.
- H_0 is rejected when the statistic falls beyond the critical value; otherwise the decision is to fail to reject, and the null is never accepted.
- A Type I error rejects a true null with probability α ; a Type II error fails to reject a false null with probability β ; power is $1 - \beta$.
- With the sample size unchanged, a lower α makes a Type II error more likely and reduces power.

QUICK CHECK

Question 2

A researcher keeps her sample size unchanged but tightens the significance level of her test from 5% to 1%. As a result, the probability of:

- A a Type II error increases.
- B rejecting a true null hypothesis increases.
- C failing to reject a false null hypothesis decreases.

Answer: A

Lowering the significance level lowers the probability of a Type I error. For a fixed sample size, the rejection region shrinks, so a false null is less likely to be rejected: the probability of a Type II error rises and the power of the test falls.

Why the other options are wrong

- **B.** Rejecting a true null is a Type I error, whose probability is the significance level; it falls from 5% to 1%.
- **C.** Failing to reject a false null is a Type II error; its probability *increases* when α is reduced with no change in sample size. A null is never 'accepted'; a test can only fail to reject it.

Key takeaway With n fixed, a lower α means a higher β and a lower power.

MODULE 7.3

Other Parametric Hypothesis Tests

LOS 7.b — Other parametric hypothesis tests

Every test below uses the same four steps as the test of a single mean (hypotheses, test statistic, significance level, decision rule). What changes is the **test statistic** and the distribution its critical values come from.

Summary of the tests

KEY CONCEPT

What is tested	Test statistic	Distribution / degrees of freedom	Key assumptions
Single mean	$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$	t, $n - 1$ (z if n is large)	normal population, or a large sample
Difference between means (independent samples)	$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2/n_1 + s_p^2/n_2}},$ with pooled variance $s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$	t, $n_1 + n_2 - 2$	independent samples, normal populations, equal (unknown) population variances
Mean difference (paired comparisons test, dependent samples)	$t = \frac{\bar{d} - \mu_{d0}}{s_d/\sqrt{n}}$	t, $n - 1$ (n = number of pairs)	paired observations, normal differences

What is tested	Test statistic	Distribution / degrees of freedom	Key assumptions
Single variance	$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$	chi-square , $n-1$	normal population
Equality of two variances	$F = \frac{s_1^2}{s_2^2}$ (larger variance on top)	F , $n_1 - 1$ and $n_2 - 1$	normal populations, independent samples
Correlation equals zero	$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$	t , $n-2$	normally distributed variables

Difference between means vs. paired comparisons

- **Independent samples** (e.g., returns of small-cap funds vs. returns of large-cap funds): test $H_0 : \mu_1 - \mu_2 = 0$ with the difference-in-means t-test, $df = n_1 + n_2 - 2$. These df belong to the **pooled** version, which assumes the two population variances are equal and combines both samples into one variance estimate s_p^2 . A test that does not assume equal variances uses the unpooled standard error $\sqrt{s_1^2/n_1 + s_2^2/n_2}$, with df from a separate (Welch) approximation that are usually smaller than $n_1 + n_2 - 2$. When $n_1 = n_2$, the pooled and unpooled standard errors are identical, and when the sample variances are similar the two t-values are close.
 - Exam convention: the difference-in-means statistic may be shown with the unpooled standard error $\sqrt{s_1^2/n_1 + s_2^2/n_2}$ and still be compared with a t critical value for $n_1 + n_2 - 2$ df.
 - Current practice: $n_1 + n_2 - 2$ df go with the pooled standard error (equal variances assumed); the unpooled (Welch) test uses fewer df.
- **Dependent samples** (the *same* members measured twice, e.g., each fund's return before and after a fee cut): compute the difference for each pair, then run a single-mean test on those differences. This is the **paired comparisons test**, with $df = n - 1$.

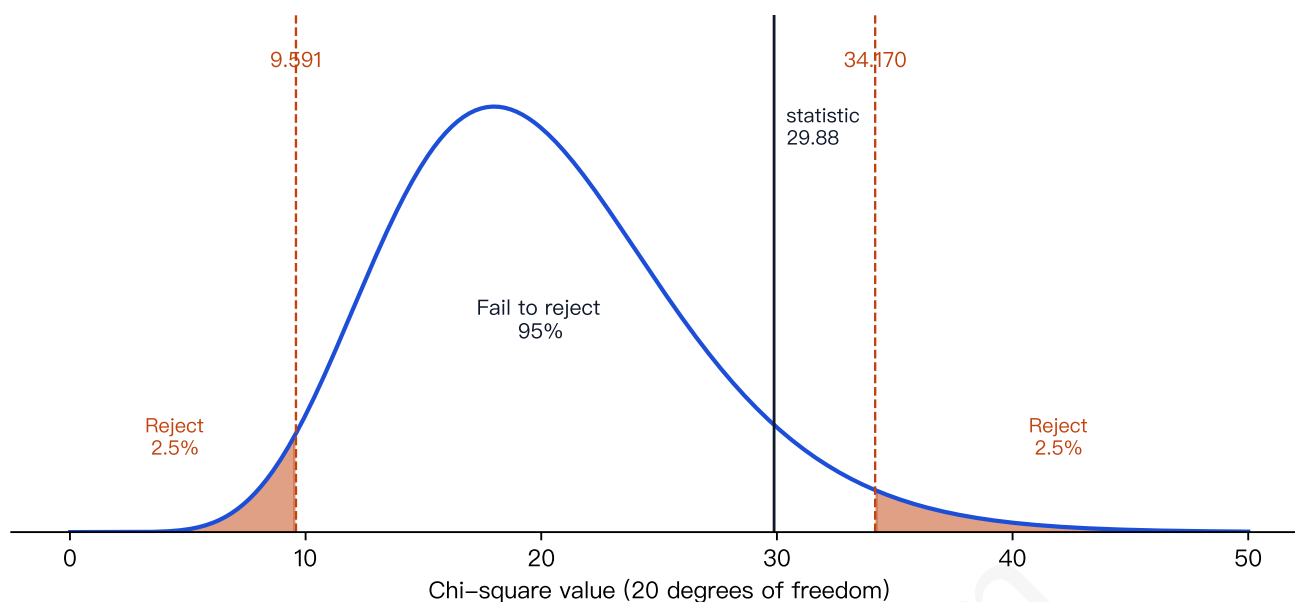
Example (independent samples). $\bar{X}_1 = 1.9$, $s_1 = 1.5$, $n_1 = 25$; $\bar{X}_2 = 1.2$, $s_2 = 1.3$, $n_2 = 31$. Assuming equal population variances, $s_p^2 = (24 \times 1.5^2 + 30 \times 1.3^2)/54 = 1.939$ and $t =$

$0.7/\sqrt{1.939/25 + 1.939/31} = 1.87$; $df = 54$, 5% two-tailed critical value 2.005, so the null of equal means is not rejected. (The unpooled standard error would give $t = 1.84$; the conclusion is the same.)

Example (paired). With 16 pairs, $\bar{d} = 1.2$ and $s_d = 2.0$, $t = 1.2/(2.0/4) = 2.40$. This exceeds $t_{15, 0.025} = 2.131$, so a zero mean difference is rejected.

Test of a single variance (chi-square)

The **chi-square distribution** is asymmetric (right-skewed) and bounded below by zero. For a two-tailed test at α , find a lower critical value (area $1 - \alpha/2$ to its right) and an upper one (area $\alpha/2$ to its right). *Example.* With 21 observations, $s = 5.5\%$ and $H_0 : \sigma = 4.5\%$, $\chi^2 = 20 \times 5.5^2/4.5^2 = 29.88$. With 20 df and $\alpha = 5\%$ the critical values are 9.591 and 34.170, so the null is not rejected.



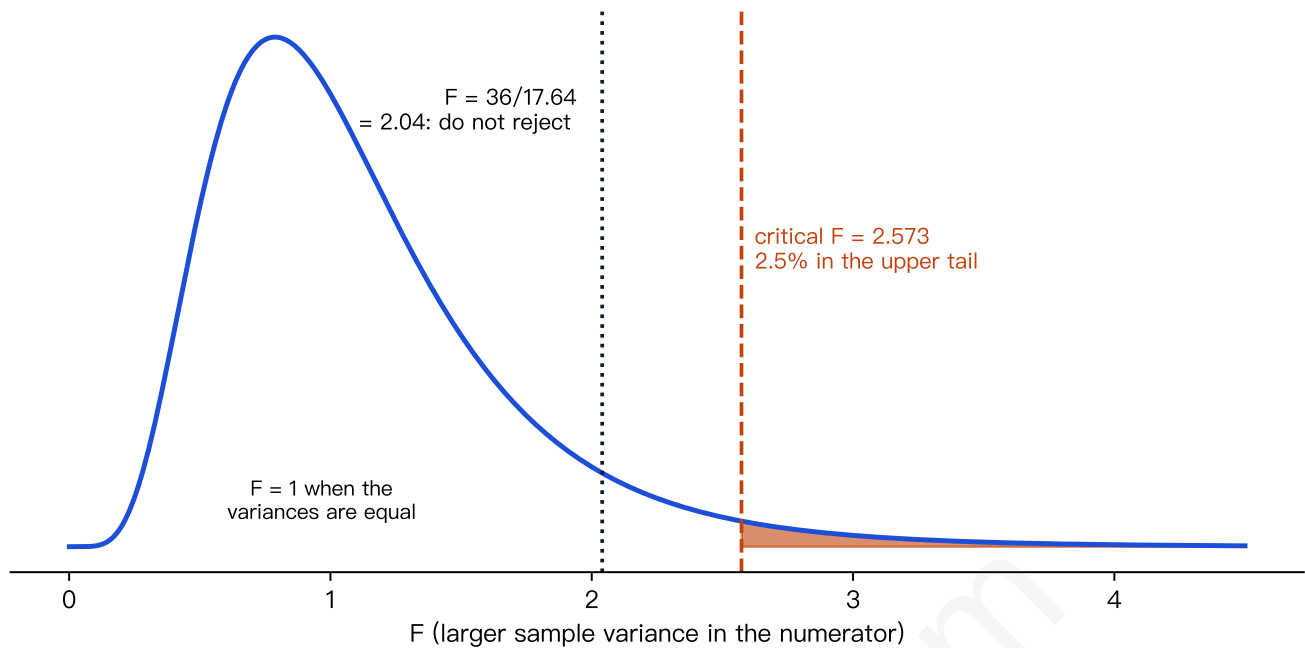
Two-tailed chi-square test of a variance at 5% with 20 degrees of freedom

A hypothesis about a standard deviation is written in terms of the variance. With returns in decimals, a claim that σ is below 3% becomes $H_0 : \sigma^2 \geq 0.0009$ versus $H_a : \sigma^2 < 0.0009$, and only the lower critical value is used.

Test of two variances (F-test)

The **F-distribution** is right-skewed, bounded below by zero and defined by two degrees of freedom (numerator $n_1 - 1$, denominator $n_2 - 1$). $F = 1$ when the sample variances are equal. Putting the larger variance in the numerator means only the upper critical value matters (with $\alpha/2$ in that tail for a two-tailed test).

- $H_0 : \sigma_1^2 = \sigma_2^2$ (the variances are *not* different) vs. $H_a : \sigma_1^2 \neq \sigma_2^2$.
- If the F-statistic exceeds the critical F-value, reject H_0 and conclude that the variances are significantly different. *Example.* With $s_1 = 6.0$ ($n_1 = 16$) and $s_2 = 4.2$ ($n_2 = 21$), $F = 36/17.64 = 2.04$. The critical value $F_{15,20}$ with 2.5% in the upper tail is 2.573, so the null is not rejected.



F-distribution with 15 and 20 degrees of freedom: rejection region above 2.573 and the example statistic 2.04

Test of a correlation coefficient

The **correlation coefficient** (the **Pearson correlation**) gauges how strong a linear relationship is and lies between -1 (perfect negative) and $+1$ (perfect positive); zero means no linear relationship. $H_0 : \rho = 0$ vs. $H_a : \rho \neq 0$. The statistic needs only the sample correlation r and the sample size n ; it follows a t-distribution with $n - 2$ degrees of freedom. It grows with $|r|$ and with n , so a modest correlation can be significant in a large sample: $r = 0.25$ gives $t = 2.56$ with $n = 100$ but only $t = 0.86$ with $n = 13$. *Example.* With $r = 0.30$ and $n = 50$, $t = 0.30\sqrt{48}/\sqrt{0.91} = 2.18$. This exceeds $t_{48, 0.025} = 2.011$, so $\rho = 0$ is rejected.

When a table is given, select the row for $n - 2$ df and the column for the stated (two-tailed) significance level. The null is rejected at a given level only if $|t|$ exceeds that column's value.

p-values and p-hacking

The **p-value** is the smallest significance level at which the null can be rejected (the probability, if H_0 is true, of a statistic at least as extreme as the one observed). Reject H_0 when p-value $< \alpha$. **p-hacking** is manipulating data or the testing process to manufacture a statistically significant result, e.g.:

- testing only the best periods, or deleting inconvenient observations as "outliers";
- **data mining**: running many tests until one rejects the null (with $\alpha = 5\%$, about 1 in 20 tests of true nulls rejects by chance);
- using smoothed (e.g., appraisal-based) data to understate risk.

Because it produces rejections of nulls that are actually true, p-hacking raises the frequency of **Type I errors** (false positives). It is not the same as legitimately choosing a significance level or seeking a more powerful test.

COMMON EXAM TRAPS

- df: correlation test $n - 2$; single mean, paired test and single variance $n - 1$; difference in means $n_1 + n_2 - 2$ (pooled, equal population variances assumed); F-test $n_1 - 1, n_2 - 1$.

- Chi-square tests one variance; F tests two variances; t tests means and correlations.
- The same subjects measured twice call for a paired comparisons test rather than a difference-in-means test.
- Rejecting H_0 in an F-test means concluding the variances are different.

EXAM SHORTCUTS

- Choose the test from what is tested: a mean uses t when the population variance is unknown (z when it is known, or as a large-sample approximation), a test that a correlation is zero uses t when both variables are normally distributed, one variance uses chi-square, and two variances use F.
- With the larger variance in the numerator of the F-statistic, only the upper critical value is needed.
- When a p-value is given, reject H_0 if it is below α ; no critical value is needed.

BOTTOM LINE

- Independent samples with equal variances use the pooled difference-in-means t-test with $n_1 + n_2 - 2$ df; the same subjects measured twice use the paired comparisons test with $n - 1$ df.
- A single variance is tested with $\chi^2 = (n - 1)s^2/\sigma_0^2$ and $n - 1$ df; the chi-square distribution is right-skewed and bounded below by zero.
- Two variances are tested with $F = s_1^2/s_2^2$ and $n_1 - 1, n_2 - 1$ df.
- A correlation is tested with $t = r\sqrt{n-2}/\sqrt{1-r^2}$ and $n - 2$ df; the statistic rises with $|r|$ and with n .
- The p-value is the smallest significance level at which H_0 can be rejected.
- p-hacking raises the frequency of Type I errors.

QUICK CHECK

Question 3

An analyst tests whether a currency strategy's mean daily return is zero and obtains a p-value of 0.021. Which conclusion is *most accurate*?

- A** The null hypothesis is rejected at the 1% significance level.
- B** There is a 2.1% probability that the null hypothesis is true.
- C** The null hypothesis can be rejected at any significance level above 2.1%.

Answer: C

The p-value is the smallest significance level at which the null hypothesis can be rejected. The null is rejected whenever the chosen α exceeds the p-value, so here it is rejected at 5% or at 2.5% but not at 1%.

Why the other options are wrong

- **A.** At 1%, the p-value of 0.021 is above $\alpha = 0.01$, so the null is not rejected.
- **B.** The p-value is computed assuming the null is true: it is the probability, under H_0 , of a test statistic at least as extreme as the one observed. It does not measure the probability that H_0 itself is true.

Key takeaway Reject H_0 when the p-value is below α . A p-value of 0.021 rejects at 5% but not at 1%.

MODULE 7.4

Nonparametric Hypothesis Tests

LOS 7.c — Parametric vs. nonparametric tests

When is a test nonparametric?

	Parametric test	Nonparametric test
What is tested	A parameter of a distribution (mean, variance, correlation)	Something that is not a parameter (ranks, counts in categories), or a parameter when assumptions fail
Assumptions	Specific distribution (e.g., normal population)	Few or no distributional assumptions
Typical data	Values measured on a scale	Ranks or categorical data
Accuracy	Tends to be more accurate when its assumptions hold	More flexible, but tends to be less accurate when the parametric assumptions hold

Use a nonparametric test when (1) the data are ranks or categories, (2) the distributional assumptions of a parametric test do not hold (e.g., a small sample from a clearly non-normal population, so the CLT cannot be relied on), or (3) the question is not about a parameter at all, for example whether two sets of ranks are related or whether a set of observations is consistent with a particular distribution.

Spearman rank correlation test

The **Spearman rank correlation** measures the strength of a *monotonic* relation between two sets of ranks (e.g., a fund's performance rank this year vs. next year). With integer ranks and d_i = difference between the two ranks of item i :

KEY CONCEPT

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

Example. For six funds with rank differences of 1, -1, 0, 2, -1 and -1, $\sum d_i^2 = 8$ and $r_s = 1 - 48/210 = 0.77$.

Significance is tested with the same statistic as a Pearson correlation, $t = r_s \sqrt{n-2} / \sqrt{1-r_s^2}$, which for large samples ($n > 30$) follows a t-distribution with $n - 2$ df. A test of whether performance ranks persist is therefore a **nonparametric test**.

Test of independence using a contingency table

A **contingency table** counts observations classified by two categorical characteristics (e.g., issuer sector × credit-rating change). The **chi-square test of independence** asks whether the two characteristics are related.

- H_0 : the two characteristics are independent. H_a : they are not independent.
- **Expected frequency** of cell (i, j) if independent. Under independence an observation's row says nothing about its column, so each row total is split across the columns in the same proportions as the column totals:

$$E_{i,j} = \frac{(\text{total for row } i) \times (\text{total for column } j)}{\text{total number of observations}}$$

- **Test statistic** (sum over all cells of squared deviations, each scaled by the expected count):

$$\chi^2 = \sum_i \sum_j \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}}, \quad \text{df} = (r - 1)(c - 1)$$

- Reject independence if χ^2 exceeds the critical **chi-square** value (right tail only).
- The **standardized residual** (Pearson residual) of a cell is $(O_{i,j} - E_{i,j}) / \sqrt{E_{i,j}}$; its square is that cell's contribution to χ^2 .
 - Exam convention: each standardized residual is approximately standard normal, and cells with values beyond ± 2 are flagged as contributing significantly to the lack of independence.
 - Current practice: because the row and column totals are estimated from the data, the raw residual has a variance below 1, so ± 2 works as a screening rule of thumb. The *adjusted* residual, which divides $O_{i,j} - E_{i,j}$ by $\sqrt{E_{i,j}(1 - \text{row share})(1 - \text{column share})}$, is the one that is approximately standard normal.

Example. A sample of 220 bonds is classified by issuer size and rating action over a year:

Issuer size	Upgraded	Unchanged	Downgraded	Total
Small	42	38	20	100
Large	18	62	40	120
Total	60	100	60	220

Expected (small, upgraded) = $100 \times 60/220 = 27.27$. Summing $(O - E)^2/E$ over the six cells gives $\chi^2 = 20.38$ with $(2 - 1)(3 - 1) = 2$ df. The 5% critical value is 5.991 (1%: 9.210), so independence is rejected. The standardized residual for (small, upgraded) is $(42 - 27.27)/\sqrt{27.27} = 2.82 > 2$: by the ± 2 rule of thumb, small issuers were upgraded notably more often than independence implies.

COMMON EXAM TRAPS

- The df for a contingency table are $(r - 1)(c - 1)$; $r + c - 2$ and $r \times c$ are common errors. Count only the category rows and columns and leave out the "Total" row and column. For the 2×3 bond table the df are 2; $r + c - 2$ gives 3 and $r \times c$ gives 6.
- Expected counts come from the row and column totals rather than from the observed cell.
- The statistic divides each squared deviation by the expected count; the plain sum of deviations is always zero.
- Rank data (e.g., performance rankings) call for a nonparametric Spearman test; category counts call for a chi-square test on a contingency table.
- Standardized residuals are screened against the ± 2 rule of thumb cell by cell, while the overall statistic is compared with a chi-square critical value.

EXAM SHORTCUTS

- Rank data point to the Spearman test; category counts point to a chi-square test on a contingency table.
- Contingency table df are $(r - 1)(c - 1)$, counting only category rows and columns and leaving out the totals.

BOTTOM LINE

- A nonparametric test is used for ranks or categories, when the assumptions of a parametric test do not hold, or when the question is not about a parameter.
- The Spearman rank correlation is $r_s = 1 - 6 \sum d_i^2 / [n(n^2 - 1)]$, tested in large samples ($n > 30$) with the same t-statistic as a Pearson correlation.
- In a chi-square test of independence the expected count is row total $\times$ column total $\div$ total, $\chi^2 = \sum (O - E)^2 / E$ with $(r - 1)(c - 1)$ df, and only the right tail is used.
- A standardized residual is $(O - E) / \sqrt{E}$, and cells beyond ± 2 are flagged by the rule of thumb.

QUICK CHECK**Question 4**

A credit researcher compiles the following default data for bond issuers in three sectors:

Bond issuers by sector and default status

Sector	No default	Technical default	Actual default	Total
Energy	94	16	15	125
Healthcare	108	9	13	130
Utilities	126	8	16	150
Total	328	33	44	405

Assuming sector and default status are independent, the expected number of energy issuers in technical default is *closest* to:

- A
- B
- C

Answer: A

Under independence, the expected count for a cell is (row total $\times$ column total) / total number of observations. For energy issuers in technical default, use the Energy row total (125) and the Technical default column total (33).

$$E_{i,j} = \frac{\text{total for row } i \times \text{total for column } j}{\text{total number of observations}} = \frac{125 \times 33}{405} = \frac{4,125}{405} = 10.2$$

Why the other options are wrong

- **B.** About 14 (13.6) uses the Actual default column total (44) instead of the Technical default column total: $125 \times 44/405 = 13.6$.
- **C.** 16 is the *observed* number of energy issuers in technical default; the question asks for the count expected under independence.

Key takeaway Expected count = row total × column total / grand total.

Practice questions

Question 5

A multi-asset portfolio is worth €40 million, and its daily returns have a standard deviation of 1.2%. Using the parametric method and assuming a daily mean return of zero, the portfolio's daily 5% value at risk (VaR) is *closest* to:

- A** €0.79 million.
- B** €0.94 million.
- C** €1.12 million.

Answers: 5. A

Why each answer is right and why each other option is wrong: credguild.com/cfa/level-1/notes/quantitative-methods/null-hypothesis/. The full bank has 67 questions on this reading.

QUANTITATIVE METHODS · READING 8

Portfolio Standard Deviation

What you'll learn

LOS 8.a Calculate and interpret the expected return, variance, standard deviation, covariance and correlation of portfolio returns, including multiperiod returns and the effects of portfolio drift and diversification.

LOS 8.b Describe, calculate and interpret the minimum-variance portfolio, the efficient frontier and portfolios that combine a risky portfolio with the risk-free asset.

LOS 8.c Explain optimal portfolio selection given an investor's risk aversion and the capital allocation line, the extension to the market portfolio and the capital market line, and the CAPM.

MODULE 8.1

Portfolio Expected Return

This reading computes a portfolio's expected return, variance and standard deviation from its weights, standard deviations and covariances or correlations, and shows how lower correlation produces a diversification benefit. It then derives the minimum-variance portfolio and the efficient frontier, combines a risky portfolio with a risk-free asset along the capital allocation line, selects an optimal portfolio from an investor's risk aversion and indifference curves, and extends the analysis to the capital market line, beta and the CAPM.

LOS 8.a — Portfolio expected return, historical returns and portfolio drift

Why portfolio statistics matter

Managing a portfolio means working with a small set of statistics: the **expected returns**, **standard deviations** and **correlations** of the holdings. These can be historical (built from realized past returns) or forward-looking (built from forecasts or models). Historical data are a sensible anchor, but a short history can give biased estimates; forecasts can blend history with economic views. Either way, the estimates must be revisited regularly.

Expected return of a portfolio

The **expected return** of a portfolio is the **weighted average** of the expected returns of its assets, where each weight is the asset's share of total portfolio value:

KEY CONCEPT

$$E(R_P) = \sum_{i=1}^N w_i E(R_i), \quad \sum_{i=1}^N w_i = 1$$

Example. A portfolio holds 30% in a fund expected to earn 5% and 70% in a fund expected to earn 9%, so $E(R_P) = 0.30(5\%) + 0.70(9\%) = 7.8\%$.

The **historical return** of a portfolio for one period is the same weighted average applied to realized returns:

$$R_{P,t} = \sum_{i=1}^N w_{i,t-1} R_{i,t}$$

where $w_{i,t-1}$ is the weight at the *start* of period t .

Multiperiod returns: arithmetic vs. geometric

KEY CONCEPT

Measure	Formula	Comment
Sum of period returns (cumulative arithmetic return)	$\sum_{t=1}^T R_{P,t}$	Adds the period returns; ignores compounding
Arithmetic average	$\bar{R}_P = \frac{1}{T} \sum_{t=1}^T R_{P,t}$	Ignores compounding
Cumulative (compound) return	$R_{P,T} = \prod_{t=1}^T (1 + R_{P,t}) - 1$	Assumes income is reinvested and no capital is added or withdrawn
Geometric average	$\bar{R}_{g,P} = \left[\prod_{t=1}^T (1 + R_{P,t}) \right]^{1/T} - 1$	Compound growth rate per period

If the portfolio is not rebalanced, the cumulative return over the whole period equals the weighted average of each asset's cumulative return using the initial weights: $R_{P,T} = \sum_i w_{i,0} \left[\prod_{t=1}^T (1 + R_{i,t}) - 1 \right]$. With or without drift, the formulas in the table are applied to the period-by-period portfolio returns $R_{P,t}$.

Portfolio drift and style drift

Market moves change the weights: winners grow and losers shrink. This is **portfolio drift**. With no rebalancing, each period's return must be computed with the new (drifted) weights, which come from the start-of-period market values.

Style drift is a portfolio moving away from its stated strategy. It can come from portfolio drift (a strong equity market pushes a balanced portfolio overweight in stocks) or from changes in the holdings themselves (a small-cap stock that grows into the mid-cap category). Portfolio drift can sometimes offset style drift, but managers still monitor and rebalance to keep a portfolio aligned with its goals.

KEY CONCEPT

How drift changes a portfolio depends on the starting weights.

- Exam convention: in an equal-weighted portfolio, drift raises the weights of the high-return securities, so performance tilts toward higher-growth or higher-volatility assets; in a portfolio that is not equally weighted, drift tends to enlarge the already large allocations, so the portfolio behaves like a **momentum** strategy and its **concentration risk** may rise.
- Current practice: in any unrebalanced portfolio, whichever assets outperform gain weight and underperformers lose weight. Whether concentration or overall risk rises depends on *which* assets outperform: if the largest positions (or the more volatile assets) do best, concentration and risk

increase; if smaller positions outperform, the weights can move closer together (e.g., an 80/20 portfolio whose assets return 0% and 100% drifts to about 67/33). A high realized return does not by itself mean the asset is high-volatility.

WORKED EXAMPLE — RETURNS WITH CHANGING WEIGHTS

A client invests \$50,000 in Asset A and \$50,000 in Asset B and never rebalances. Returns: A earns 10% in Period 1 and -4% in Period 2; B earns 2% and then 8%.

1. Period 1 return = $0.50(10\%) + 0.50(2\%) = 6.00\%$. End values: A = \$55,000, B = \$51,000 (total \$106,000).
2. New weights: A = $55,000/106,000 = 51.89\%$; B = 48.11%.
3. Period 2 return = $0.5189(-4\%) + 0.4811(8\%) = 1.77\%$.
4. Cumulative return = $(1.06)(1.0177) - 1 = 7.88\%$ (check: end value \$52,800 + \$55,080 = \$107,880). Arithmetic average = $(6.00\% + 1.77\%)/2 = 3.89\%$; the simple sum of the two period returns is 7.77%, below the compound 7.88%.

COMMON EXAM TRAPS

- Expected return is a weighted average of expected returns only; do not plug in variances or correlations.
- Applying the original weights to a later period of an unbalanced portfolio gives the wrong return: 2.00% instead of 1.77% for Period 2 in the example.
- A holding that grows from small cap into the mid-cap category causes style drift; portfolio drift refers only to weights shifting with market values.

EXAM SHORTCUTS

- For a portfolio that is not rebalanced, the cumulative return equals the initial-weight average of each asset's cumulative return, so the cumulative figure needs no period-by-period weights.

BOTTOM LINE

- A portfolio's expected return is the weighted average $E(R_P) = \sum w_i E(R_i)$, with each weight equal to the asset's share of total portfolio value and the weights summing to 1.
- A portfolio's historical return for one period is $\sum w_{i,t-1} R_{i,t}$, using the weights at the start of the period.
- The cumulative (compound) return $\prod (1 + R_{P,t}) - 1$ and the geometric average include compounding, while the sum and the arithmetic average of the period returns ignore it.
- Without rebalancing, market moves shift the weights (portfolio drift), so each period's return must use the drifted weights from start-of-period market values.
- Style drift is a portfolio moving away from its stated strategy, either through portfolio drift or through changes in the holdings themselves, such as a small-cap stock growing into the mid-cap category.

- Exam convention: drift raises the weights of high-return securities in an equal-weighted portfolio and enlarges already large allocations in a portfolio that is not equally weighted, so it behaves like a momentum strategy with possibly higher concentration risk; current practice: in any unbalanced portfolio the outperformers gain weight, and whether concentration or risk rises depends on which assets outperform.

QUICK CHECK

Question 1

Nadia Okafor builds the two-fund portfolio shown in the table below and does not rebalance it at any point. Okafor's portfolio return for Year 2 is *closest* to:

Okafor portfolio: starting values and annual returns

Holding	Value at start of Year 1	Return in Year 1	Return in Year 2
Equity fund	\$60,000	20%	-10%
Bond fund	\$40,000	2%	6%

- A -2.00%.
- B -3.60%.
- C -4.21%.

Answer: C

Without rebalancing, the weights drift with market values (portfolio drift). Each period's portfolio return must use the weights at the start of that period, so the Year 2 return is computed with the weights at the end of Year 1, not the original 60/40 split.

1. Values at the end of Year 1

Equity fund: $\$60,000 \times 1.20 = \$72,000$; bond fund: $\$40,000 \times 1.02 = \$40,800$; total = $\$112,800$.

2. Weights at the start of Year 2

$$w_{\text{equity}} = \frac{72,000}{112,800} = 0.6383, \quad w_{\text{bond}} = \frac{40,800}{112,800} = 0.3617$$

3. Year 2 portfolio return

$$R_{P,2} = 0.6383(-10\%) + 0.3617(6\%) = -6.383\% + 2.170\% = -4.21\%$$

Check: end values $\$64,800 + \$43,248 = \$108,048$; $108,048/112,800 - 1 = -4.21\%$.

Why the other options are wrong

- A. -2.00% is the equal-weighted average of the two Year 2 returns, $(-10\% + 6\%)/2$; it ignores both the original and the drifted weights.

- **B.** -3.60% uses the original weights, $0.60(-10\%) + 0.40(6\%)$. That would be right only if the portfolio had been rebalanced back to 60/40 at the end of Year 1.

Key takeaway No rebalancing means the weights drift: recompute each weight from start-of-period market values before weighting the period's returns.

MODULE 8.2

Portfolio Risk Measures

LOS 8.a — Variance and standard deviation of a portfolio

Portfolio variance

Risk is measured by how far returns scatter around their mean: the **variance** and its square root, the **standard deviation**. Because the standard deviation is in the same units as returns, it is the usual headline risk number.

For N assets, portfolio variance depends on the weights, each asset's variance and every pair's **covariance**:

$$\sigma_P^2 = \sum_{i=1}^N \sum_{j=1}^N w_i w_j \text{Cov}(R_i, R_j)$$

For two assets the formula becomes:

KEY CONCEPT

$$\sigma_P^2 = w_1^2 \sigma_1^2 + w_2^2 \sigma_2^2 + 2w_1 w_2 \text{Cov}_{1,2}$$

$$\sigma_P = \sqrt{\sigma_P^2}$$

Covariance and correlation

Covariance is often given indirectly through the **correlation** (correlation coefficient) $\rho_{1,2}$, which is bounded between -1 and +1:

KEY CONCEPT

$$\text{Cov}_{1,2} = \rho_{1,2} \sigma_1 \sigma_2 \implies \sigma_P^2 = w_1^2 \sigma_1^2 + w_2^2 \sigma_2^2 + 2w_1 w_2 \rho_{1,2} \sigma_1 \sigma_2$$

For forward-looking (expected) values the identity does not hold mechanically, because joint variability and nonlinear interactions among the three inputs also matter. Exam calculations use the identity above.

Special cases

KEY CONCEPT

Correlation	Portfolio standard deviation (two assets, positive weights)	Diversification benefit
$\rho = +1$	$\sigma_P = w_1\sigma_1 + w_2\sigma_2$ (weighted average)	None
$-1 < \rho < +1$	Below the weighted average	Some; larger as ρ falls
$\rho = -1$	$\sigma_P = w_1\sigma_1 - w_2\sigma_2 $	Maximum; zero risk only when $w_1\sigma_1 = w_2\sigma_2$

WORKED EXAMPLE

A portfolio is 20% in Asset J (standard deviation 25%) and 80% in Asset K (standard deviation 10%); the correlation is 0.20.

$$\begin{aligned}\sigma_P^2 &= (0.20)^2(0.25)^2 + (0.80)^2(0.10)^2 + 2(0.20)(0.80)(0.20)(0.25)(0.10) \\ &= 0.0025 + 0.0064 + 0.0016 = 0.0105 \quad \sigma_P = \sqrt{0.0105} = 10.25\%\end{aligned}$$

The weighted average of the two standard deviations is $0.20(25\%) + 0.80(10\%) = 13\%$; the gap (13% vs. 10.25%) is the diversification benefit.

WORKED EXAMPLE — COVARIANCE GIVEN DIRECTLY, THREE-ASSET EXTENSION

If the covariance is given, use it as is. Two assets with $\sigma_1^2 = 0.0400$, $\sigma_2^2 = 0.0225$ and $\text{Cov}_{1,2} = -0.0030$, held 60/40:

$$\begin{aligned}\sigma_P^2 &= 0.36(0.0400) + 0.16(0.0225) + 2(0.60)(0.40)(-0.0030) = 0.0144 + 0.0036 - 0.00144 = \\ &0.01656, \\ \sigma_P &= 12.87\%\end{aligned}$$

The negative covariance makes the third term negative, so the portfolio is far less risky than a 60/40 average of the standard deviations ($0.6 \times 20\% + 0.4 \times 15\% = 18\%$) would suggest. With three assets, the variance has three weighted-variance terms and three covariance terms, each doubled:

$$\sigma_P^2 = \sum_{i=1}^3 w_i^2 \sigma_i^2 + 2w_1w_2\text{Cov}_{1,2} + 2w_1w_3\text{Cov}_{1,3} + 2w_2w_3\text{Cov}_{2,3}$$

Calculator tip

Compute each of the three terms separately, add them, then take the square root last. Keep returns in decimals throughout (0.25 rather than 25) or in percent throughout; mixing the two gives wrong answers.

COMMON EXAM TRAPS

- Forgetting the factor 2 in the covariance term. In the J/K example, this gives $\sqrt{0.0097} = 9.85\%$ instead of 10.25%.
- Using variances instead of standard deviations inside $\rho\sigma_1\sigma_2$. In the J/K example, $0.20 \times 0.0625 \times 0.0100$ as the covariance gives a variance of 0.00894 and a standard deviation of 9.46%, instead of 10.25%.
- Reporting the variance as if it were the standard deviation (e.g., 0.0105 read as "1.05%").
- Averaging the standard deviations when $\rho < 1$: that is correct only for perfect positive correlation. In the J/K example, this gives 13% instead of 10.25%.
- Assuming $\rho = -1$ always means zero risk: it does only at one specific set of weights.

EXAM SHORTCUTS

- For two assets with positive weights, the portfolio standard deviation equals the weighted average of the two standard deviations only when $\rho = +1$ and is below it for any lower correlation, so with $\rho < +1$ an answer at or above the weighted average can be ruled out.

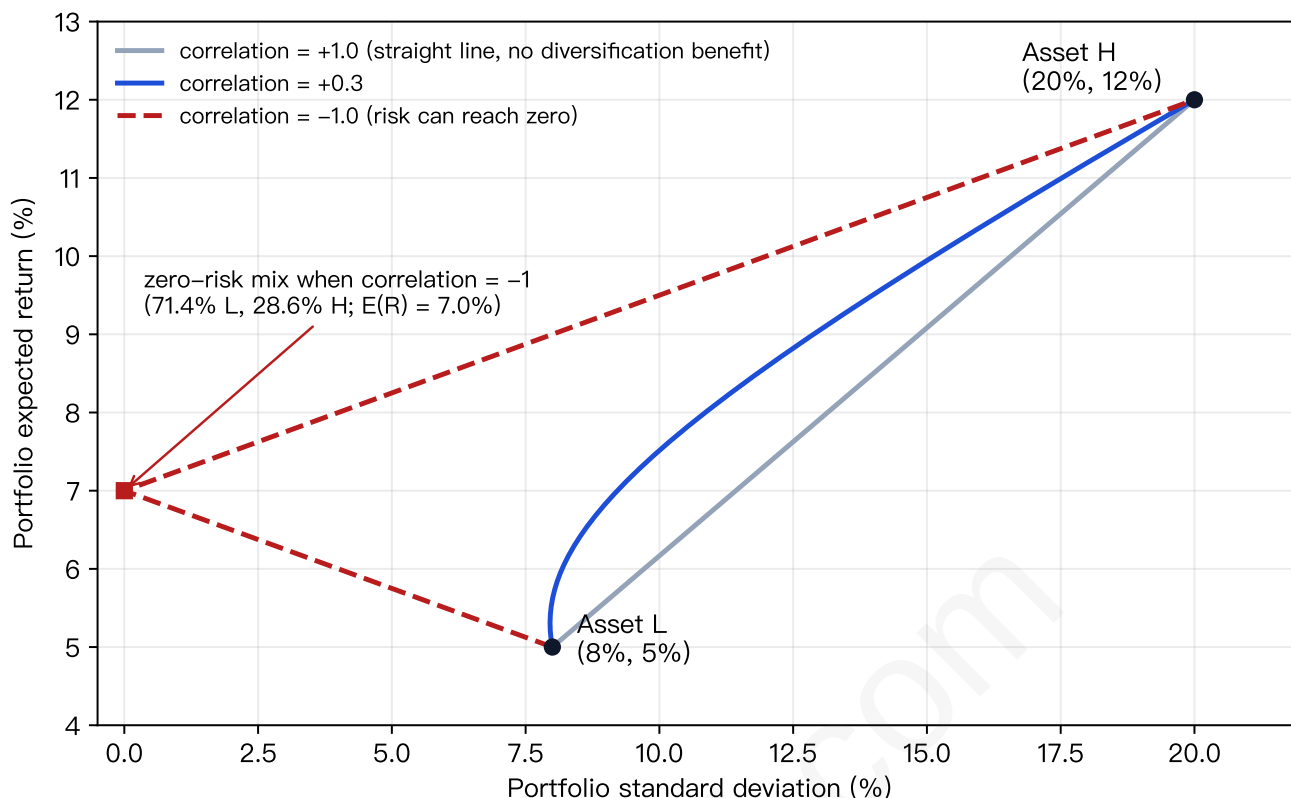
BOTTOM LINE

- Two-asset portfolio variance is $\sigma_P^2 = w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\text{Cov}_{1,2}$, and the portfolio standard deviation is its square root.
- Covariance can be written $\text{Cov}_{1,2} = \rho_{1,2}\sigma_1\sigma_2$, which exam calculations use, although for forward-looking expected values the identity does not hold mechanically.
- With positive weights, a correlation of +1 gives a portfolio standard deviation equal to the weighted average of the standard deviations (no diversification benefit), and any lower correlation gives less, with a larger benefit as ρ falls.
- With $\rho = -1$, $\sigma_P = |w_1\sigma_1 - w_2\sigma_2|$, which is zero only when $w_1\sigma_1 = w_2\sigma_2$.
- A negative covariance makes the cross term negative, so the portfolio is much less risky than the weighted average of the standard deviations suggests.

MODULE 8.3

Correlation and Diversification Benefits**LOS 8.a — Correlation, diversification and portfolios of many assets****Correlation drives the diversification benefit**

Changing the weights of two assets traces out a curve of attainable risk–return combinations. When the assets are less than perfectly correlated, the curve bows to the left: moving from the riskier asset toward the safer one cuts risk faster than it cuts return. This is the **diversification benefit**, and it grows as correlation falls. At $\rho = +1$ the combinations lie on a straight line; at $\rho = -1$ one mix has zero risk.



Two-asset portfolios under different return correlations

Of all the feasible mixes, the efficient ones are those with the highest return for their level of risk, which lie on the upper part of the curve.

More assets: the variance-covariance matrix

With N assets, the variance-covariance matrix has N variances on its diagonal and

KEY CONCEPT

$$\text{number of unique covariances} = \frac{N^2 - N}{2}$$

So 20 assets require $(400 - 20)/2 = 190$ covariances. Normalizing the covariance matrix gives a **correlation matrix**, which helps an investor spot the pairs with the lowest correlations (the most useful diversifiers). Using correlations, the general formula is

$$\sigma_P^2 = \sum_{i=1}^N \sum_{j=1}^N w_i w_j \rho_{i,j} \sigma_i \sigma_j, \quad \rho_{i,i} = 1$$

Example. A portfolio holds 40% in Asset 1, 35% in Asset 2 and 25% in Asset 3, with standard deviations of 16%, 10% and 24%. The correlations are $\rho_{1,2} = 0.25$, $\rho_{1,3} = 0.60$ and $\rho_{2,3} = -0.20$. Each covariance is $\rho_{i,j} \sigma_i \sigma_j$ (for example, $0.25 \times 0.16 \times 0.10 = 0.0040$), and the diagonal holds the variances:

Covariance matrix	Asset 1	Asset 2	Asset 3
Asset 1	0.0256	0.0040	0.02304
Asset 2	0.0040	0.0100	-0.0048
Asset 3	0.02304	-0.0048	0.0576

	Asset 1	Asset 2	Asset 3	
Asset 1	σ_1^2 0.0256	Cov1,2 0.0040	Cov1,3 0.02304	Diagonal (blue): 3 variances Above the diagonal (orange): 3 unique covariances Below the diagonal: the same covariances again, so each pair gets the factor 2 in σ_p^2 n assets: $n(n-1)/2$ unique covariances (3 here)
Asset 2	Cov1,2 0.0040	σ_2^2 0.0100	Cov2,3 -0.0048	
Asset 3	Cov1,3 0.02304	Cov2,3 -0.0048	σ_3^2 0.0576	

The three-asset covariance matrix: variances on the diagonal, each covariance appearing twice

Each pair appears twice in the matrix, which is where the factor 2 in the cross terms comes from:

$$\sigma_P^2 = (0.004096 + 0.001225 + 0.0036) + (0.00112 + 0.004608 - 0.00084) = 0.013809, \quad \sigma_P = 11.75\%$$

The first bracket holds the weighted variances $w_i^2 \sigma_i^2$ and the second the terms $2w_i w_j \text{Cov}_{i,j}$. The result is well below the 15.9% weighted average of the three standard deviations because every correlation is below +1.

Expected portfolio variance from scenarios

A forward-looking variance can be built from economic scenarios m with probabilities p_m that sum to 1:

1. Portfolio return in each scenario: $R_{P,m} = \sum_i w_i R_{i,m}$.
2. Expected portfolio return: $E(R_P) = \sum_m p_m R_{P,m}$, which equals the weighted average of the assets' expected returns, $\sum_i w_i E(R_i)$.
3. Expected variance: $\sigma_P^2 = \sum_m p_m [R_{P,m} - E(R_P)]^2$, then take the square root.

Example. Two equally likely scenarios give portfolio returns of 12% and 2%. Then $E(R_P) = 7\%$ and $\sigma_P^2 = 0.5(5)^2 + 0.5(-5)^2 = 25$ (in %²), so $\sigma_P = 5\%$.

Large portfolios: variance approaches average covariance

For an **equally weighted** portfolio of N assets, the exact variance is

KEY CONCEPT

$$\sigma_P^2 = \frac{\bar{\sigma}^2}{N} + \frac{N-1}{N} \overline{\text{Cov}}$$

where $\bar{\sigma}^2$ is the average variance and $\overline{\text{Cov}}$ the average pairwise covariance. As N grows, each asset's own variance matters less and less and the portfolio variance approaches the **average covariance**. For a large portfolio, the approximation writes this as the **average correlation** times the **average variance**:

$$\sigma_P^2 \approx \bar{\rho} \bar{\sigma}^2$$

- Exam convention: for an equally weighted portfolio, a more exact variance is $\sigma_P^2 = \bar{\sigma}^2/N + (1 - 1/N) \bar{\rho} \bar{\sigma}^2$, and $\bar{\rho} \bar{\sigma}^2$ is its large- N approximation.
- Current practice: that more exact form is exact only when all assets have the same variance, because only then does $\bar{\rho} \bar{\sigma}^2$ equal the average covariance. In general the exact variance is $\bar{\sigma}^2/N + (N-1)/N$ times the average covariance, as above.

With unequal variances the gap can be large. For example, two assets with standard deviations of 10% and 30% and a correlation of 0.5 have a covariance of 0.015, but $\bar{\rho} \bar{\sigma}^2 = 0.5 \times 0.05 = 0.025$. *Example.* With 40 assets, an average variance of 0.05 and an average correlation of 0.25, $\sigma_P^2 \approx 0.25 \times 0.05 = 0.0125$, $\sigma_P \approx 11.2\%$. If all 40 assets had the same variance of 0.05, the finite- N term would add $0.75 \times 0.05/40 = 0.0009$, giving 0.0134 (11.6%).

Systematic vs. unsystematic risk

KEY CONCEPT

	Unsystematic risk	Systematic risk
Other names	Unique, diversifiable, firm-specific	Market, nondiversifiable
Removed by diversification?	Yes	No
Rewarded with higher expected return?	No	Yes

A portfolio of roughly 20 to 30 well-chosen assets captures most of the diversification benefit; beyond about 30 assets the extra benefit of each new holding is close to zero, and what remains is systematic risk.

COMMON EXAM TRAPS

- The approximation gives a variance; take the square root for standard deviation. In the 40-asset example, reading 0.0125 as 1.25% instead of taking the square root, about 11.2%.
- Number of covariances is $(N^2 - N)/2$, not N^2 . For 20 assets, N^2 gives 400 instead of 190.

- Adding assets reduces risk only to the extent that their correlations are below +1; with correlations of +1, more assets bring no diversification benefit.

EXAM SHORTCUTS

- The expected portfolio return from scenarios equals $\sum w_i E(R_i)$, the assets' expected returns weighted by portfolio weight, so it can be found without the scenario portfolio returns; the variance still needs them.

BOTTOM LINE

- The diversification benefit grows as correlation falls: two-asset combinations lie on a straight line at $\rho = +1$, bow to the left when correlation is below +1, and include one zero-risk mix at $\rho = -1$.
- With N assets the variance-covariance matrix holds N variances and $(N^2 - N)/2$ unique covariances.
- Expected portfolio variance from scenarios is $\sum_m p_m [R_{P,m} - E(R_P)]^2$, with scenario probabilities that sum to 1.
- An equally weighted portfolio of N assets has variance $\bar{\sigma}^2/N + [(N - 1)/N] \overline{\text{Cov}}$, which approaches the average covariance as N grows and is approximated for a large portfolio by $\bar{\rho} \bar{\sigma}^2$.
- Unsystematic risk is removed by diversification and is not rewarded with higher expected return; systematic risk is not removed and is rewarded.
- Roughly 20 to 30 well-chosen assets capture most of the diversification benefit, and beyond about 30 assets each new holding adds close to none.

QUICK CHECK

Question 2

Brookline Endowment holds an equally weighted portfolio of 50 stocks. The average variance of returns of the individual stocks is 0.09, and the average correlation between pairs of stocks is 0.30. Using the approximation for a portfolio with a large number of assets, the standard deviation of the portfolio's returns is *closest* to:

- A 2.7%.
- B 9.0%.
- C 16.4%.

Answer: C

For an equally weighted portfolio of many assets, the individual variances matter little and portfolio variance is approximately the average correlation times the average variance. The standard deviation is the square root of that variance.

$$\sigma_P^2 \approx \bar{\rho} \bar{\sigma}^2 = 0.30 \times 0.09 = 0.027$$

$$\sigma_P \approx \sqrt{0.027} = 0.164 = 16.4\%$$

The approximation ignores the finite- N term. If every stock had the same variance of 0.09, the exact equal-weight formula would add $(1 - \bar{\rho})\sigma^2/N = 0.70(0.09)/50 = 0.00126$, giving $\sigma_P^2 = 0.02826$ and $\sigma_P = 16.8\%$, still closest to 16.4%. With unequal variances, the exact result depends on the average covariance, which the given averages do not pin down, so the question asks for the approximation. The average single stock has a standard deviation of $\sqrt{0.09} = 30\%$; diversification cuts this to about 16%, and the remainder is systematic risk.

Why the other options are wrong

- **A.** 2.7% is the approximate portfolio *variance* (0.027) quoted as if it were the standard deviation.
- **B.** 9.0% multiplies the average correlation by the average standard deviation ($0.30 \times 30\%$). The approximation multiplies the correlation by the average *variance*, and the square root is taken afterward.

Key takeaway For a large portfolio, $\sigma_P^2 \approx \bar{\rho}\bar{\sigma}^2$. Diversification removes unsystematic risk, but the average covariance (systematic risk) remains.

MODULE 8.4

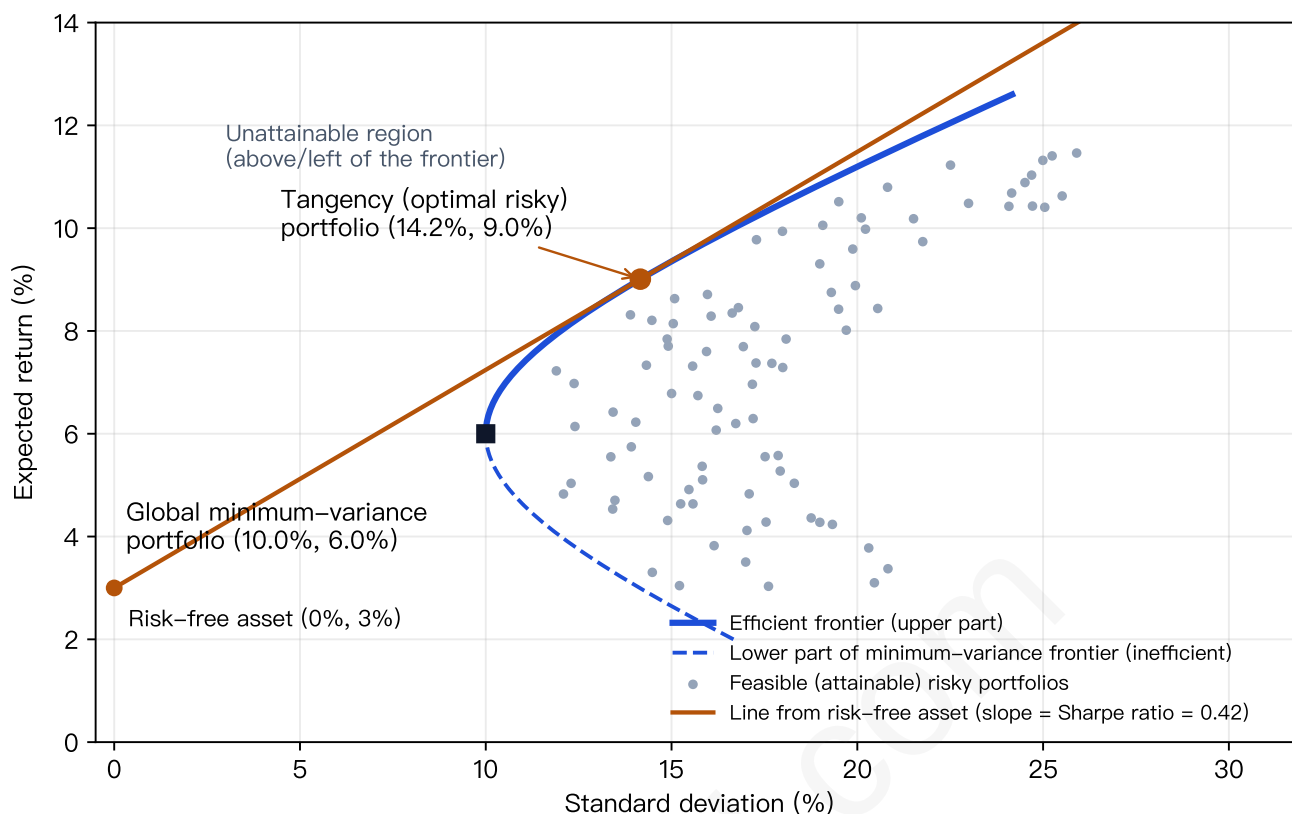
The Minimum-Variance Portfolio and the Efficient Frontier

LOS 8.b — The minimum-variance portfolio and the efficient frontier

Feasible set, minimum-variance frontier and efficient frontier

Every mix of the available risky assets is a **risky portfolio**; together they form the set of feasible (attainable) portfolios.

- The **minimum-variance frontier** is the set of portfolios with the lowest risk for each level of expected return. It is the left boundary of the feasible set.
- The **global minimum-variance portfolio** is the point on that frontier with the lowest risk of all.
- The **efficient frontier** is the set of portfolios that offer the most expected return at any given level of risk. On a graph it is the upper part of the minimum-variance frontier, beginning at the global minimum-variance portfolio and rising from there. The lower part is inefficient: another portfolio offers more return for the same risk.
- Points to the left of the minimum-variance frontier, including every point above the efficient frontier, are **unattainable** with the available assets. Attainable points below the efficient frontier are **inefficient**.



Feasible portfolios, minimum-variance frontier, efficient frontier and the tangency portfolio

In a theoretical universe of all investable assets, the upper segment is called the **Markowitz efficient frontier**.

Minimum-variance portfolio of two risky assets

KEY CONCEPT

$$w_A = \frac{\sigma_B^2 - \text{Cov}_{A,B}}{\sigma_A^2 + \sigma_B^2 - 2\text{Cov}_{A,B}}, \quad w_B = 1 - w_A$$

When both weights are positive and the other inputs are held constant, a higher variance of an asset lowers its weight in the minimum-variance portfolio, and a lower variance raises it.

Example. With $\sigma_A = 15\%$, $\sigma_B = 25\%$ and $\rho = 0.20$, $\text{Cov} = 0.20(0.15)(0.25) = 0.0075$.

$$w_A = \frac{0.0625 - 0.0075}{0.0225 + 0.0625 - 0.0150} = \frac{0.0550}{0.0700} = 78.6\%, \quad w_B = 21.4\%$$

The less risky asset gets the larger weight. Expected returns do not enter the formula.

Adding a risk-free asset

A **risk-free asset** has a return with zero variance. Exam convention: the risk-free asset has zero correlation with any risky asset. Current practice: because its return never varies, its covariance with every asset is zero and the correlation coefficient is, strictly, undefined. Either way the covariance term drops out of the portfolio variance.

The straight line from the risk-free rate that just touches (is tangent to) the efficient frontier identifies the **optimal risky portfolio**, also called the **tangency portfolio**: the efficient portfolio with the highest **Sharpe ratio**,

KEY CONCEPT

$$\text{Sharpe ratio} = \frac{E(R_P) - R_f}{\sigma_P}$$

Combining the risk-free asset with a risky portfolio P :

$$E(R_C) = w_P E(R_P) + (1 - w_P) R_f, \quad \sigma_C = w_P \sigma_P$$

To reach a target return, solve for the risky weight:

$$w_P = \frac{r_{\text{target}} - R_f}{E(R_P) - R_f}$$

Example. With $R_f = 4\%$, a risky portfolio with a 12% expected return and a 20% standard deviation, and a 10% target, $w_P = 6/8 = 0.75$ and $\sigma_C = 0.75 \times 20\% = 15\%$. Weights between 0 and 100% mean lending at the risk-free rate; a risky weight above 100% means borrowing at the risk-free rate.

Portfolio optimization

KEY CONCEPT

Approach	What it does
Return-constrained optimization	Minimizes risk for a given level of return
Risk-constrained optimization	Maximizes expected return for a given level of risk (finds points on the efficient frontier)
Risk-adjusted return optimization	Maximizes the Sharpe ratio

The key inputs are **expected returns**, **volatilities** (standard deviations) and **asset correlations**. Investor-specific factors (risk tolerance, time horizon, cash needs) and constraints such as no short selling or position limits can move weights away from the unconstrained optimum.

COMMON EXAM TRAPS

- In the minimum-variance formula, use variances (σ^2) rather than standard deviations, and keep the covariance term. In the example, standard deviations give $(0.25 - 0.0075)/(0.15 + 0.25 - 0.015) = 63.0\%$ instead of 78.6%.
- With a risk-free asset, combined risk is just $w_P \sigma_P$; do not add a covariance term.

- The risky weight for a target return is $(r_{\text{target}} - R_f)/(E(R_P) - R_f)$; dividing the target by $E(R_P)$ ignores the return on the risk-free asset. In the example, $10/12 = 0.833$ instead of 0.75.

EXAM SHORTCUTS

- Expected returns do not enter the two-asset minimum-variance weights, so any expected returns given in such a question can be set aside.

BOTTOM LINE

- The minimum-variance frontier holds the lowest-risk portfolio for each level of expected return, and the global minimum-variance portfolio is its point of lowest risk.
- The efficient frontier is the upper part of the minimum-variance frontier from the global minimum-variance portfolio upward; points to the left of the frontier are unattainable and attainable points below the efficient frontier are inefficient.
- For two risky assets, $w_A = (\sigma_B^2 - \text{Cov}_{A,B})/(\sigma_A^2 + \sigma_B^2 - 2\text{Cov}_{A,B})$ and $w_B = 1 - w_A$ give the minimum-variance portfolio.
- The optimal risky (tangency) portfolio is where a line from the risk-free rate touches the efficient frontier, the efficient portfolio with the highest Sharpe ratio $(E(R_P) - R_f)/\sigma_P$.
- Combining the risk-free asset with portfolio P gives $E(R_C) = w_P E(R_P) + (1 - w_P)R_f$ and $\sigma_C = w_P \sigma_P$, with $w_P = (r_{\text{target}} - R_f)/(E(R_P) - R_f)$ for a target return and a weight above 100% meaning borrowing at the risk-free rate.
- Return-constrained optimization minimizes risk for a given return, risk-constrained optimization maximizes expected return for a given risk, and risk-adjusted return optimization seeks the highest Sharpe ratio.

MODULE 8.5

Risk Aversion

LOS 8.c — Risk aversion, utility and indifference curves

Assumptions behind optimal portfolio choice

- Investors are rational and **risk averse**; they maximize satisfaction (**utility**) and prefer more return and less risk.
- A single-period time horizon.
- No taxes or transaction costs.
- Investors can borrow and lend at a constant **risk-free rate**.

Three attitudes toward risk

KEY CONCEPT

Attitude	Behavior	Coefficient A	Indifference curves
Risk seeking	Prefers an uncertain outcome to a certain one with the same expected value; may accept a lower expected return (e.g., casino gambling)	$A < 0$	Slope downward
Risk neutral	Indifferent between a gamble and a sure amount with the same expected value; cares only about return	$A = 0$	Flat (horizontal)
Risk averse	Prefers the certain outcome; needs extra return to bear extra risk	$A > 0$	Slope upward

A **fair gamble** (expected gain of zero) raises the expected utility of a risk-seeking investor, leaves a risk-neutral investor indifferent, and lowers the utility of a risk-averse investor. Most real-world investors are risk averse; risk neutrality is a useful theoretical case (e.g., large institutions making many small trades whose variance averages out).

Utility function

A common **utility function** is

KEY CONCEPT

$$U = E(r) - \frac{1}{2}A\sigma^2$$

where A is the investor's **coefficient of risk aversion** and σ^2 the variance of returns (use decimals). A higher A means a bigger utility penalty for risk. Portfolio theory assumes every investor is risk averse ($A > 0$). Research suggests A typically lies between 2 and 3 and is rarely below 1. **Mean-variance utility**, $U_P = U[E(R_P), \sigma_P]$, depends only on expected return and risk and ignores other factors such as total wealth.

Example. With $A = 4$, an expected return of 10% and a standard deviation of 15%, $U = 0.10 - 0.5(4)(0.0225) = 0.055$, i.e., 5.5%. The investor would be equally happy with a certain 5.5% return.

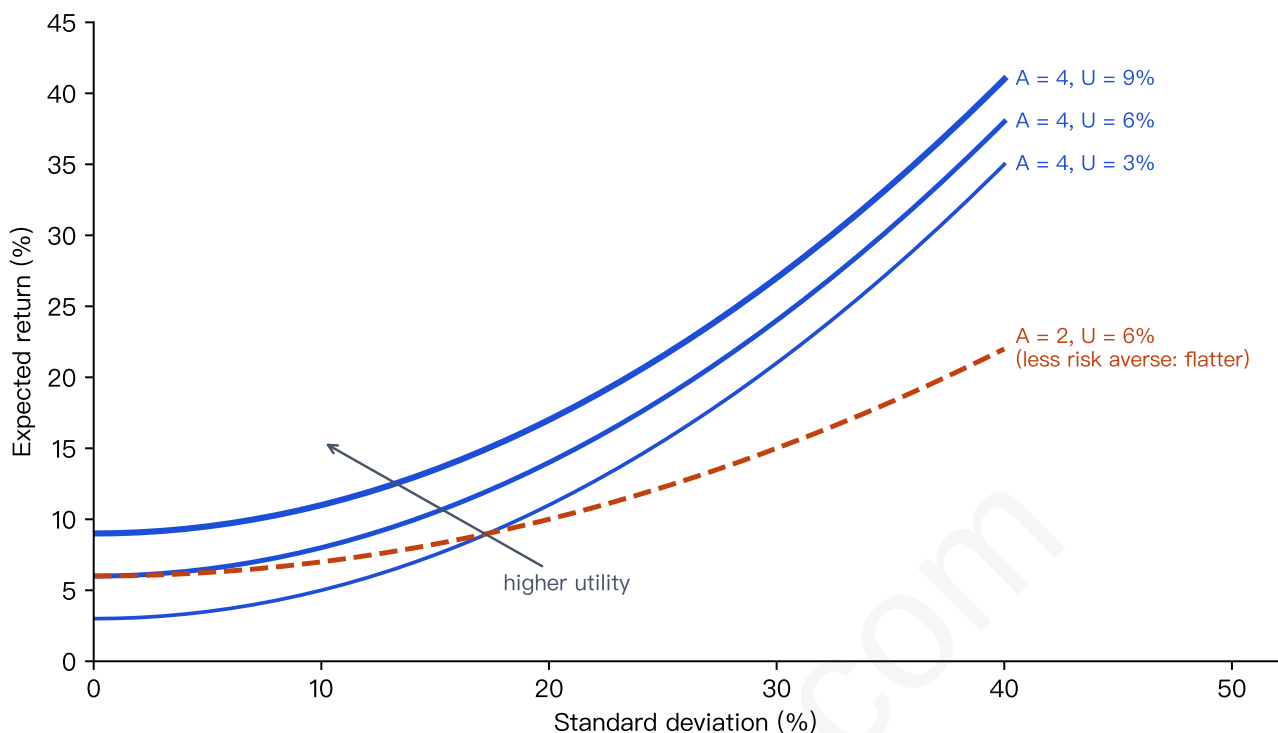
Risk tolerance is the willingness to accept risk for higher expected return; it is the inverse of risk aversion and can change with age, wealth and economic conditions, and differs across countries and cultures. **Absolute risk aversion** looks at how utility changes with extra risk holding wealth constant; **relative risk aversion** looks at how the *share* of wealth put at risk changes as wealth changes.

Indifference curves

An **indifference curve** plots the risk–return combinations that give an investor the same expected utility. For a risk-averse investor the curves slope upward and are convex; curves higher and to the left represent higher utility. The slope measures how much extra return is required per unit of extra risk:

- A more risk-averse (less risk-tolerant) investor has steeper curves.

- A less risk-averse (more risk-tolerant) investor has flatter curves.



Indifference curves: higher curves give higher utility, and a less risk-averse investor has flatter curves (illustrative utility levels)

COMMON EXAM TRAPS

- The utility function uses the variance rather than the standard deviation, together with the factor $\frac{1}{2}$. In the example, using the standard deviation gives $0.10 - 0.5(4)(0.15) = -0.20$ and dropping the $\frac{1}{2}$ gives 0.010, instead of 0.055.
- The highest expected return or the lowest risk is not necessarily the highest-utility choice; compute U for each.

BOTTOM LINE

- Optimal portfolio theory assumes rational, risk-averse investors who maximize utility, a single-period horizon, no taxes or transaction costs, and borrowing and lending at a constant risk-free rate.
- A risk-seeking investor has $A < 0$ and downward-sloping indifference curves, a risk-neutral investor has $A = 0$ and flat curves, and a risk-averse investor has $A > 0$ and upward-sloping curves.
- Utility is $U = E(r) - \frac{1}{2}A\sigma^2$ with the variance in decimals, so a higher coefficient of risk aversion A means a bigger penalty for risk.
- Risk tolerance, the inverse of risk aversion, can change with age, wealth and economic conditions.
- A risk-averse investor's indifference curves slope upward and are convex, curves higher and to the left give higher utility, and a more risk-averse investor has steeper curves.

QUICK CHECK

Question 3

An adviser measures a client's utility with the function $U = E(r) - \frac{1}{2}A\sigma^2$ and estimates the client's coefficient of risk aversion, A , at 3. Based on the table below, which fund gives this client the highest utility?

Funds under consideration

Fund	Expected return	Standard deviation of returns
Northgate Growth Fund	11%	20%
Westbury Balanced Fund	8%	12%
Eastholm Income Fund	6%	5%

- A Northgate Growth Fund.
- B Westbury Balanced Fund.
- C Eastholm Income Fund.

Answer: B

Utility rewards expected return and penalizes variance in proportion to the coefficient of risk aversion. Computing the utility of each fund (with returns and standard deviations in decimals and the standard deviation squared) shows that the balanced fund offers the best trade-off for this client.

$$U = E(r) - \frac{1}{2}(3)\sigma^2 = E(r) - 1.5\sigma^2$$

Fund	$E(r)$	σ^2	U
Northgate Growth	0.11	0.0400	$0.11 - 0.0600 = 0.0500$
Westbury Balanced	0.08	0.0144	$0.08 - 0.0216 = 0.0584$
Eastholm Income	0.06	0.0025	$0.06 - 0.00375 = 0.05625$

Westbury Balanced has the highest utility (5.84%).

Why the other options are wrong

- **A.** Northgate Growth has the highest expected return, but its large variance creates the biggest utility penalty ($1.5 \times 0.04 = 0.06$), leaving a utility of only 5.00%.
- **C.** Eastholm Income has the lowest risk and is the choice that results if the standard deviation is used instead of the variance in the penalty term. With the variance, its utility (5.625%) is slightly below that of the balanced fund.

Key takeaway Use σ^2 (in decimals) and the $\frac{1}{2}$ in the utility function; neither the highest return nor the lowest risk is automatically the best choice.

MODULE 8.6

Capital Allocation Line and Capital Market Line

LOS 8.c — Capital allocation line, optimal portfolio and the capital market line

The capital allocation line (CAL)

Combining a risky portfolio with the risk-free asset (assuming zero risk for the risk-free asset, a risky portfolio whose return and risk exceed the risk-free rate, only these two assets available, and unlimited borrowing and lending at the risk-free rate) gives a straight line in risk–return space:

KEY CONCEPT

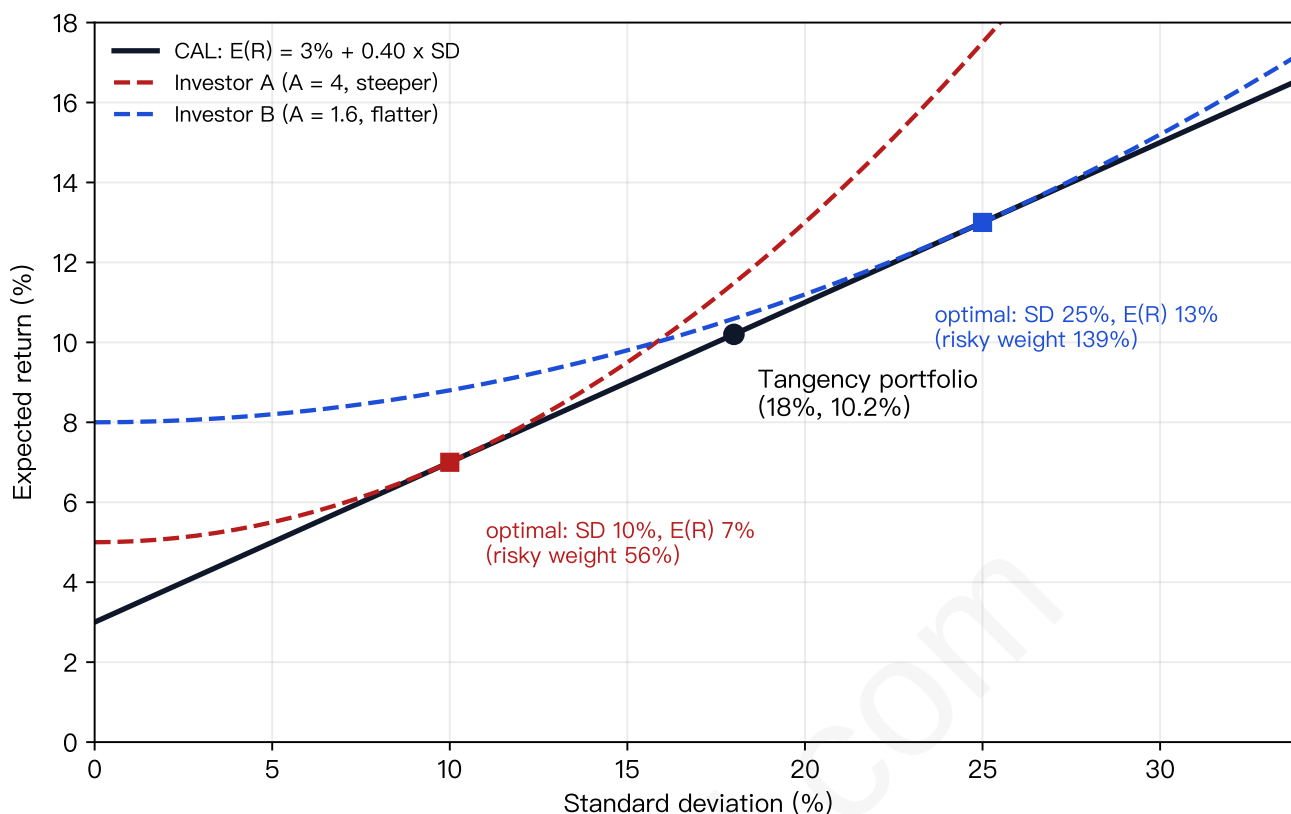
$$E(R_C) = R_f + \left[\frac{E(R_P) - R_f}{\sigma_P} \right] \sigma_C$$

This is the risky portfolio's **capital allocation line (CAL)**. It starts at the risk-free rate on the vertical axis, and its slope is the risky portfolio's Sharpe ratio. Points between the intercept and the risky portfolio involve lending (a positive weight in the risk-free asset); points beyond it involve borrowing.

Example. Let $R_f = 3\%$; the tangency portfolio has an expected return of 10.2% and a standard deviation of 18% (Sharpe ratio = $7.2/18 = 0.40$). A client with 40% in the risky portfolio has $E(R) = 0.40(10.2\%) + 0.60(3\%) = 5.88\%$ and $\sigma = 0.40(18\%) = 7.2\%$; check with the CAL: $3\% + 0.40(7.2\%) = 5.88\%$. A client who borrows 30% of wealth at 3% and puts 130% in the risky portfolio has $E(R) = 1.30(10.2\%) - 0.30(3\%) = 12.36\%$ and $\sigma = 1.30(18\%) = 23.4\%$; check: $3\% + 0.40(23.4\%) = 12.36\%$. Both portfolios lie on the same line.

Choosing the optimal portfolio

The investor's **optimal portfolio** is the point where the CAL touches the highest attainable indifference curve. Higher curves would be better but cannot be reached; lower curves are dominated.



Optimal portfolios on a CAL for a more and a less risk-averse investor

In the figure, the more risk-averse Investor A (steeper indifference curves) reaches the highest attainable curve at a standard deviation of 10% (about 56% in the risky portfolio, the rest lent at the risk-free rate). The less risk-averse Investor B (flatter curves) chooses 25% (about 139% in the risky portfolio, financed partly by borrowing). Both points lie on the same CAL: risk aversion changes where on the line an investor sits, while the slope of the line stays the same.

The capital market line (CML)

The **capital market line (CML)** is the special CAL whose risky portfolio is the **market portfolio**, which holds all risky assets weighted by market value:

$$E(R_P) = R_f + \left[\frac{E(R_M) - R_f}{\sigma_M} \right] \sigma_P$$

The market portfolio sits on the efficient frontier and is **mean-variance efficient** (the highest Sharpe ratio possible). The **mutual fund theorem** states that every investor can obtain an efficient portfolio by combining the **risk-free asset** with the **market portfolio**; investors differ only in how much of each they hold.

KEY CONCEPT

	CAL	CML
Risky portfolio	Any risky portfolio (e.g., the investor's own tangency portfolio)	The market portfolio

	CAL	CML
Expectations	Heterogeneous — each investor can have a different CAL	Homogeneous — one CML for all investors
Slope	Sharpe ratio of that risky portfolio	Sharpe ratio of the market portfolio

In practice a portfolio of "all risky assets" cannot be held, so a broad market index is used as the proxy.

COMMON EXAM TRAPS

- Every portfolio on a CAL has the same Sharpe ratio as that CAL's risky portfolio, however much is lent or borrowed.
- The mutual fund theorem combines the risk-free asset with the market portfolio specifically; another efficient portfolio or the minimum-variance portfolio does not qualify.
- Owning "about 30 stocks" diversifies away most unsystematic risk; that rule of thumb is a different idea from the mutual fund theorem.

EXAM SHORTCUTS

- Every portfolio on a CAL with a positive allocation to the risky portfolio has the Sharpe ratio of that risky portfolio, however much is lent or borrowed, so its Sharpe ratio can be read from the risky portfolio directly.

BOTTOM LINE

- Under its assumptions (a riskless risk-free asset, a risky portfolio with return and risk above the risk-free rate, only these two assets, and unlimited borrowing and lending at the risk-free rate), the CAL is $E(R_C) = R_f + [(E(R_P) - R_f)/\sigma_P]\sigma_C$, with intercept R_f and slope equal to the risky portfolio's Sharpe ratio.
- Points on the CAL between the intercept and the risky portfolio involve lending at the risk-free rate, and points beyond the risky portfolio involve borrowing.
- An investor's optimal portfolio is where the CAL touches the highest attainable indifference curve; a more risk-averse investor chooses a lower-risk point on the same line.
- The CML is the CAL whose risky portfolio is the market portfolio, which holds all risky assets weighted by market value and is mean-variance efficient.
- Under the mutual fund theorem, every investor can obtain an efficient portfolio by combining the risk-free asset with the market portfolio.
- A CAL allows heterogeneous expectations, so each investor can have a different CAL, while the CML assumes homogeneous expectations and is the same for all investors.

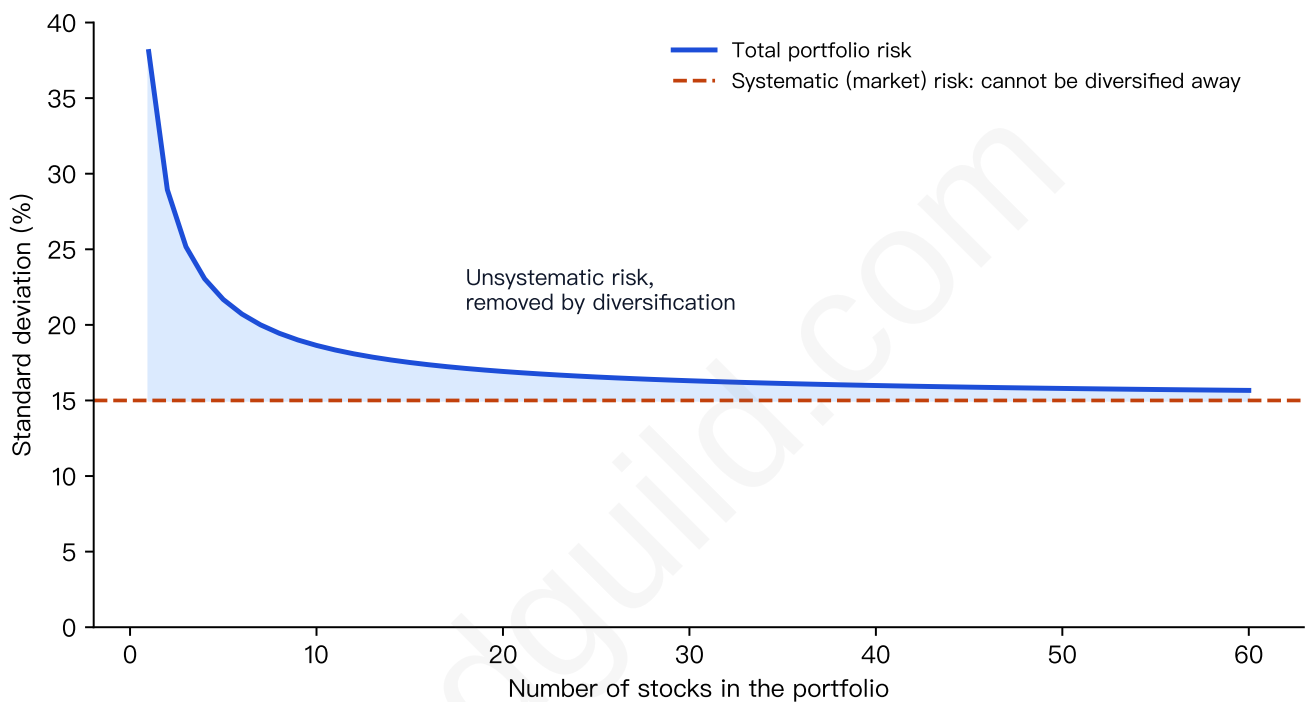
MODULE 8.7

The Capital Asset Pricing Model

LOS 8.c — Systematic risk, beta and the capital asset pricing model (CAPM)

Only systematic risk is priced

The market portfolio holds every risky asset, so all unsystematic risk has been diversified away within it and only systematic risk remains (Module 8.3). Since unsystematic risk can be removed at no cost, the market rewards investors only for bearing systematic risk, and the CAPM prices assets on that basis.



Diversification: portfolio risk falls toward systematic risk as stocks are added (illustrative figures)

Return-generating models and the CAPM

Return-generating models estimate expected returns from one or more factors (macroeconomic factors such as GDP growth or inflation, or fundamental factors such as earnings growth or firm size). The **capital asset pricing model (CAPM)** is a single-factor model: expected return depends only on exposure to systematic risk.

The CAPM assumes **homogeneous expectations**; investors hold efficient portfolios that maximize their expected utility; investors are **price takers**; and markets are in **equilibrium** (no mispriced assets).

KEY CONCEPT

$$E(R_i) = R_f + \beta_i [E(R_M) - R_f]$$

$E(R_M) - R_f$ is the **market risk premium**, the extra return expected for holding the market instead of the risk-free asset.

Beta

Beta measures an asset's sensitivity to market (systematic) risk:

KEY CONCEPT

$$\beta_i = \frac{\text{Cov}(R_i, R_M)}{\sigma_M^2} = \frac{\rho_{i,M} \sigma_i \sigma_M}{\sigma_M^2} = \frac{\rho_{i,M} \sigma_i}{\sigma_M}$$

The market portfolio has a beta of 1.

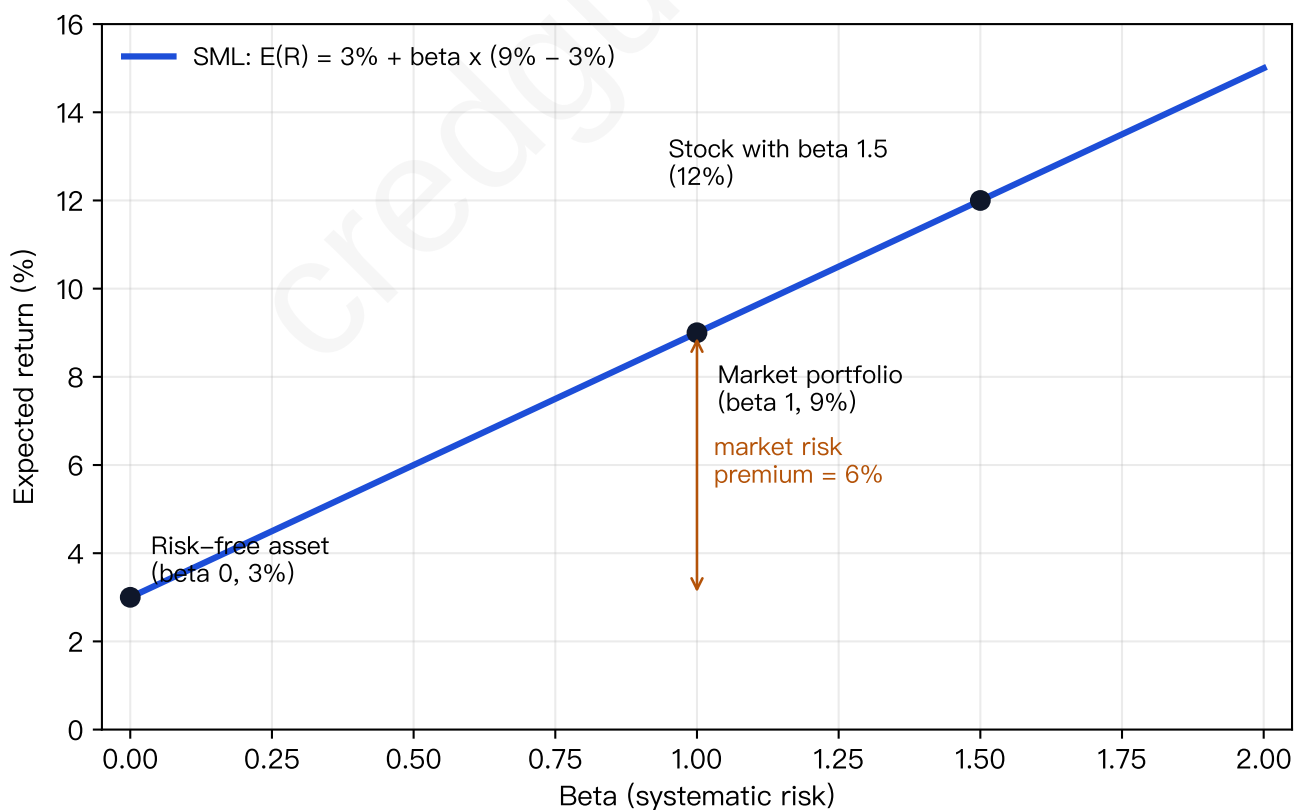
- Exam convention: a beta above 1 indicates an expected return above the market's, and a beta below 1 an expected return below it.
- Current practice: this holds only when the market risk premium is positive; with a negative premium the ordering reverses.

Beta can be estimated by regression (see the reading on simple linear regression).

Example. With $\rho_{i,M} = 0.5$, $\sigma_i = 36\%$ and $\sigma_M = 20\%$, $\beta = 0.5 \times 36/20 = 0.90$. A stock can be almost twice as volatile as the market, as here, and still have a beta below 1 if its correlation with the market is low.

The security market line (SML)

The **security market line (SML)** is the graph of the CAPM: expected return on the vertical axis, **beta** on the horizontal axis, intercept R_f , slope equal to the market risk premium.



The security market line

Example (from the figure). With $R_f = 3\%$, $E(R_M) = 9\%$ and $\beta = 1.5$, $E(R) = 3\% + 1.5(6\%) = 12\%$.

KEY CONCEPT

	CML	SML
Horizontal axis	Total risk (standard deviation)	Systematic risk (beta)
Applies to	Efficient portfolios only	Any asset or portfolio
Slope	Market Sharpe ratio	Market risk premium

Limitations of portfolio theory

Asset returns are not normally distributed (skewness, kurtosis); correlations are unstable, especially in market stress; markets are less than perfectly efficient; models are very sensitive to estimation error; the future may not look like the past. Practical constraints ignored by the theory include transaction costs from rebalancing and taxes. Taxes on distributions and on capital appreciation differ across portfolios and can change the optimal portfolio, and realizing capital losses to offset capital gains changes after-tax, risk-adjusted realized returns.

COMMON EXAM TRAPS

- If the question gives the market risk premium, multiply it by beta directly; if it gives the market return, subtract R_f first. In the SML example, using the 9% market return as the premium gives $3\% + 1.5(9\%) = 16.5\%$ instead of 12%.
- Do not forget to add the risk-free rate back. In the SML example, $1.5 \times 6\% = 9\%$ instead of 12%.
- Beta uses the correlation and both standard deviations: $\beta = \rho\sigma_i/\sigma_M$ rather than σ_i/σ_M . In the example, $36/20 = 1.80$ instead of 0.90.

BOTTOM LINE

- Because unsystematic risk can be diversified away at no cost, the market rewards investors only for bearing systematic risk.
- The CAPM is a single-factor model, $E(R_i) = R_f + \beta_i[E(R_M) - R_f]$, that assumes homogeneous expectations, utility-maximizing investors holding efficient portfolios, price takers and markets in equilibrium.
- Beta is $\text{Cov}(R_i, R_M)/\sigma_M^2 = \rho_{i,M}\sigma_i/\sigma_M$, and the beta of the market portfolio is 1.
- Exam convention: a beta above 1 indicates an expected return above the market's and a beta below 1 one below it; current practice: this ordering holds only when the market risk premium is positive.
- The SML plots expected return against beta with intercept R_f and slope equal to the market risk premium and applies to any asset, while the CML plots against total risk and applies only to efficient portfolios.
- Portfolio theory is limited by non-normal returns, unstable correlations, imperfectly efficient markets, sensitivity to estimation error and a future that may differ from the past, and it ignores transaction costs and taxes.

QUICK CHECK

Question 4

Using the capital asset pricing model (CAPM) and the data in the table below, the expected return on Gorham Textiles shares is *closest* to:

Inputs for Gorham Textiles

Input	Value
Beta of Gorham Textiles	0.8
Market risk premium	6%
Risk-free rate	3%

- A 7.8%.
- B 5.4%.
- C 4.8%.

Answer: A

Under the CAPM, an asset's expected return equals the risk-free rate plus beta times the market risk premium. The market risk premium is given directly, so it is multiplied by beta without subtracting the risk-free rate again.

$$E(R_i) = R_f + \beta_i [E(R_M) - R_f] = 3\% + 0.8(6\%) = 3\% + 4.8\% = 7.8\%$$

Why the other options are wrong

- **B.** 5.4% treats 6% as the market *return* and subtracts the risk-free rate again: $3\% + 0.8(6\% - 3\%)$. The 6% is already the market risk premium.
- **C.** 4.8% is only the asset's risk premium, $0.8 \times 6\%$; the risk-free rate must be added.

Key takeaway Check whether the question gives the market return or the market risk premium before plugging into the CAPM.

Why each answer is right and why each other option is wrong: credguild.com/cfa/level-1/notes/quantitative-methods/portfolio-standard-deviation/. The full bank has 23 questions on this reading.

QUANTITATIVE METHODS · READING 9

Simulation of Financial Asset Prices and Returns

What you'll learn

LOS 9.a Describe historical simulation, its assumptions, strengths and weaknesses, and how it is used in investment applications such as VaR and bond portfolios.

LOS 9.b Describe bootstrap resampling (sampling with replacement), its strengths and weaknesses, and its investment applications.

LOS 9.c Describe Monte Carlo simulation, including multivariate simulation with the Cholesky decomposition, its strengths and weaknesses, and its investment applications.

MODULE 9.1

Historical Simulation

This reading describes three ways to simulate asset prices and returns: historical simulation, which applies actual past changes to current positions; bootstrap resampling, which draws from the observed sample with replacement; and Monte Carlo simulation, which draws from distributions the analyst specifies. For each method it covers the assumptions, investment applications such as VaR and option valuation, and the main strengths and weaknesses.

LOS 9.a — Historical simulation

What a simulation does

KEY CONCEPT

A **simulation** is a predictive modeling technique that estimates the range of possible outcomes of a complex system (for example, a portfolio's return distribution, its risk, or future cash flows and liquidity) by running it through many scenarios. Every simulation follows the same four steps:

1. **Set parameters:** state the objective, choose the output variables, and define the calculations linking inputs to outputs.
2. **Generate scenarios:** create the input values for each trial (simple or drawn from complex distributions).
3. **Evaluate scenarios:** run each scenario through the calculation engine to produce outcome paths.
4. **Compile and interpret results:** aggregate the outcomes (e.g., into a distribution) and draw conclusions.

Historical simulation (this module), bootstrap resampling (Module 9.2) and Monte Carlo simulation (Module 9.3) differ mainly in step 2, where the scenarios come from.

Historical simulation

Historical simulation applies actual past changes in prices, returns or risk factors to the current portfolio positions. It assumes that **past data represent the future** (past performance is a reliable guide to future outcomes and risks, and the historical record stands in for the population of outcomes). When the parameters are set, the chosen dates and risk factors may need simplifying assumptions so that all returns cover the same time frame. Equity inputs are usually logarithmic (continuously compounded) price returns, and percentage returns work as well; for bonds, the simulated yields serve as discount rates for the future cash flows.

KEY CONCEPT**Strengths**

Easy to explain — narratives are built around real past events

Nonparametric — no need to force the data into a distribution such as the normal; handles **excess kurtosis**, **skewness** and **fat tails**, and captures extreme events and drift better than a parametric model

Captures **volatility clustering** (calm and turbulent periods tend to persist)

Weaknesses

Assumes the past repeats; weak when there are **regime changes** or unprecedented events

Can only produce outcomes that actually occurred in the lookback period

Results depend on the quality and completeness of the data

Missing data and proxies

Gaps in the historical record arise from thinly traded instruments, asynchronous global markets (holidays, outages, trading halts), and securities that did not exist during part of the period (e.g., a recent IPO). The fix is a **proxy**: a substitute instrument, or a weighted combination of instruments with data available, that stands in for the missing observations.

Application: value at risk (VaR)**KEY CONCEPT**

Value at risk (VaR) is the minimum loss expected over a specified period in the worst 5% (or 1%) of outcomes. Equivalently, it is the maximum potential loss at the matching level of confidence (95% or 99%); larger losses still occur in the tail. Using historical simulation for a long-only portfolio:

1. *Set parameters*: compute daily (usually logarithmic) returns of each asset over the lookback period.
2. *Generate scenarios*: apply each day's returns to today's positions to get one profit-or-loss (P&L) figure per historical day.
3. *Evaluate scenarios*: build the P&L distribution (e.g., a histogram).
4. *Compile and interpret*: rank the P&L from worst to best; the 95% VaR is the fifth percentile of the distribution. If the percentile falls between two observations, use either neighbor or a weighted average.

Example (historical vs. parametric VaR). A \$4 million equity portfolio is revalued under each of the last 500 daily return scenarios. Five percent of 500 is 25 outcomes, so the fifth percentile lies between the 25th-worst P&L (–\$118,000) and the 26th-worst (–\$114,000). The 1-day 95% historical VaR is therefore about \$116,000 (the average of the two neighbors; either neighbor is also acceptable). A parametric VaR that assumes normal returns, a zero daily mean and a daily standard deviation of 1.5% gives $1.645 \times 0.015 \times \$4,000,000 = \$98,700$. The historical figure is larger because the actual return distribution has a fatter left tail than the normal distribution allows.

Because VaR focuses on the loss tail, a few large gains barely affect it. For daily VaR, analysts usually assume a mean return of zero, because a daily drift is hard to distinguish statistically from zero (the null hypothesis of no directional move) and daily volatility dwarfs any trend.

Application: bonds

A bond's historical prices are misleading because its interest-rate sensitivity (duration) shrinks as it approaches maturity: a 10-year zero-coupon bond bought five years ago now behaves like a 5-year bond. Historical simulation for bonds therefore uses historical changes in yields (for example, benchmark yields for the same remaining maturity) and applies them to today's cash flows. To simulate a bond with three years to maturity, use yield changes that were observed on bonds that had three years left to maturity at the time.

COMMON EXAM TRAPS

- Historical simulation assumes past data represent the future. Fitting a statistical model to the population describes Monte Carlo simulation, and resampling with replacement describes bootstrapping.
- Historical 95% VaR is the fifth percentile of the simulated P&L distribution (about the 25th-worst of 500 outcomes) rather than the single worst outcome.
- A proxy fills missing data; it is not a risk factor or a hypothetical scenario.

EXAM SHORTCUTS

- With N historical scenarios, the 95% VaR lies between the $(0.05N)$ th-worst P&L and the next-worst one, so only those two outcomes need to be located (either one or a weighted average of the two is acceptable).

BOTTOM LINE

- Every simulation sets parameters, generates scenarios, evaluates them and compiles the results; historical simulation, bootstrapping and Monte Carlo simulation differ mainly in where the scenarios come from.
- Historical simulation applies actual past changes in prices, returns or risk factors to the current portfolio positions and assumes that past data represent the future.
- Historical simulation is nonparametric and captures skewness, excess kurtosis, fat tails and volatility clustering, but it can only produce outcomes from the lookback period and is weak when there are regime changes or unprecedented events.
- Missing historical data are filled with a proxy, a substitute instrument or a weighted combination of instruments with data available.
- VaR is the minimum loss over a specified period in the worst 5% (or 1%) of outcomes; the 95% historical VaR is the fifth percentile of the simulated P&L distribution, and daily VaR usually assumes a mean return of zero.
- Because a bond's duration shrinks as it approaches maturity, historical simulation for bonds applies past yield changes for the same remaining maturity to today's cash flows instead of using the bond's own past prices.

QUICK CHECK

Question 1

Which of the following is *most likely* an advantage of historical simulation compared with a parametric approach that assumes normally distributed returns?

- A It remains reliable after a regime change that makes the future unlike the past.
- B It captures fat tails and volatility clustering without imposing a distributional assumption.
- C It can generate outcomes more extreme than any observed during the lookback period.

Answer: B

Historical simulation is nonparametric: it uses actual past changes, so the scenarios automatically carry the fat tails, skewness and volatility clustering present in the data. A normal-distribution model tends to understate extreme outcomes.

Why the other options are wrong

- **A.** Reliance on the past is the main weakness of historical simulation; after a regime change, past data may no longer represent the future.
- **C.** Historical simulation can only reproduce outcomes that actually occurred in the lookback period; simulating conditions never seen before is an advantage of Monte Carlo simulation.

Key takeaway Historical simulation is nonparametric, captures fat tails and volatility clustering and is easy to explain; its limit is that it can only replay what actually happened.

MODULE 9.2

Bootstrap Resampling

LOS 9.b — Bootstrap resampling

How bootstrapping works

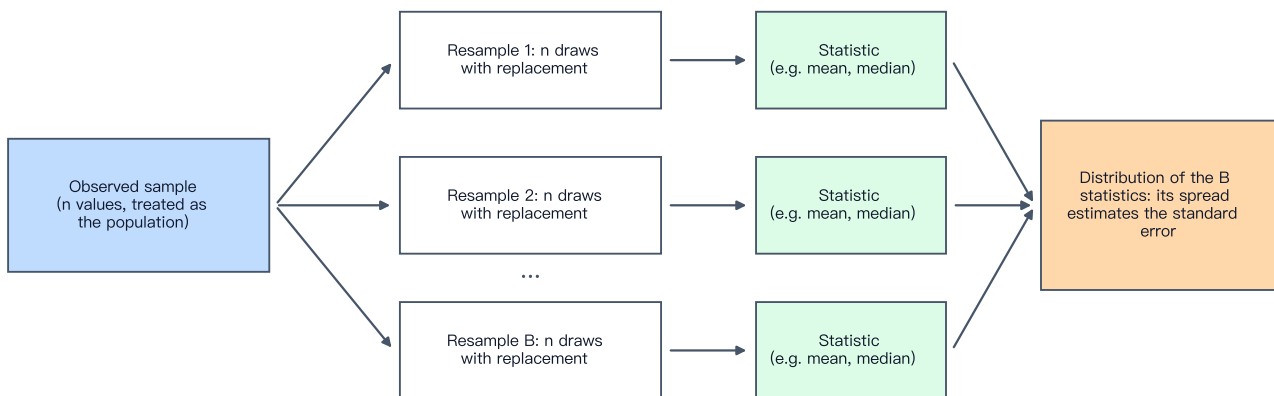
KEY CONCEPT

Bootstrap resampling (bootstrapping) is a **nonparametric** simulation method. It treats the observed historical **sample as a proxy for the population** and creates many new datasets by drawing observations from it at random **with replacement**: each observation is drawn, used, and put back so that it can be drawn again. Drawing a random value in this way is called **pulling**.

- With a dataset of n observations, each observation has a probability of $1/n$ of being selected on every draw.
- Within one scenario, the same observation can be drawn as many times as there are draws in that scenario.
- Across scenarios, the same observation can appear in many samples (or in none).

Example. To simulate a 6-month path from 60 historical monthly returns, draw six random integers from 1 to 60 and look up the corresponding returns. Any single month has a $1/60$ chance on each draw and could, in principle, appear up to six times in the same path (though that is very unlikely). Repeat thousands of times.

Each draw: every observation has probability $1/n$, and it goes back before the next draw



How bootstrap resampling builds a sampling distribution

Bootstrapping vs. historical simulation

The two methods treat the data differently. Historical simulation replays the past dataset directly, as if it were the population, so each scenario is a sequence of events that actually happened. Bootstrapping treats the observed sample as a proxy for the population and builds new combinations of past observations by resampling with replacement. Module 9.3 compares all three simulation methods side by side.

Bootstrapping needs no theoretical model. Each resample gives one value of the statistic of interest (a mean, say), so thousands of resamples trace out an approximate sampling distribution of that statistic even when the original dataset is small. The spread of that distribution makes the stability of an estimate easy to assess. The repeated resampling takes somewhat more computation than historical simulation. Because it preserves the empirical distribution of the data while reshuffling it, it reduces the risk of overfitting to one particular historical sequence. It suits finance problems in which the population distribution is unknown but can be inferred from the sample.

Applications

- **European-style call option values at maturity.** A European call can be exercised only at maturity and is valuable only if the underlying price ends above the exercise price, so the objective is the option's value at maturity. Resample historical returns to build many price paths over the investment period, compute the option payoff at the end of each path, and study the resulting distribution and its mean. This avoids the constant-volatility and normality assumptions of theoretical option-pricing models. Because of the option's leverage and limited downside, the distribution of option values can be more skewed than the distribution of share prices. The result is the distribution of terminal payoffs under historical (real-world) returns. It is not by itself the option's current no-arbitrage price: discounting a real-world average payoff at the risk-free rate would not give that price (Module 9.3 sets out the conditions for Monte Carlo pricing).
- **Estimating the volatility of target date funds** (which shift toward less risky assets as the retirement date approaches): bootstrap monthly returns, compute the standard deviation of each simulated series, and compare ranges across funds to judge which is more aggressive.

Strengths and weaknesses

KEY CONCEPT

Strengths

Minimal assumptions — no need to specify a distribution such as the normal

Simple to implement

Statistically robust — reliable estimates of the variability of a statistic, even from small samples

Reliable standard errors and confidence intervals, less influenced by outliers

Adaptable to many statistics (mean, median, correlation)

Works with the sample at hand; no large external datasets needed

Weaknesses

Sensitive to sample quality — any bias in the original sample is carried into every resample

Computationally intensive for thousands of resamples on large datasets

No analytical insight — estimates, but no theoretical explanation

Limited extrapolation — cannot generate outcomes outside the range of observed data (weak for extreme tail events)

Not suited to autocorrelated time series — random resampling breaks the time-sequence dependence

Bootstrapping is useful when the data show time-varying volatility, skewness or fat tails that are hard to model parametrically, and when historical patterns are expected to persist. It complements analytic methods, which explain the underlying relationships in more detail.

COMMON EXAM TRAPS

- Bootstrapping makes no distributional assumption; Monte Carlo is the method that requires one.

BOTTOM LINE

- Bootstrapping is a nonparametric method that uses the sample at hand as a proxy for the population and builds new datasets by drawing from it at random with replacement, so each of n observations has a $1/n$ chance on every draw.
- In bootstrapping the same observation can be drawn several times within one scenario and can appear in many samples or in none.
- Historical simulation replays the past dataset as if it were the population, while bootstrapping builds new combinations of past observations, which reduces the risk of overfitting to one historical sequence.
- Bootstrapping can give the distribution of a European call's payoff at maturity under historical returns, which is not by itself the option's current no-arbitrage price, and can estimate the volatility of target date funds.
- Bootstrapping needs minimal assumptions and works with small samples, but it carries any bias in the original sample into every resample, cannot generate outcomes outside the observed range and is not suited to autocorrelated time series.

QUICK CHECK**Question 2**

An analyst builds each bootstrap scenario for an eight-quarter horizon by drawing from a dataset of 40 historical quarterly returns. Within a single scenario, any one historical observation can be drawn:

- A** up to eight times.
- B** only once.
- C** up to 40 times.

Answer: A

Bootstrapping samples with replacement: after each draw the observation is returned to the dataset. An eight-quarter scenario needs eight draws, so the same observation could in principle be pulled on every one of them, although with a $1/40$ chance per draw this is extremely unlikely.

Each scenario consists of 8 draws of a random integer from 1 to 40, each mapped to one historical quarterly return. The probability that a given observation is chosen on any single draw is $1/40$, and the maximum number of appearances in one scenario equals the number of draws, 8.

Why the other options are wrong

- **B.** "Only once" describes sampling *without* replacement. Bootstrap resampling replaces each observation, so repeats are possible.
- **C.** Forty is the size of the dataset; a scenario of eight quarters contains only eight draws.

Key takeaway Sampling with replacement: repeats are possible, capped by the number of draws in the scenario.

MODULE 9.3

Monte Carlo Simulation**LOS 9.c — Monte Carlo simulation****The method**

Monte Carlo simulation generates scenarios by random sampling from probability distributions specified by the analyst (normal, lognormal, or any other shape with any parameters) rather than from raw historical observations. Running thousands of trials produces a **simulated frequency distribution** of prices or returns, from which risk and performance measures are estimated. The analyst should run as many trials as practical: more trials give more precise estimates, and the number of trials is limited essentially by computational power and time. Because the analyst controls the input assumptions, the sensitivity of the output to each assumption can be tested ("what if" analysis).

Applications

- **Portfolio performance** over a future horizon: expected return and standard deviation to a chosen level of confidence.

- **Valuing securities without an analytic (closed-form) pricing formula**, such as path-dependent options (e.g., **Asian-style options**, whose payoff depends on an average price), mortgage-backed securities (the borrowers' prepayment option acts like an embedded call) and convertible bonds.
- **European call options.**
 - Exam convention: a Monte Carlo value for a European call should match the Black-Scholes-Merton value, and it gets closer as more trials are run.
 - Current practice: this holds only when the simulated price dynamics match the Black-Scholes-Merton assumptions: a lognormal price with constant volatility and, for pricing, an expected return equal to the risk-free rate less any dividend yield (risk-neutral valuation). The average payoff discounted at the risk-free rate then converges to the Black-Scholes-Merton value as the number of trials grows. A lognormal assumption alone is not enough: simulating with a real-world expected return and then discounting at the risk-free rate would not reproduce that value.

Steps for an option: (1) set the horizon, risk-free rate, the price distribution (e.g., lognormal), volatility and expected return; (2) draw random numbers to create price paths; (3) compute the payoff on each path over thousands of trials; (4) average the payoffs and discount at the risk-free rate.

Multivariate simulation and the Cholesky decomposition

When several variables move together (a portfolio, or an option on several assets), the simulated inputs must have the right correlation. Under the usual lognormal price model, the assets' log returns are assumed to be jointly normal, and the inputs are each asset's volatility and the correlations between returns. The **Cholesky decomposition** factorizes the covariance (or correlation) matrix and maps independent standard normal draws into correlated normal variables, which are then used to generate the simulated paths. For two variables with independent draws Z_1 and Z_2 (mean 0, standard deviation 1):

KEY CONCEPT

$$X_1 = \sigma_1 Z_1, \quad X_2 = \sigma_2 (\rho Z_1 + \sqrt{1 - \rho^2} Z_2)$$

Example. With $\sigma_1 = 10\%$, $\sigma_2 = 15\%$, $\rho = 0.8$ and draws $Z_1 = 0.5$ and $Z_2 = 1.0$, $X_1 = 5\%$; $X_2 = 15\%(0.8 \times 0.5 + 0.6 \times 1.0) = 15\%$.

Strengths and weaknesses

KEY CONCEPT

Strengths

Wide applicability — assets, risk factors, portfolios; nonlinear relationships, time-varying volatility, fat tails

Flexibility for nonnormal distributions (skewness, excess kurtosis)

Weaknesses

Complexity — needs modeling expertise and a time-consuming set-up

Model dependency — results are only as good as the assumed distributions and parameters

Strengths

Precision given enough trials

Correlation handling among many assets

Forward-looking — can simulate hypothetical or extreme conditions never seen in history

Weaknesses

Computational intensity — slow and costly for real-time use

Data requirements — extensive data to estimate parameters such as volatilities and correlations

Gives a **statistical estimate**, not an exact answer, and cannot provide the cause-and-effect **insight of analytic methods**

Monte Carlo is usually run alongside analytic methods. Where no analytic method can be applied, it can serve as the valuation tool itself, and its robustness is checked by changing the assumptions and rerunning.

Comparing the three methods**KEY CONCEPT**

Feature	Historical simulation	Bootstrapping	Monte Carlo simulation
Data source	Historical dataset	Observed sample	Assumed (modeled) distribution
Sampling	Directly from history	Resampling with replacement	Random draws from a specified distribution
Assumption	Past data represent the future	Sample represents the population	Process generating outcomes can be modeled statistically
Typical uses	Risk management, stress tests	Confidence intervals, standard errors	Complex systems, long-term outcomes
Flexibility	Limited to historical patterns	Limited by sample size	Highly flexible, hypothetical scenarios
Uncertainty captured	Only what was observed in history	Variability from resampling	Hypothetical uncertainty from the modeled inputs
Computational intensity	Low to moderate	Moderate	High

COMMON EXAM TRAPS

- The Cholesky decomposition generates correlated paths from an assumed correlation; it does not estimate correlations or covariances.

EXAM SHORTCUTS

- Identify the method from the source of the scenarios: actual past changes mean historical simulation, resampling with replacement from the observed sample means bootstrapping, and random draws from

an assumed distribution mean Monte Carlo simulation.

BOTTOM LINE

- Monte Carlo simulation draws scenarios from probability distributions specified by the analyst and runs thousands of trials to build a simulated frequency distribution, with more trials giving more precise estimates.
- Monte Carlo simulation is used to estimate portfolio performance and to value securities without a closed-form pricing formula, such as Asian-style options, mortgage-backed securities and convertible bonds.
- Exam convention: a Monte Carlo value for a European call should match the Black-Scholes-Merton value, getting closer as more trials are run; current practice: this holds only when the simulated prices are lognormal with constant volatility and a risk-neutral expected return, with the average payoff discounted at the risk-free rate.
- The Cholesky decomposition maps independent standard normal draws into correlated normal variables using an assumed correlation; it does not estimate correlations or covariances.
- Monte Carlo simulation is flexible, forward-looking and handles nonnormal distributions and correlations, but it is complex, model-dependent, computationally intensive and data-hungry, and it gives a statistical estimate without the cause-and-effect insight of analytic methods.
- Historical simulation assumes past data represent the future, bootstrapping assumes the sample represents the population, and Monte Carlo simulation assumes the process generating outcomes can be modeled statistically.

QUICK CHECK

Question 3

Compared with historical simulation and bootstrap resampling, an advantage of Monte Carlo simulation is that it:

- A** has lower data requirements.
- B** produces results that are easier to explain.
- C** can model hypothetical scenarios.

Answer: C

Monte Carlo scenarios are generated from distributions chosen by the analyst, so the method can simulate hypothetical or extreme conditions that never appeared in the historical record. Historical simulation and bootstrapping can only reuse what actually happened.

Why the other options are wrong

- **A.** Monte Carlo usually has heavy data requirements, because volatilities, correlations and other parameters must be estimated.
- **B.** Monte Carlo models are more complex and harder to communicate than methods that replay or resample actual history.

Key takeaway Monte Carlo is forward-looking: it can model scenarios with no historical precedent.

Why each answer is right and why each other option is wrong: credguild.com/cfa/level-1/notes/quantitative-methods/simulation-of-financial-asset-prices-and-returns/. The full bank has 16 questions on this reading.

credguild.com

QUANTITATIVE METHODS · READING 10

Least Squares Regression

What you'll learn

LOS 10.a Describe and interpret simple linear regression and estimate its coefficients using the least squares criterion.

LOS 10.b Describe the assumptions of simple linear regression, detect violations from residuals, and evaluate goodness of fit, coefficient tests and ANOVA results.

LOS 10.c Calculate and interpret predicted values, the standard error of estimate and prediction intervals, and describe log-lin, lin-log and log-log functional forms.

LOS 10.d Estimate and interpret CAPM beta and related statistics with a simple linear regression of excess returns on the market risk premium.

MODULE 10.1

Linear Regression Basics

This reading fits a simple linear regression by least squares, interprets its slope and intercept, and checks the model's assumptions with residual plots. It then measures fit with ANOVA, the standard error of estimate, R^2 and the F-test, tests the coefficients, builds prediction intervals, chooses among log-transformed functional forms, and estimates a stock's CAPM beta by regression.

LOS 10.a — Simple linear regression and the least squares criterion

Purpose and vocabulary

Simple linear regression explains the **variation** of one **dependent variable** with the variation of a single **independent variable**. *Variation* is the total squared deviation from the mean, $\sum_{i=1}^n (Y_i - \bar{Y})^2$; it is related to, but not the same as, the variance (the variance divides the variation by $n - 1$).

	Dependent variable (Y)	Independent variable (X)
Role	The variable being explained	The variable doing the explaining
Other names	Explained, endogenous, predicted variable	Explanatory, exogenous, predicting variable
Example: using GDP growth to forecast stock returns	Stock returns	GDP growth

A model with one dependent variable and one independent variable is a *simple* regression; a model with one dependent variable and more than one independent variable is a **multiple regression** (Level II).

The model and the regression line

KEY CONCEPT

$$Y_i = b_0 + b_1 X_i + \varepsilon_i \quad (\text{population model})$$

$$\hat{Y}_i = \hat{b}_0 + \hat{b}_1 X_i \quad (\text{estimated regression line, the line of best fit})$$

The hat marks an estimate. The **residual** for observation i , $Y_i - \hat{Y}_i$, is the gap between the actual value and the value the fitted line predicts for that X .

- Exam convention: the residual, the **error term** and the **disturbance term** are treated as the same thing, written ε_i .
- Current practice: the error term ε_i is the unobservable deviation in the population model, and the residual is its sample estimate.

Least squares criterion

Of all the lines that could be drawn through the scatter plot, **ordinary least squares (OLS)** regression picks the one that minimizes the **sum of squared errors (SSE)**, found by squaring each vertical gap between an observed value of the dependent variable and its fitted value, then adding the squares:

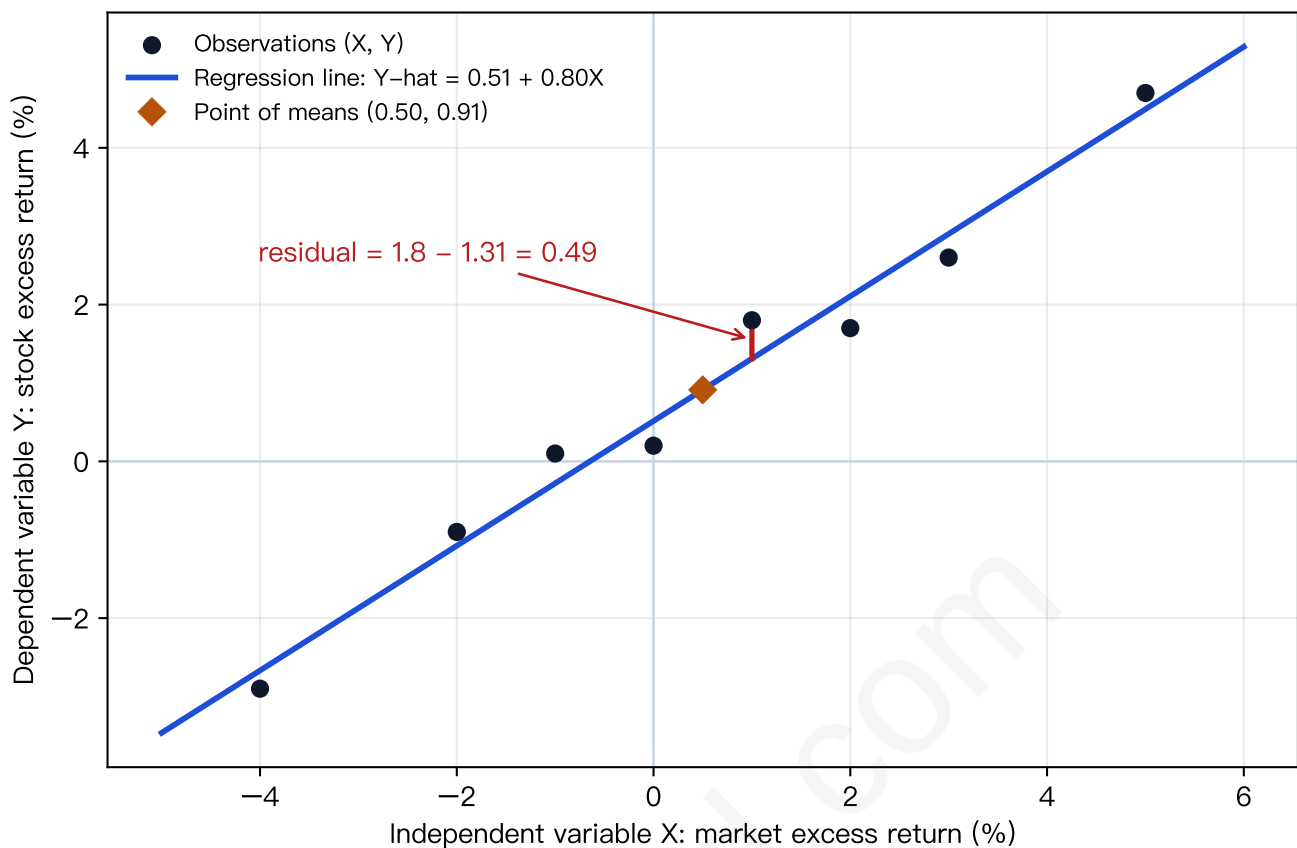
$$\text{SSE} = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

The resulting estimates are:

KEY CONCEPT

$$\hat{b}_1 = \frac{\text{Cov}(X, Y)}{\sigma_X^2}, \quad \hat{b}_0 = \bar{Y} - \hat{b}_1 \bar{X}$$

The intercept formula shows that the regression line always passes through the point of means $(\bar{X}, \bar{Y})$. The covariance and the variance share the divisor $n - 1$, which cancels, so from raw data the slope is $\hat{b}_1 = \sum(X_i - \bar{X})(Y_i - \bar{Y}) / \sum(X_i - \bar{X})^2$.



Scatter plot with the least-squares regression line

WORKED EXAMPLE (DATA IN THE FIGURE)

Eight monthly observations give $\bar{X} = 0.50\%$, $\bar{Y} = 0.9125\%$, $\text{Cov}(X, Y) = 6.593$ and $\sigma_X^2 = 8.286$ (both in $\%^2$). The underlying sums are $\sum(X_i - \bar{X})(Y_i - \bar{Y}) = 46.15$ and $\sum(X_i - \bar{X})^2 = 58.0$, and $46.15/58.0$ gives the same slope.

$$\hat{b}_1 = 6.593/8.286 = 0.7957 \approx 0.80, \quad \hat{b}_0 = 0.9125 - 0.7957(0.50) = 0.51\%$$

For the observation at $X = 1\%$ with $Y = 1.8\%$: $\hat{Y} = 0.51 + 0.80(1) = 1.31\%$, so the residual is $1.8 - 1.31 = 0.49\%$.

Interpreting the coefficients

- **Slope coefficient $\hat{b}_1$** : how much the dependent variable is expected to move when the independent variable rises by one unit. In the example, a market excess return 1 percentage point higher goes with a stock excess return 0.80 point higher. When Y is a stock's excess return and X the market's, the slope is the stock's **beta**.
- **Intercept $\hat{b}_0$** : the predicted value of Y when X equals zero (where the line crosses the vertical axis).
- A large slope does not by itself mean a strong relationship; significance must be tested (t-test or confidence interval).

Types of data

KEY CONCEPT

Type	Definition	Example
Time series	Observations of one variable at regular intervals over time	A company's quarterly EPS for five years
Cross-sectional data	Observations of many subjects at one point in time	Today's closing prices of 30 stocks in an index
Panel data	Cross-sectional observations repeated over time	Quarterly EPS of 10 companies over five years

COMMON EXAM TRAPS

- The slope divides the covariance by the variance of X, the independent variable. Dividing by the variance of Y is a common error.
- OLS minimizes the squared differences between actual and predicted values of Y. It does not compare coefficients or X values.
- The slope is the change in Y per unit change in X. Reading it as the change in X per unit of Y reverses the roles.

EXAM SHORTCUTS

- The regression line passes through $(\bar{X}, \bar{Y})$, so once the slope is known the intercept is $\bar{Y} - \hat{b}_1 \bar{X}$ with no further sums.

BOTTOM LINE

- Simple linear regression explains the variation of a dependent variable Y, $\sum(Y_i - \bar{Y})^2$, with the variation of a single independent variable X.
- Ordinary least squares chooses the line that minimizes the sum of squared errors, $SSE = \sum(Y_i - \hat{Y}_i)^2$.
- The OLS estimates are $\hat{b}_1 = \text{Cov}(X, Y) / \sigma_X^2$ and $\hat{b}_0 = \bar{Y} - \hat{b}_1 \bar{X}$, so the fitted line passes through the point of means.
- The slope is the expected change in Y for a one-unit rise in X, and the intercept is the predicted value of Y when X equals zero.
- The residual $Y_i - \hat{Y}_i$ is the gap between the actual and the fitted value; exam convention treats the residual, error term and disturbance term as the same thing, while current practice treats the residual as the sample estimate of the unobservable error term.
- Time series data follow one variable over time, cross-sectional data cover many subjects at one point in time, and panel data repeat cross-sectional observations over time.

QUICK CHECK**Question 1**

A credit analyst gathers data on 80 bond issuers at a single date to see whether an issuer's interest coverage ratio helps explain the yield spread on its bonds. In her simple linear regression, the dependent variable is:

- A** the interest coverage ratio, because it is the explanatory variable.
- B** the yield spread, because it is the variable being explained.
- C** neither, because data from a single date cannot be used in a regression.

Answer: B

The dependent variable is the one whose variation the model explains, here the yield spread. The interest coverage ratio does the explaining, so it is the independent (explanatory) variable. Observations of many issuers at one date are cross-sectional data, which a simple linear regression can use as readily as time-series data.

Why the other options are wrong

- **A.** An explanatory variable is the independent variable. Making the coverage ratio the dependent variable reverses the roles set by the analyst's question.
- **C.** Data on many subjects at one point in time are cross-sectional data, and a regression can be run on them. Time-series data are not required.

Key takeaway Dependent variable (Y): explained, endogenous, predicted. Independent variable (X): explanatory, exogenous, predicting. The research question decides which is which.

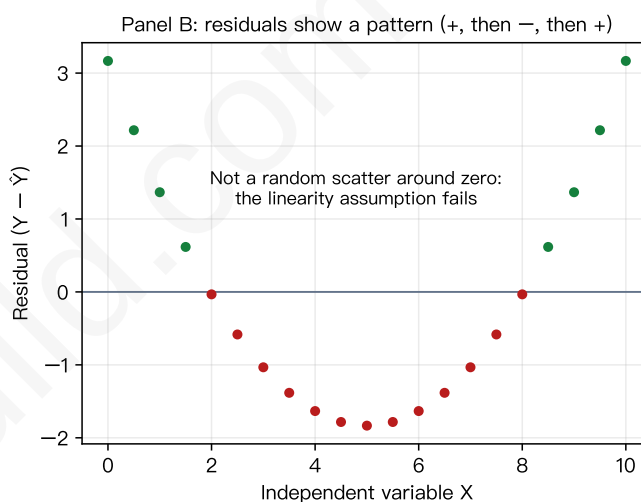
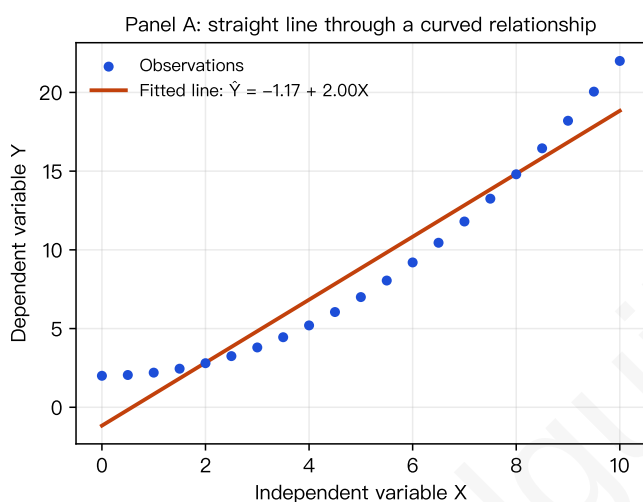
MODULE 10.2

Analysis of Variance (ANOVA) and Goodness of Fit**LOS 10.b — Assumptions, residual analysis, ANOVA and goodness of fit****Assumptions of simple linear regression****KEY CONCEPT**

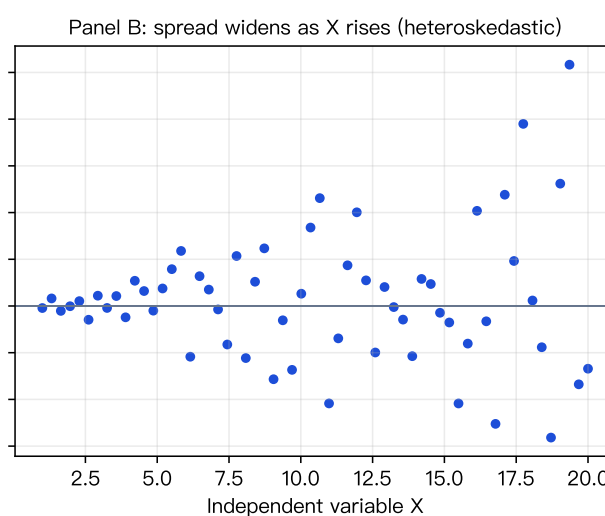
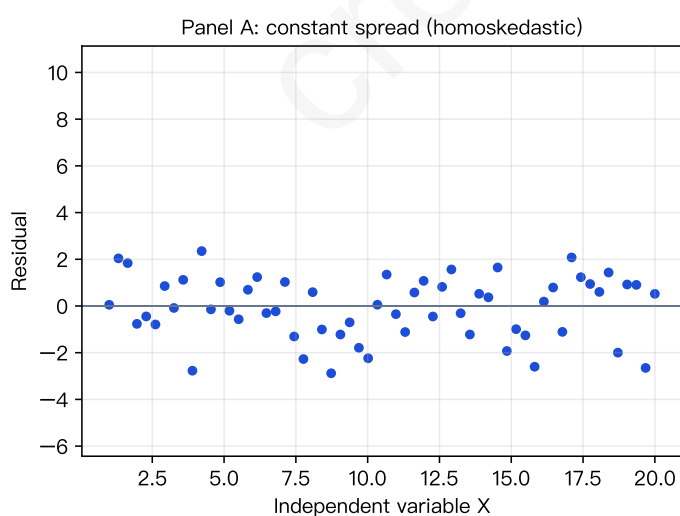
1. **Linearity:** the relationship between X and Y is linear.
2. **Homoskedasticity:** the residuals have the same variance for every observation.
3. **Independence:** residuals are uncorrelated with one another (the X–Y pairs are independent).
4. **Normality:** the residual term is normally distributed.

Violations show up in plots of the residuals against X (or against time). In a well-specified model the residuals scatter randomly around zero with no pattern.

Violation	What the residual plot shows	Name
Nonlinear relationship	A systematic pattern, e.g., residuals positive, then negative, then positive as X rises	—
Residual variance not constant	Spread of residuals widens (or narrows) with X, or changes over time	Heteroskedasticity
Residuals correlated	Recurring pattern, e.g., large errors every December (seasonality); the estimated variances of the coefficients are then wrong	Autocorrelation
Residuals not normal	Skewed or fat-tailed residuals; hypothesis tests may be unreliable in small samples (with a large sample, the central limit theorem means the estimates may still be valid)	—



A straight line fitted to a curved relationship, and the residual pattern it leaves



Residual plots: homoskedastic vs. heteroskedastic

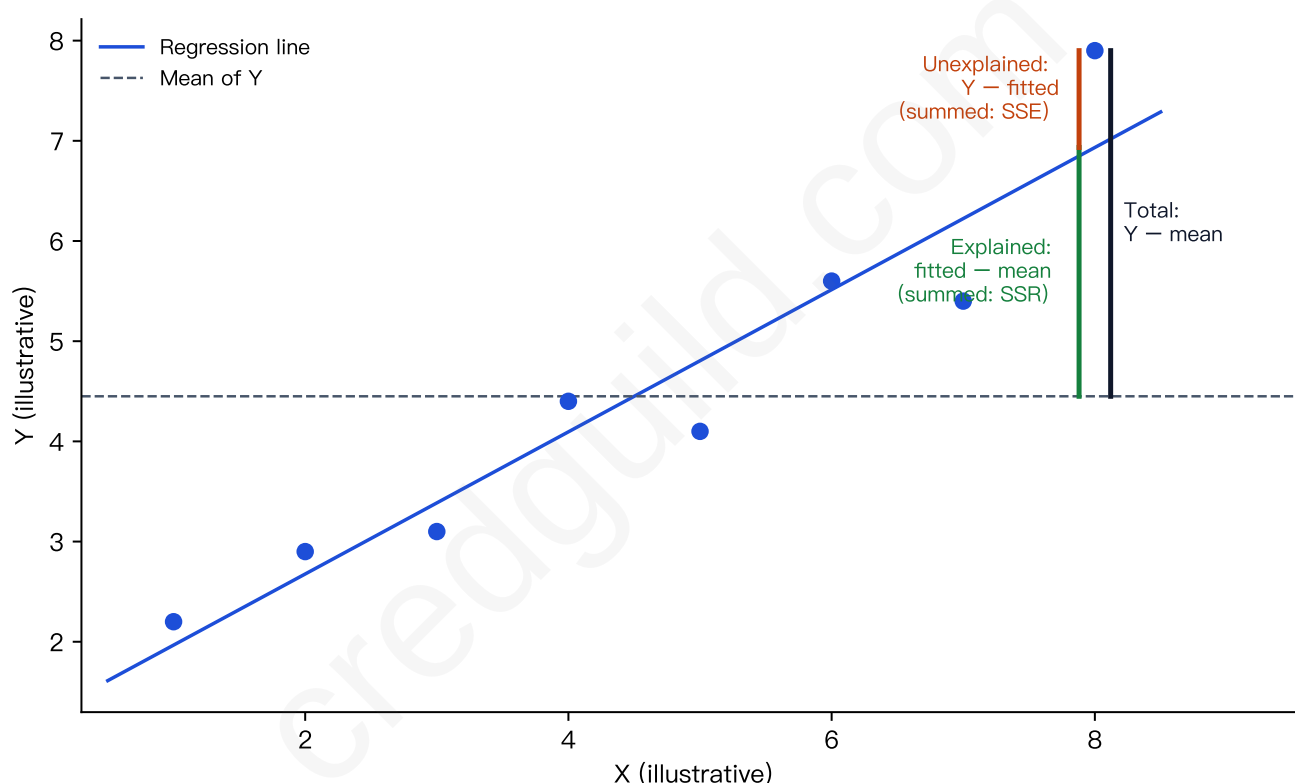
Outliers are observations far from the line. They can pull the estimated coefficients so that the line fits the other points poorly.

Analysis of variance (ANOVA)

Analysis of variance (ANOVA) splits the total variation of the dependent variable into an explained and an unexplained part:

- **Total sum of squares (SST)** = $\sum (Y_i - \bar{Y})^2$: total variation.
- **Sum of squares regression (SSR)** = $\sum (\hat{Y}_i - \bar{Y})^2$: variation explained by the regression.
- **Sum of squared errors (SSE)** = $\sum (Y_i - \hat{Y}_i)^2$: unexplained variation (also called the residual sum of squares).

$$SST = SSR + SSE$$



Total variation split into explained and unexplained parts (illustrative data)

The sample variance of Y is its total variation divided by its degrees of freedom, $SST/(n - 1)$.

- **Mean square regression (MSR)** = SSR/k ; with one independent variable ($k = 1$), $MSR = SSR$.
- **Mean squared error (MSE)** = $SSE/(n - k - 1) = SSE/(n - 2)$ in a simple regression.

KEY CONCEPT

Source of variation	Degrees of freedom	Sum of squares	Mean sum of squares
Regression (explained)	$k = 1$	SSR	$MSR = SSR/1$
Error (unexplained)	$n - 2$	SSE	$MSE = SSE/(n - 2)$

Source of variation	Degrees of freedom	Sum of squares	Mean sum of squares
Total	$n - 1$	SST	

Goodness-of-fit measures

$$SEE = \sqrt{MSE} = \sqrt{\frac{SSE}{n - 2}}$$

The **standard error of estimate (SEE)**, also called the **root mean squared error**, is the standard deviation of the residuals; the lower it is, the better the fit.

KEY CONCEPT

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

The **coefficient of determination** (R^2) measures what fraction of the dependent variable's total variation the independent variable accounts for. In a simple regression $R^2 = r^2$, so the **correlation coefficient** is $r = \pm\sqrt{R^2}$, with the sign of the slope. When only R^2 and SSR are known, $SST = SSR/R^2$ and $SSE = SST - SSR$, which then gives the MSE and the SEE.

F-test

The F-test compares the variation the regression explains, per degree of freedom, with the variation it leaves unexplained, per degree of freedom. A large ratio means the independent variable explains a lot relative to the noise.

$$F = \frac{MSR}{MSE} \quad df = 1 \text{ (numerator), } n - 2 \text{ (denominator)}$$

The **F-test** asks whether the independent variables, as a group, explain the variation of Y; with several independent variables its null is that all slope coefficients are zero. With one independent variable, it tests $H_0 : b_1 = 0$ vs. $H_a : b_1 \neq 0$, the same hypothesis as the t-test of the slope. Reject H_0 if $F > F_c$ (a one-tailed test using the upper tail of the F-distribution). Rejecting H_0 means X and Y have a significant linear relationship.

WORKED EXAMPLE

$n = 30$; $SSR = 64$ and $SSE = 112$, so $SST = 176$. $MSE = 112/28 = 4$; $SEE = \sqrt{4} = 2$; $R^2 = 64/176 = 0.364$; $r = \sqrt{0.364} = 0.60$ (if the slope is positive); $F = 64/4 = 16$, well above the 5% critical value of about 4.2 for 1 and 28 degrees of freedom, so the slope is significant.

t-test of a regression coefficient

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

To test significance, set $H_0 : b_1 = 0$ (two-tailed) and reject if $|t| > t_c$. The test of the slope is equivalent to testing the correlation, $t = r\sqrt{n-2}/\sqrt{1-r^2}$. With one independent variable, the F-statistic equals the square of this t-statistic, so the F-test and the t-test of the slope always reach the same decision (in the worked example above, $F = 16$ corresponds to $t = 4$). The intercept can be tested the same way ($H_0 : b_0 = 0$).

Indicator variables

An **indicator variable (dummy variable)** takes the value **1** when a condition holds and **0** otherwise (e.g., 1 in recession quarters). Regressing Y on the indicator tests whether the condition matters: the slope equals the **difference in the mean of Y** between periods with and without the condition, and testing $b_1 = 0$ is equivalent to a difference-in-means test. The intercept is the mean of Y in periods without the condition (indicator = 0), and intercept plus slope is the mean in periods with it. *Example.* Regressing a company's quarterly earnings on a recession indicator tests whether earnings are cyclical; for cyclical earnings the slope is expected to be negative (lower earnings in recession quarters). If quarterly EPS averaged 1.40 in expansion quarters and 0.95 in recession quarters, the fitted line is $\overline{\text{EPS}} = 1.40 - 0.45D$, where $D = 1$ in a recession quarter.

COMMON EXAM TRAPS

- $R^2 = \text{SSR}/\text{SST}$. SSE/SST is the unexplained share ($1 - R^2$), and SSR/SSE is not a standard measure. In the worked example, $112/176 = 0.636$ is the unexplained share; R^2 is 0.364.
- F is a ratio of mean squares, MSR/MSE. Dividing SSR by SSE ignores the degrees of freedom. In the worked example, $64/112 = 0.571$ instead of $F = 16$.
- SEE divides SSE by $n - 2$, the degrees of freedom, before taking the square root. Dividing by n understates it. In the worked example, $\sqrt{112/30} = 1.93$ instead of 2.
- A negative slope gives a negative r even though R^2 is positive. The standard error of the slope is a different statistic and is not a correlation.

EXAM SHORTCUTS

- With one independent variable, $F = t^2$ for the test of $H_0 : b_1 = 0$ against $H_a : b_1 \neq 0$, so at the same significance level the F-test and the two-tailed slope t-test reach the same decision and only one needs computing.
- When only R^2 and SSR are given, $\text{SST} = \text{SSR}/R^2$ and $\text{SSE} = \text{SST} - \text{SSR}$, which lead straight to the MSE and the SEE.
- In a simple regression $r = \pm\sqrt{R^2}$, taking the sign of the slope.

BOTTOM LINE

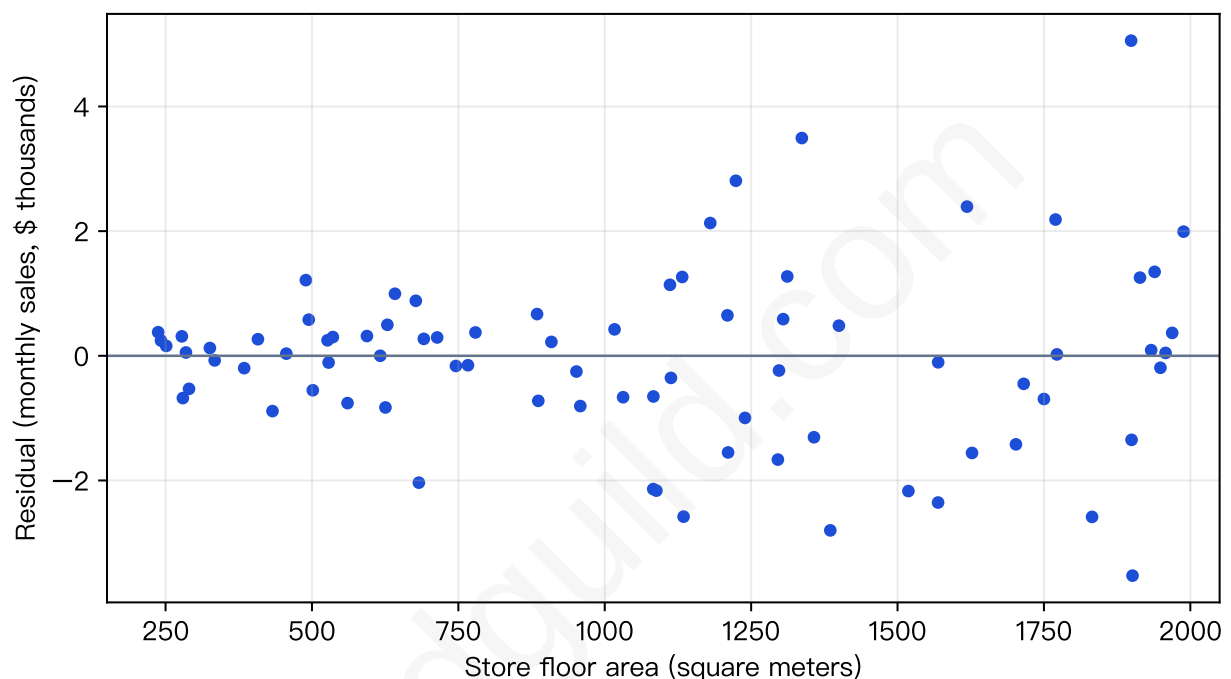
- Simple linear regression assumes a linear relationship, homoskedastic residuals, residuals uncorrelated with one another, and normally distributed residuals.
- Residual variance that widens or narrows with X or over time is heteroskedasticity, and residuals correlated with one another, such as a recurring seasonal pattern, show autocorrelation.
- ANOVA splits total variation into explained and unexplained parts, $\text{SST} = \text{SSR} + \text{SSE}$.
- In a simple regression $\text{MSE} = \text{SSE}/(n - 2)$, $\text{SEE} = \sqrt{\text{MSE}}$, and $R^2 = \text{SSR}/\text{SST} = 1 - \text{SSE}/\text{SST}$, which equals r^2 .
- $F = \text{MSR}/\text{MSE}$ with 1 and $n - 2$ degrees of freedom is a one-tailed test, and with one independent variable it tests $H_0 : b_1 = 0$, the same hypothesis as the slope's t-test.

- Regressing Y on an indicator variable gives a slope equal to the difference in the mean of Y between periods with and without the condition, and an intercept equal to the mean when the indicator is 0.

QUICK CHECK

Question 2

An analyst regresses the monthly sales of 80 stores on each store's floor area. The figure below plots the regression residuals against floor area. The pattern *most likely* indicates that the regression's residual term:



Residuals plotted against store floor area

- A does not have a constant variance.
- B is not normally distributed.
- C is not independently distributed.

Answer: A

The residuals fan out as floor area increases: their spread (variance) grows with the independent variable. A residual variance that is not constant across observations is heteroskedasticity, a violation of the homoskedasticity assumption.

Why the other options are wrong

- B.** The plot shows residuals centered on zero with a changing spread; nothing in it points specifically to a nonnormal shape, and normality is judged from the shape of the residuals' distribution. How their spread changes with X is a separate question.
- C.** Non-independence (autocorrelation) shows up as residuals that are correlated with one another, such as a recurring seasonal pattern over time. A widening spread with X is a variance problem.

Key takeaway Residual spread widening with X = heteroskedasticity (non-constant residual variance).

MODULE 10.3

Predicted Values and Functional Forms of Regression

LOS 10.c — Predicted values, prediction intervals and functional forms

Predicted values

A **predicted value** of the dependent variable plugs a forecast of the independent variable, X_p , into the estimated equation:

$$\hat{Y} = \hat{b}_0 + \hat{b}_1 X_p$$

Example: $\hat{Y} = 1.2 + 0.9X$ and $X_p = 4$: $\hat{Y} = 1.2 + 3.6 = 4.8$. Include the intercept and keep the sign of each term.

Confidence (prediction) interval for a predicted value

KEY CONCEPT

$$\hat{Y} \pm t_c s_f \quad df = n - 2$$

where s_f is the **standard error of the forecast**. At Level I it is usually given; if not:

KEY CONCEPT

$$s_f^2 = \text{SEE}^2 \left[1 + \frac{1}{n} + \frac{(X_p - \bar{X})^2}{(n-1)s_X^2} \right]$$

The forecast is more precise (smaller s_f) when the SEE is smaller, the sample is larger, X_p is closer to $\bar{X}$, and the independent variable has a greater variance.

Example (continued): $s_f = 1.5$, $n = 27$, so $df = 25$ and the two-tailed 5% critical value is 2.060. The 95% interval is $4.8 \pm 2.060(1.5) = 4.8 \pm 3.09$, i.e., 1.71 to 7.89.

Computing s_f when it is not given (a separate regression): $\text{SEE} = 2.0$, $n = 27$, $\bar{X} = 3$, $s_X^2 = 4$, $X_p = 4$:

$$s_f^2 = 2.0^2 \left[1 + \frac{1}{27} + \frac{(4-3)^2}{26 \times 4} \right] = 4(1 + 0.0370 + 0.0096) = 4.187, \quad s_f = 2.05$$

The standard error of the forecast is always larger than the SEE: the forecast carries both the residual uncertainty and the uncertainty in the estimated coefficients.

Interval	Centered on	Width driven by
Confidence interval for a coefficient	$\hat{b}_1$ (or $\hat{b}_0$)	Standard error of that coefficient
Prediction interval for Y	$\hat{Y}$ at X_p	Standard error of the forecast s_f

Functional forms

If the relationship between X and Y is not linear, taking natural logarithms of the dependent variable, the independent variable or both can make it linear. For example, EPS growing at a roughly constant percentage rate over time traces an upward curve.

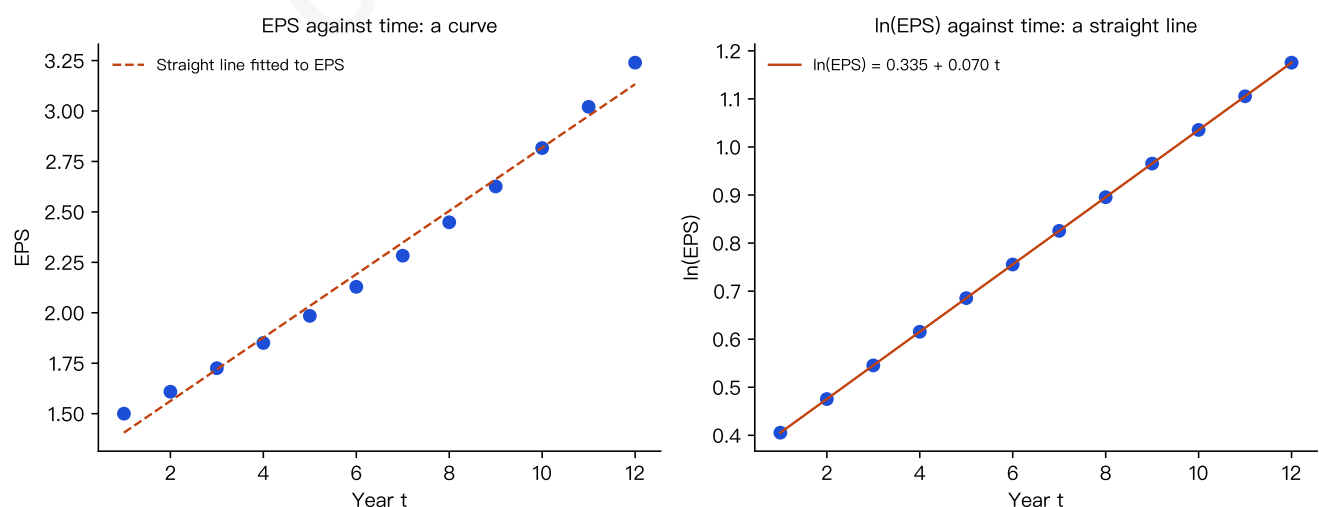
KEY CONCEPT

Model	Equation	Slope interpretation	Typical use
Log-lin model	$\ln Y_i = b_0 + b_1 X_i + \varepsilon_i$	Relative (%) change in Y for an absolute change in X	Constant growth rate over time
Lin-log model	$Y_i = b_0 + b_1 \ln X_i + \varepsilon_i$	Absolute change in Y for a relative (%) change in X	Y responds to proportional changes in X
Log-log model	$\ln Y_i = b_0 + b_1 \ln X_i + \varepsilon_i$	Relative change in Y for a relative change in X (an elasticity)	Both variables in proportional terms

Example. In a log-lin regression of a company's annual EPS on a time index (1, 2, 3, ...), a slope of 0.07 means EPS grows by roughly 7% a year (exactly $e^{0.07} - 1 = 7.25\%$). It does not mean EPS rises by 0.07 currency units a year.

Choose the form by the nature of the variables and by comparing goodness-of-fit measures and residual plots.

- Exam convention: a higher R^2 , a higher F-statistic and a lower SEE indicate the better-fitting functional form.
- Current practice: R^2 and SEE are directly comparable only when the dependent variable is the same; a model of $\ln Y$ and a model of Y measure fit on different scales.



Log-lin model: EPS growing about 7% a year (illustrative)

COMMON EXAM TRAPS

- Prediction intervals use $df = n - 2$ in a simple regression (two estimated parameters), not $n - 1$. In the example, 26 degrees of freedom give a critical value of 2.056 and an interval of 1.72 to 7.88 instead of 1.71 to 7.89.
- A prediction interval uses the standard error of the forecast s_f . The standard error of the slope belongs to the interval for the slope coefficient.
- "Log" is attached to the variable that is logged: log-lin means Y is logged and X is linear.

EXAM SHORTCUTS

- The standard error of the forecast is always larger than the SEE, so a prediction interval narrower than $\hat{Y} \pm t_c \times \text{SEE}$ can be ruled out.

BOTTOM LINE

- A predicted value is $\hat{Y} = \hat{b}_0 + \hat{b}_1 X_p$, using the intercept and the sign of each term.
- The prediction interval for Y is $\hat{Y} \pm t_c s_f$ with $n - 2$ degrees of freedom, where s_f is the standard error of the forecast.
- The forecast is more precise when the SEE is smaller, the sample is larger, X_p is closer to $\bar{X}$ and the independent variable has a greater variance.
- A log-lin model ($\ln Y$ on X) has a slope giving the relative change in Y for an absolute change in X, a lin-log model the absolute change in Y for a relative change in X, and a log-log model an elasticity.
- Exam convention: a higher R^2 , a higher F-statistic and a lower SEE identify the better-fitting functional form; current practice: R^2 and SEE are directly comparable only when the dependent variable is the same.

QUICK CHECK**Question 3**

An analyst regresses a fertilizer producer's monthly stock returns on monthly changes in a crop price index, using 32 observations. The estimated coefficients are in the table below. She expects next month's change in the index to be -0.015 and computes a standard error of the forecast of 0.028 . Critical t-values with 30 degrees of freedom are 1.310 with 10% in one tail and 1.697 with 5% in one tail. The 90% prediction interval for next month's return is *closest* to:

Regression of monthly stock returns on crop price changes

	Coefficient	Standard error
Intercept	0.0040	0.0035
Crop price change	0.6500	0.2100

A -0.0533 to 0.0418 .

B -0.0424 to 0.0309.

C -0.0338 to 0.0613.

Answer: A

A prediction interval is the predicted value plus or minus the critical t-value times the standard error of the forecast. A two-sided 90% interval leaves 5% in each tail, so with $n - 2 = 30$ degrees of freedom the critical value is 1.697.

- Predicted return: $\hat{Y} = 0.0040 + 0.6500(-0.015) = 0.0040 - 0.00975 = -0.00575$
- Half-width: $t_c \times s_f = 1.697 \times 0.028 = 0.0475$
- Interval: -0.00575 ± 0.0475 , that is, -0.0533 to 0.0418

Calculator: $0.65 \times 0.015 [=]$ gives 0.00975; $1.697 \times 0.028 [=]$ gives 0.047516.

Why the other options are wrong

- **B.** This interval uses 1.310, the critical value with 10% in one tail. A two-sided 90% interval splits the 10% between the two tails, so each tail holds 5%.
- **C.** This interval is centered on $0.0040 + 0.00975 = 0.01375$, which drops the minus sign of the forecast change in the index.

Key takeaway Prediction interval = $\hat{Y} \pm t_c s_f$, with $n - 2$ degrees of freedom and $\alpha/2$ in each tail. Keep the sign of the forecast X when computing $\hat{Y}$.

MODULE 10.4

Estimating CAPM Values With Simple Linear Regression

LOS 10.d — Estimating the CAPM with simple linear regression

The CAPM as a one-factor regression

The capital asset pricing model states

$$E(R_i) = R_f + \beta_i [E(R_M) - R_f]$$

Because the only explanatory variable is the **market risk premium**, the CAPM is a **one-factor risk model** and can be estimated with simple linear regression, using past observations as proxies for expected values.

Subtracting R_f from both sides gives the estimable form:

KEY CONCEPT

$$Z_{i,t} = b_0 + b_1 Z_{M,t} + \varepsilon_t, \quad Z_{i,t} = R_{i,t} - R_{f,t}, \quad Z_{M,t} = R_{M,t} - R_{f,t}$$

KEY CONCEPT

Regression element	CAPM meaning
Dependent variable	The asset's excess return ($R_i - R_f$)
Independent variable	The market risk premium ($R_M - R_f$)
Slope $\hat{b}_1$	The asset's estimated beta
Intercept $\hat{b}_0$	Return not explained by market exposure (zero if the CAPM holds exactly)

Steps

1. **Collect data:** the asset's returns; a **security market index** as the proxy for the **market return**; yields on **short-term government debt** as the proxy for the **risk-free rate**. At least 30 observations is the usual minimum (e.g., 1–3 years of weekly data or 3–5 years of monthly data).
2. Compute the asset's excess returns and the market risk premium for each period.
3. Run the least-squares regression of excess returns on the market risk premium.
4. Use $\hat{\beta}$ and an expected market risk premium to estimate the expected return: $E(R_i) = R_f + \hat{\beta}[E(R_M) - R_f]$.

Testing and interpreting beta

Example: the slope is 1.25 with a standard error of 0.10 (large sample, so use z-values).

- Beta above 1: the stock's excess returns amplify the market's; below 1: they are dampened.
- Test $H_0 : \beta = 0$: $t = 1.25/0.10 = 12.5 > 1.96$, so reject; beta is significantly different from zero.
- 95% confidence interval: $1.25 \pm 1.96(0.10) = 1.054$ to 1.446 .
- Test $H_0 : \beta = 1$ (does the stock move one-for-one with the market?): $t = (1.25 - 1)/0.10 = 2.5 > 1.96$, so reject; beta is significantly different from 1, and the estimate lies above it. This matches the interval above, which does not contain 1.
- A one-sided question, such as whether beta exceeds 1, uses $H_0 : \beta \leq 1$ versus $H_a : \beta > 1$, the same statistic $(\hat{b}_1 - 1)/s_{\hat{b}_1}$ and a one-tailed critical value.
- With $R_f = 3\%$ and an expected market risk premium of 5%: $E(R_i) = 3\% + 1.25(5\%) = 9.25\%$.

Changing inputs: $\Delta E(R_i) = \Delta R_f + \hat{\beta} \Delta[E(R_M) - R_f]$. With $\hat{\beta} = 0.8$, a 0.4% rise in R_f combined with a 0.4% fall in the market risk premium changes expected return by $0.4\% - 0.8(0.4\%) = +0.08\%$; for a beta above 1 the same changes would lower it.

Reading a typical output

	Coefficient	Standard error	t-statistic
Intercept	0.05	0.08	0.63
Market risk premium (slope)	1.25	0.10	12.50

The slope row carries beta and its test; the intercept row is not significant here ($t = 0.63$), consistent with the CAPM's prediction of a zero intercept. The R^2 of such a regression tells how much of the asset's excess-return variation is systematic (market-driven); $1 - R^2$ is the share that is unsystematic.

Choosing the sample

Longer samples give more observations and more precise estimates but risk including periods when the company's business mix, leverage and therefore beta were different. More frequent data (daily or weekly) add observations quickly but can be noisy for thinly traded assets. The choice depends on the purpose: a short-term speculator favors recent high-frequency data; a buy-and-hold investor favors monthly or quarterly data spanning at least one full business cycle.

WORKED EXAMPLE: ESTIMATING AND TESTING A CAPM BETA

Sixty monthly observations of a stock's excess return and the market risk premium give a covariance of 0.00168 and a variance of the market risk premium of 0.00140. The regression reports a standard error of 0.16 for the slope, and with 60 observations the sample is large, so the two-tailed 5% critical value is 1.96. Estimate beta, test at the 5% level whether beta differs from 1, build the matching confidence interval for beta, and estimate the stock's expected return with a risk-free rate of 2.8% and an expected market risk premium of 5.5%.

Step 1. Beta is the slope coefficient: $\hat{b}_1 = 0.00168/0.00140 = 1.20$.

Step 2. Test $H_0 : \beta = 1$ against $H_a : \beta \neq 1$: $t = (1.20 - 1)/0.16 = 1.25$. Because 1.25 is below 1.96, H_0 is not rejected, so beta is not significantly different from 1.

Step 3. The 95% confidence interval is $1.20 \pm 1.96(0.16)$, or 0.886 to 1.514. It contains 1, which agrees with Step 2.

Step 4. Expected return: $E(R_i) = 2.8\% + 1.20(5.5\%) = 9.4\%$.

Result. The estimated beta is 1.20, not significantly different from 1 at the 5% level, and the CAPM expected return is 9.4%.

COMMON EXAM TRAPS

- The dependent variable is the asset's excess return and the independent variable is the market risk premium. A regression of raw returns on raw returns is not the CAPM specification.
- Beta is the slope coefficient. The intercept says nothing about beta, so a negative intercept does not make beta negative.
- Confidence interval for beta: slope $\pm$ critical value $\times$ standard error of the slope. The intercept's standard error plays no part. With the output above, using the intercept's 0.08 gives 1.093 to 1.407 instead of 1.054 to 1.446.

EXAM SHORTCUTS

- A two-tailed 95% confidence interval for beta that excludes 1 means $H_0 : \beta = 1$ is rejected at the 5% level, and one that includes 1 means it is not, so either result gives the other.

BOTTOM LINE

- The CAPM is estimated by regressing the asset's excess return $R_i - R_f$ on the market risk premium $R_M - R_f$; the slope is the asset's beta and the intercept is zero if the CAPM holds exactly.
- A security market index proxies the market return, short-term government debt yields proxy the risk-free rate, and at least 30 observations is the usual minimum.

- Beta is tested with $t = (\hat{b}_1 - \text{hypothesized value}) / s_{\hat{b}_1}$, and its confidence interval is the slope $\pm$ critical value $\times$ the slope's standard error.
- The expected return is $E(R_i) = R_f + \hat{\beta}[E(R_M) - R_f]$, and a change in inputs gives $\Delta E(R_i) = \Delta R_f + \hat{\beta} \Delta[E(R_M) - R_f]$.
- The R^2 of the CAPM regression is the share of the asset's excess-return variation that is systematic, and $1 - R^2$ is the unsystematic share.
- Longer samples give more precise estimates but may include periods with a different beta, and more frequent data can be noisy for thinly traded assets.

QUICK CHECK

Question 4

To estimate a stock's beta, an analyst regresses the stock's excess returns on the market risk premium and obtains an intercept of -0.15 and a slope of 0.72 . Based on this model, the stock's beta is:

- A negative.
- B between 0 and 1.
- C greater than 1.

Answer: B

In the regression $(R_i - R_f) = b_0 + b_1(R_M - R_f) + \varepsilon$, the slope coefficient is the estimate of the asset's beta. The slope is 0.72 , so beta lies between 0 and 1: the stock's excess returns tend to move less than one-for-one with the market risk premium.

$$(R_i - R_f) = -0.15 + 0.72(R_M - R_f) \Rightarrow \hat{\beta} = \hat{b}_1 = 0.72$$

Why the other options are wrong

- **A.** The negative number is the intercept, which is not beta. Beta is the slope, 0.72 .
- **C.** A beta above 1 would require a slope above 1; the estimated slope is 0.72 .

Key takeaway Beta = slope of the excess-return regression; the intercept is irrelevant to beta.

Why each answer is right and why each other option is wrong: credguild.com/cfa/level-1/notes/quantitative-methods/least-squares-regression/. The full bank has 29 questions on this reading.

QUANTITATIVE METHODS · READING 11

Machine Learning in Finance

What you'll learn

LOS 11.a Describe how big data, machine learning and artificial intelligence are used in financial data science, fintech and investment management.

MODULE 11.1

Introduction to Financial Data Science

This reading describes how financial data science draws on big data: its types, sources and characteristics, and how the data are processed and visualized. It then explains data mining and machine learning (supervised, unsupervised and reinforcement), overfitting and underfitting, the split into training, validation and test datasets, neural networks, and natural language processing and large language models.

LOS 11.a — Big data, machine learning and AI in finance

What financial data science is for

Financial data science applies both quantitative and qualitative analysis to generate insight into a specific financial question, for example how customers behave or where market prices may move next. Among its main uses in finance are (1) evaluating investment opportunities, (2) optimizing portfolios and (3) mitigating risk.

Fintech is where finance and technology meet. Of most interest in investment management are the fintech developments that aid quantitative analysis. The word is also used more broadly for technology developments applied to financial services and for the companies that build them.

The objective is the insight itself. Data-quality work, described below under the characteristics of big data and under curation, supports that objective but is not the objective.

Big data: types and sources

Big data is the broad term for all the potentially useful information being generated. By format it is:

Type	What it looks like	Examples
Structured data	Rows and columns in a table or database	Financial statement line items, daily closing prices
Semi-structured data	Not a table, but tagged so it can be stored in an organized way	HTML code of a web page
Unstructured data	No predefined organization; usually needs ML/AI or special code to use	Social media posts, voice recordings, pictures, sensor output

Data come from **traditional sources** (markets, company financial reports, regulatory filings, government statistics) or **alternative data** sources:

- **Individuals:** social media activity, clicks, searches, time spent on web pages (mostly unstructured and growing fast).
- **Business processes:** transaction and corporate records such as sales data or banking records (usually structured). They can act as leading or real-time indicators, whereas quarterly reports give a lagging view.

- **Sensors:** smartphones, connected cars, smart buildings, satellites. The network of these devices is the **Internet of Things (IoT)**; it produces the largest volumes, mostly unstructured.

Buying alternative data or collecting it by **web scraping** (automated extraction from websites) raises legal and ethical issues: some of what is captured may be personal information protected by privacy rules.

The "Vs" of big data

KEY CONCEPT

Characteristic	Question it answers	Vocabulary to recognize
Volume	How much data?	megabytes → gigabytes → terabytes → petabytes
Velocity	How fast are data communicated?	low latency = real time; high latency = periodic or delayed
Variety	In what structures do the data exist?	structured, semi-structured, unstructured
Veracity	Are the data reliable and credible?	fake posts, errors, doubtful sources

A terabyte is 1,000 gigabytes and a petabyte is 1,000 terabytes. Faster data generate more of it: volumes rise from batch collection (megabytes) through periodic (gigabytes) and near-real-time (terabytes) to real-time feeds (petabytes). More complex structures also need more storage.

The first three are the classic **3 Vs** (volume, velocity, variety); **veracity** is the fourth V that matters especially for financial data. Other data-quality concerns are outliers, bad or missing data, sampling biases, and whether the amount collected is sufficient and appropriate for the intended use.

Data science: processing and visualization

Data science is the field concerned with extracting information from big data. Its processing methods:

Method	What happens	Do not confuse with
Capture	Collecting data and converting them into a usable form	Transfer
Curation	Ensuring data quality: fixing errors, adjusting for bad or missing data	Capture
Storage	Archiving data and accessing them later; choice depends on structure and latency	Search
Search	Querying stored data to find what is needed	Storage
Transfer	Moving data from the source or storage medium to the analytical tool	Capture

A **data lake** is a central repository that keeps structured, semi-structured and unstructured data in their raw, original format. **Distributed computing frameworks** let data be processed and analyzed on many machines at once.

Visualization: charts and graphs suit structured data; less structured data call for **word clouds** (show how often words occur in a text sample), **heat maps**, 3D graphics and **mind maps** (show logical relations between concepts).

Typical inputs include share prices and trading volumes, asset returns, interest rates, exchange rates, accounting figures and macroeconomic series. The work calls for expertise in finance as well as in data handling, because much of the data is sequential, time dependent and sometimes cyclical. Financial data bring special problems: noise, nonstationarity, interdependencies, nonlinearity, **fat tails** (extreme events more likely), very **high-frequency data** in trading, and too little data in some markets (e.g., emerging markets). Regulated firms must meet strict rules on data management, risk assessment and model validation; firms outside financial regulation still have to comply with business law and their contracts, although they have more freedom in how they handle data. Data privacy, thorough model testing and human review of model decisions matter for every firm.

Data mining, AI and machine learning

Data mining searches large datasets (usually structured) for patterns that can be used to predict outputs; pattern identification rests on conditional probabilities (e.g., Bayes' theorem). Its key risk is **overfitting**.

Artificial intelligence (AI) refers to computer systems that simulate human thinking. **Machine learning (ML)** is a major subcategory of AI: an algorithm receives input data, without any assumption about how they are distributed, possibly with target outputs, and learns without human help to model the outputs or to find patterns.

KEY CONCEPT

Approach	Data given	What the machine does	Typical finance use
Supervised learning	Labeled inputs and outputs	Learns to model outputs from inputs, then predicts on new data	Forecasting company or stock performance, default prediction
Unsupervised learning	Unlabeled inputs, no outputs	Describes the structure of the data and finds patterns; reduces dimensionality	Customer segmentation, anomaly detection
Reinforcement learning	Feedback on its own actions	Learns by trial and error from rewards and penalties	Portfolio optimization (sequential decisions)

A supervised model trained on a large, reliable dataset should predict well but may not adapt to new situations. Reinforcement learning is costly: designing a reward system that actually improves the model can take much training and many resources. Whatever the method, people with subject expertise are needed to choose the model, clean the inputs, limit bias and interpret the output.

Deep learning uses many layers of neural networks to find patterns of increasing complexity and may use supervised, unsupervised or semi-supervised learning (image and speech recognition, natural language processing).

Overfitting vs. underfitting

KEY CONCEPT

	Overfitting	Underfitting
Model is	Too complex	Not complex enough
Error	Treats noise as a true pattern	Treats true parameters as noise
Result	Spurious relationships, poor generalization to new data	Misses actual patterns and relationships

Other ML issues: results can be a "**black box**" (hard to explain), and **data leakage** occurs when information from outside the training dataset influences what the model learns.

Training, validation and test datasets

- **Training dataset:** the model looks for relationships here; it is the largest set, about 60%–80% of the data (upper end for large datasets).
 - **Validation dataset:** used to refine (tune) the relationships found in training.
 - **Test dataset:** used last, to judge the model's predictive ability.
- The validation and test sets share the remaining 20%–40% equally. Observations are normally assigned at random, which reduces bias, overfitting and data leakage. **Time series** data must keep their chronological order instead:
- **Time-based split:** the oldest 60%–80% of observations go to training, the next 10%–20% to validation and the most recent 10%–20% to testing.
 - **Rolling-window validation:** train on one window, validate on a later period, then shift the window forward and repeat.

Either way, the span covered by validation plus test data should match the intended forecast horizon.

Example. A credit model has 48,000 randomly ordered loan records and uses a 75% / 12.5% / 12.5% split. Training gets 36,000 records, validation 6,000 and test 6,000. If instead the data were 12 years of monthly returns (144 months) and a time-based split with the same proportions were used, months 1–108 would train the model, months 109–126 would validate it and months 127–144 would test it.

Neural networks, NLP and LLMs

Neural networks suit nonlinear relationships with many interdependencies: an input layer, several **hidden layers** of weighted nodes (**neurons**), and an output layer. They can learn in a supervised or unsupervised way, and each neuron's weight reflects its importance. They are often trained by **backpropagation** (a form of supervised learning): a **forward pass** produces a prediction that is compared with the actual outcome, the loss is computed, the **loss gradient** shows how weight changes affect the error, and **gradient descent** adjusts the weights step by step to minimize the loss.

- **Generative adversarial networks (GANs)** pit a generator network against a discriminator network to create realistic synthetic data (e.g., plausible market scenarios for stress tests, trained on historical market data). Because GANs can also produce convincing fake material, their output needs careful review.

- **Variational autoencoders (VAEs)** compress data and then reconstruct them using encoder and decoder networks. They are probabilistic generative models, used to reduce dimensionality and spot anomalies.
- **Copulas** (e.g., the Gaussian copula) are a more traditional mathematical tool for nonlinear relationships that uses **marginal probability distributions** to capture dependencies between random variables. Exam convention: the Cholesky decomposition is given as an example of a copula. Current practice: the Cholesky decomposition is a matrix factorization, used to generate the correlated normal draws in a Gaussian-copula simulation.

Natural language processing (NLP) lets computers interpret human language: speech recognition, translation, scanning employee communications for compliance, and detecting sentiment shifts in research reports or earnings-call transcripts. **Text analytics** works mainly with unstructured material, such as filings and research reports or recordings of earnings calls and surveys. NLP converts that material into usable data much more quickly than a person could and searches it for patterns; **lexical analysis**, for example, measures how often each word appears. **Large language models (LLMs)** are generative AI that produce human-like text, using neural networks to read context and meaning in qualitative inputs (nonfinancial inputs such as election polls can matter too). Their output is text such as a summary or a forecast based on tone and sentiment. They are typically trained on a very large dataset through **self-supervised learning**, an unsupervised approach in which the model generates the labels itself from raw data. They are not built to catch errors in financial data, so outputs need expert review and should be used alongside traditional data-driven models.

Common tools are **Python** (fintech apps), **R** (statistics), **Java** (runs across platforms) and **C/C++** (high-speed algorithmic trading). Common databases are **SQL** (structured data, on a server), **SQLite** (structured data, not on a server; common in mobile apps) and **NoSQL** (unstructured data).

COMMON EXAM TRAPS

- Business-process records such as sales data are usually structured, yet they count as alternative data. Structure does not decide whether a source is traditional.

EXAM SHORTCUTS

- Classify a data source by where it comes from, not by its structure: records from individuals, business processes and sensors are alternative data even when they are structured.
- Match the type of machine learning to what the algorithm is given: labeled inputs and outputs mean supervised learning, unlabeled inputs alone mean unsupervised learning, and rewards and penalties for its own actions mean reinforcement learning.

BOTTOM LINE

- Financial data science applies quantitative and qualitative analysis to gain insight into a specific financial question, with main uses in evaluating investment opportunities, optimizing portfolios and mitigating risk.
- Big data can be structured (rows and columns), semi-structured (tagged, such as HTML code) or unstructured (social media posts, voice recordings, pictures, sensor output), and comes from traditional sources or from alternative sources: individuals, business processes and sensors.
- The characteristics of big data are volume, velocity (low latency means real time), variety of structures and, for financial data especially, veracity, meaning whether the data are reliable and credible.

- Data science processes data through capture, curation (ensuring data quality), storage, search and transfer, and a data lake keeps structured, semi-structured and unstructured data in their raw format.
- Data mining searches large datasets for patterns to predict outputs, while machine learning, a subcategory of AI, learns from input data without assumptions about their distribution and without human help.
- Supervised learning uses labeled inputs and outputs, unsupervised learning finds structure in unlabeled inputs, and reinforcement learning learns by trial and error from rewards and penalties.
- Overfitting comes from a model that is too complex and treats noise as a true pattern; underfitting comes from one that is not complex enough and treats true parameters as noise.
- The training dataset is the largest, about 60%–80% of the data, with validation and test sets splitting the rest equally; observations are normally assigned at random, but time series data keep their chronological order through a time-based split or rolling-window validation.

QUICK CHECK

Question 1

A new analyst at Ottery Capital drafts three glossary entries for the firm's research team. Which entry about fintech is *most accurate*?

- A** Fintech covers only payment apps and online banking services sold to consumers.
- B** Fintech is another name for financial data science, the analysis of financial data.
- C** Fintech is the meeting of finance and technology, such as tools that aid quantitative analysis.

Answer: C

Fintech refers to the meeting of the finance and technology worlds. In investment management the developments of most interest are those that aid quantitative analysis, such as big data tools and machine learning. The term is also used more broadly for technology applied to financial services and for the firms that build it.

Why the other options are wrong

- **A.** Consumer payment and banking apps are one use of fintech. Limiting the term to them leaves out the investment tools, such as machine learning and big data processing, that matter most to analysts.
- **B.** Financial data science is the quantitative and qualitative analysis of financial information to gain insight into a financial question. Fintech is the technology that supports such work and other financial services, so the two terms are related but different.

Key takeaway Fintech is where finance meets technology; financial data science is the analysis that some of that technology makes possible.

Practice questions

Question 2

Voice recordings of earnings calls, satellite images of shipping ports and posts on social media platforms are all examples of:

- A unstructured data.
- B structured data.
- C semi-structured data.

Question 3

A data team at Larchmont Analytics is documenting the steps it uses to process data. Which of the following statements about these data processing methods is *most accurate*?

- A Search is the process of archiving data so that they can be retrieved later.
- B Curation involves cleaning the data, for example by correcting errors and dealing with missing values, so that the data are of good quality.
- C Capture is the process of moving data from a storage medium to the analytical tool that will use them.

Question 4

A lender at Brackenridge Credit trains a model on records of past borrowers, each tagged with the borrower's financial ratios and whether the borrower eventually defaulted. The model learns to map the ratios to the default outcome and is then applied to new loan applicants. This machine learning process is best described as:

- A unsupervised learning.
- B supervised learning.
- C reinforcement learning.

Answers: 2. A · 3. B · 4. B

Why each answer is right and why each other option is wrong: credguild.com/cfa/level-1/notes/quantitative-methods/machine-learning-in-finance/. The full bank has 25 questions on this reading.

Keep going at credguild.com

The notes are the start. These are free on the site and do what a PDF cannot:

- **Diagnostic test:** 36 questions weighted like the exam, a score in each topic and the topics to start with.
credguild.com/cfa/level-1/diagnostic/
- **Practice questions:** 398 free questions with full explanations, scored as you answer. A free account answers any question in the bank, up to 50 new questions a day. credguild.com/cfa/level-1/practice/
- **Notes:** every reading of the 2027 curriculum, with worked examples and exam traps, on the site and as a PDF. credguild.com/cfa/level-1/notes/
- **Study plan:** pick your exam window and weekly hours and get a week-by-week schedule.
credguild.com/cfa/level-1/plan/
- **Mock exam:** 180 questions in two sessions, on the 2027 topic weights; the first one is free with an account.
credguild.com/cfa/level-1/mock/
- **Cheat sheet and formula sheet:** every reading's bottom line and every formula, by topic.
credguild.com/cfa/level-1/cheat-sheet/

Found a mistake? Email hello@credguild.com.