# Module - Foundations for Quantitative Analysis

Yash Bhasein

19th August, 2026

*(Its recommended to go through the fa and ta modules first and then start this module)*

## Where am I starting from?

I know how to analyze an individual stock. Give me a company and I can work through its financial statements, sanity-check its valuation, and read a price chart for entry and exit signals. That skill answers one question well: is this stock, on its own, worth owning?

But investing is not just about finding good stocks. What happens when I actually have to manage a portfolio? Suppose I have ₹10,00,000 and five stocks I like - Reliance Industries, TCS, HDFC Bank, Infosys, and ITC. Buying all five doesn't automatically tell me how much of my ₹10 lakh should sit in each one, or whether the five of them, held together, behave the way I'd expect from looking at each one individually.

That gap, between analyzing a stock and managing a portfolio, is where this module starts. I'll move from thinking about single securities to thinking about portfolios, risk, pricing, market behaviour, and systematic factors.

## 1. Portfolio management:

Suppose I go ahead and buy all five stocks. If each one is genuinely a good pick, is the result automatically a good portfolio?
Not necessarily. Stock selection asks: is this asset attractive on its own? Portfolio management asks a different question: given a set of assets, how do I combine them, in what proportions, held for how long, rebalanced how often, to meet a specific objective at a level of risk I can live with.

*A good stock does not automatically create a good portfolio. The way the stocks in it interact matters as much as the quality of any single one.*

Portfolio management fundamentally revolves around balancing risk and return through strategic asset allocation. The portfolio's risk, measured by standard deviation of returns, is determined not only by individual asset risks but also by the correlation between assets.

Portfolio construction bundles four separate jobs that get run together in casual conversation, but are worth keeping distinct:

- Asset allocation: the split across asset classes (equity, debt, gold, cash) before I even pick individual names.
- Security selection: inside the equity sleeve, which specific stocks. This is the fundamental and technical analysis skill I already have.
- Position sizing: how much capital goes into each selected stock.
- Diversification: spreading capital so that no single stock, sector, or event can do disproportionate damage to the whole portfolio.

Diversification is the cornerstone strategy for risk reduction in portfolio management. By spreading investments across different assets with varying correlations, investors can reduce unsystematic (firm-specific) risk while maintaining exposure to expected returns.

Diversification's benefit doesn't keep growing forever, either. Research on portfolio construction indicates that a well-diversified portfolio typically needs only about 20 to 30 assets. Beyond that point, the marginal reduction in risk from adding more names diminishes significantly, while complexity keeps rising.

For a basic two-asset portfolio, risk is calculated as:

$$\sigma_{portfolio} = \sqrt{w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\rho_{1,2}\sigma_1\sigma_2}$$

Where:

| | | |
|---|---|---|
| $W_1$ | = | Proportion of the portfolio invested in Asset 1 |
| $W_2$ | = | Proportion of the portfolio invested in Asset 2 |
| $\sigma_1$ | = | Asset 1 standard deviation of returns |
| $\sigma_2$ | = | Asset 2 standard deviation of returns |
| $\rho_{1,2}$ | = | Correlation coefficient between the returns of Asset 1 and Asset 2 |

Portfolio Risk isn’t a weighted average. Let’s understand this with an example:

Take just two of the five holdings, RIL (Reliance Industries Ltd.) and TCS. Say, RIL has an expected return of 14% and volatility (standard deviation) of 22%; TCS has an expected

return of 13% and volatility of 19%. If I put half my money in each, is the portfolio's risk simply the average of 22% and 19%, i.e. 20.5%?

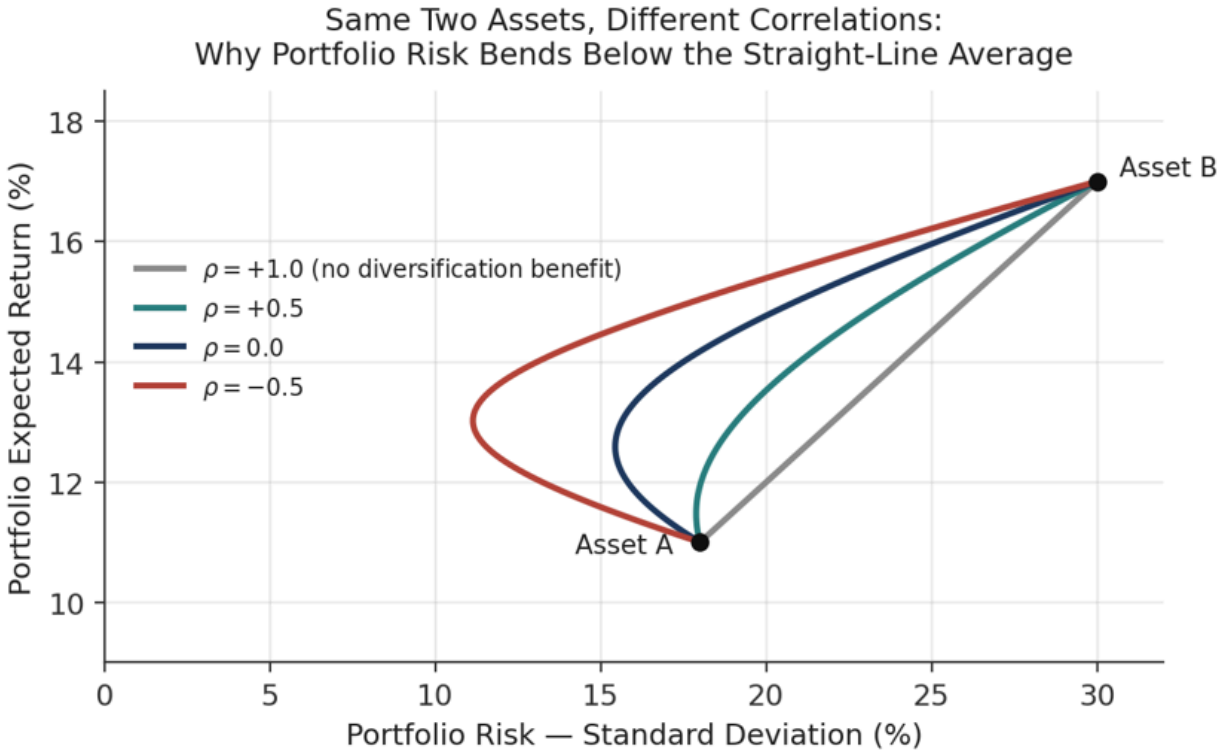
Only if the two stocks always moved in perfect lockstep. In reality they don't, some days one is up while the other is flat or down. That partial independence is exactly what diversification captures, and it is measured by covariance and correlation (the covariance rescaled to sit between −1 and +1).

Let's assume that the correlation between RIL and TCS is 0.4. Plug the values in the formula given above and you will get:

$\sigma_p^2 = (0.5^2 \times 0.22^2) + (0.5^2 \times 0.19^2) + (2 \times 0.5 \times 0.5 \times 0.22 \times 0.19 \times 0.4) = 0.0295 \rightarrow \sigma_p \approx$ **17.2%**

Compare that to the naive weighted average of the two individual volatilities: 0.5(22%) + 0.5(19%) = 20.5%. The actual portfolio risk, 17.2%, is meaningfully lower. That gap is the entire mathematical case for diversification.

*Portfolio risk depends on how assets move together, not just on how risky each asset is individually. The lower the correlation, the greater the reduction in risk from combining them.*




At ρ = +1, the combination line is straight, with no diversification benefit at all. As correlation falls, the achievable combinations bow outward and to the left, meaning that for the same expected return, lower risk becomes achievable.

Correlation analysis is essential for effective portfolio diversification. A correlation matrix shows the pairwise correlations between assets in a portfolio, with values ranging from -1 (perfect negative correlation) to +1 (perfect positive correlation).
The Correlation Matrix: The ideal correlation matrix for a portfolio has the structure:

$$\begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1n} \\ P_{21} & P_{22} & \cdots & P_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ P_{n1} & P_{n2} & \cdots & P_{nn} \end{bmatrix}$$

Where each element $P_{ij}$ represents the correlation coefficient between assets i and j. The diagonal elements $P_{ii}$ always equal 1, as each asset perfectly correlates with itself.

Interpreting Correlation Values

- Strong Positive Correlation (Close to +1): Assets tend to move in the same direction. When one asset increases in value, the other also tends to increase.
- Strong Negative Correlation (Close to -1): Assets tend to move in opposite directions, providing excellent diversification benefits.
- No Correlation (Close to 0): Assets move independently of each other, also providing good diversification benefits

## 2. Modern Portfolio Theory and Efficient Frontier

With more than two assets, every possible combination of weights produces a corresponding risk-return point, together tracing out a feasible set - the full cloud of achievable portfolios. Harry Markowitz's 1952 insight, which founded Modern Portfolio Theory (MPT) and later won him the Nobel Prize in Economics, was that within that cloud, most combinations are simply worse than others available at the same risk level.

*MPT's central claim: an investor should evaluate the risk and return of the entire portfolio together, rather than judging securities one at a time.*

Two obvious preferences define which portfolios in the feasible set are worth holding:

- Among portfolios offering the same expected return, I prefer the one with lower risk.
- Among portfolios with the same risk, I prefer the one with higher expected return.

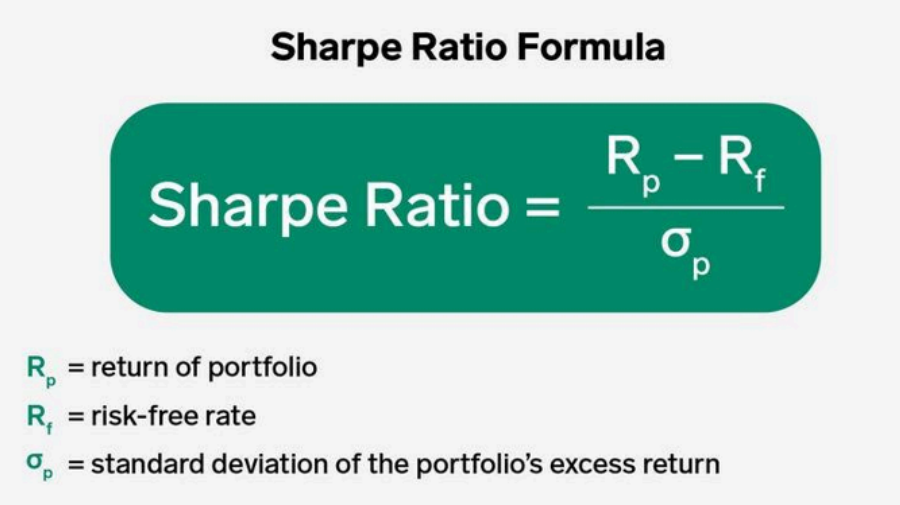
The minimum-variance portfolio is the single feasible portfolio with the lowest possible risk, regardless of return. Portfolios that satisfy both preferences above trace the efficient frontier.

The efficient frontier represents the set of optimal portfolios that offer the highest expected return for a defined level of risk or the lowest risk for a given level of expected return.

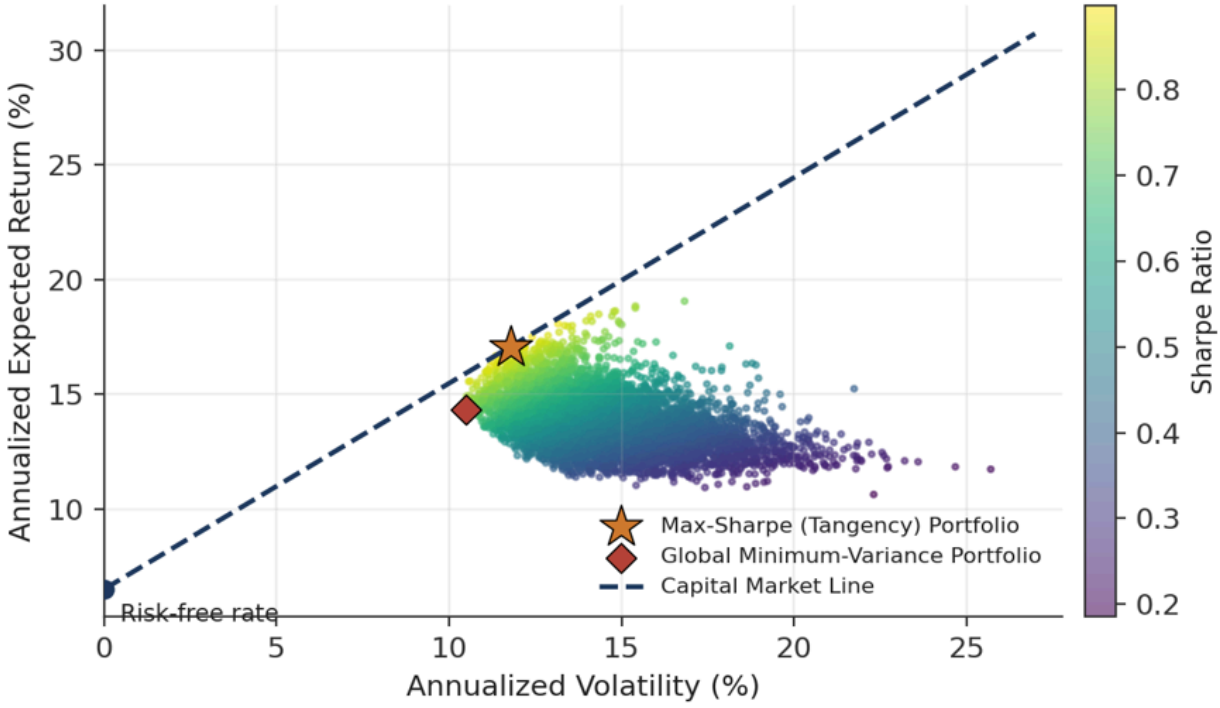
Key Characteristics of the Efficient Frontier

1. The efficient frontier forms a curved line in risk-return space, demonstrating that there is a diminishing marginal return to risk.
2. Portfolios that lie below the curve are suboptimal because they don't provide sufficient return for their risk level.
3. The curvature of the frontier illustrates the benefit of diversification - combining assets with different correlations can create portfolios with better risk-return profiles than individual assets.

**Sharpe Ratio**: The Sharpe ratio measures excess return per unit of total risk, with higher values indicating better risk-adjusted performance.




If I add a risk-free asset to the mix, the best achievable combinations of risk-free asset and risky portfolio lie along the Capital Market Line (CML), the straight line tangent to the efficient frontier from the risk-free rate. The point of tangency is the maximum-Sharpe portfolio




The efficient frontier is essentially the set of portfolios where I cannot get more return without accepting more risk and the Capital Market Line shows the best way to combine that frontier with a risk-free asset.

## 3. Risk-Return Relationships and CAPM

I understand how to construct a portfolio and locate it on the efficient frontier. But that only tells me about combinations of assets I already hold. It doesn't tell me what return I should expect from taking on risk in the first place. That is the question the Capital Asset Pricing Model (CAPM) tries to answer.

When a risk-free asset is introduced to the efficient frontier model, investors can create portfolios along the Capital Market Line (CML), which represents combinations of the risk-free asset and the optimal risky portfolio (the tangency portfolio)
The CML is expressed through the Capital Asset Pricing Model (CAPM):

$$R_i = R_f + \beta_i * (R_m - R_f)$$

Where:

- $R_i$ = Expected return of the portfolio
- $R_f$ = Risk-free rate
- $\beta_i$ = Portfolio's beta (sensitivity to market movements)
- $R_m$ = Expected market return

Beta asks a narrow, specific question: *if the market moves by 1%, how much do I expect this asset to move?* A stock with beta 1.0 moves, on average, in step with the market. A stock with beta 1.5 amplifies market moves by 50%; a stock with beta 0.6 dampens them. Beta measures an asset's sensitivity to market-wide movements. It says nothing about the company's own stock-specific news.

CAPM says that investors are only rewarded for the risk they cannot avoid through diversification.

Total risk of a stock has two parts:

- Systematic risk: Risk caused by factors that affect the whole market, such as interest rates, economic growth, or overall market sentiment. You cannot remove this risk just by owning more stocks.
- Idiosyncratic (unsystematic) risk: Risk specific to a company or industry, such as a factory fire, lawsuit, or management change. You can reduce this risk by holding a diversified portfolio.

## 4. Efficient Market Hypothesis:

If markets price systematic risk the way CAPM describes, a deeper question follows: can I actually outperform the market using information that's already public - the same fundamental and technical analysis I already know how to do? The Efficient Market Hypothesis (EMH), formalized by Eugene Fama, addresses this directly.

If past price and volume information is already reflected in prices, what does that imply for technical analysis? Under weak-form efficiency, a pattern I spot on a chart is one everyone else with the same price history could also spot, and if it's real and exploitable, that exploitation itself should erase it.
If publicly available information is already reflected in prices, what does that imply for fundamental analysis? Under semi-strong efficiency, correctly valuing a company from its financial statements isn't enough on its own, the market may already have priced in everything in those statements the moment they were released.

| Form | Information already in price | Implication |
|---|---|---|
| Weak-form | All past price and volume data | Technical analysis based on past-price patterns cannot generate consistent excess returns. |
| Semi-strong-form | All public information (statements, news, macro data) | Fundamental analysis based on public data cannot generate consistent excess returns either. |
| Strong-form | All information, public and private (insider) | Even insider information could not generate consistent excess returns, the most disputed form. |

EMH does not claim that prices are always perfectly correct, that bubbles or volatility cannot occur, or that every investor is rational. It only argues that prices quickly reflect available information, making it difficult to consistently predict price movements or corrections.

The honest view is that EMH is neither completely true nor false. Markets are efficient enough that simple strategies rarely beat them consistently after costs, but not so efficient that all systematic patterns disappear.

## 5. Correlation and Cointegration

Let's come back to the topic of Correlation. You might have heard about Statistical Arbitrage and the famous example of Pepsi and Coca-Cola. Statistical Arbitrage is also known as Pairs Trading.

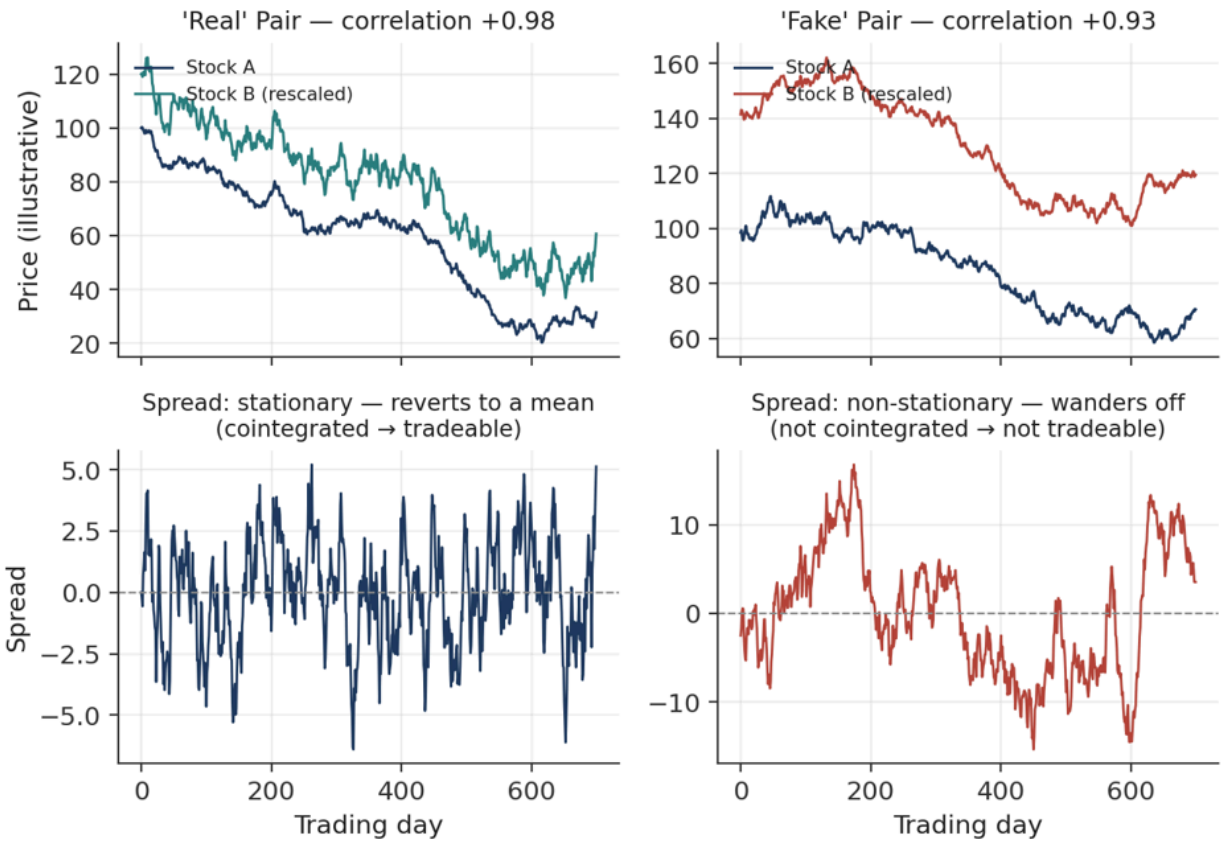
I find two stocks that seem to move together, say PepsiCo and The Coca Cola Company. Does that automatically mean I've found a stable, tradeable, long-term relationship between them? This is one of the most commonly confused ideas in quantitative finance, and getting it wrong is exactly what turns a promising pairs trade into a loss.

**Correlation** measures how closely two investments move together over time. It ranges from $-1$ to $+1$: $-1$ means they move in opposite directions, 0 means there is no clear linear relationship, and $+1$ means they move together. It simply tells us about their co-movement, not why they move together.

**Cointegration** asks a completely different question: is there a stable, long-run relationship between these two price series, such that the gap between them keeps returning to some equilibrium, even while each series individually wanders unpredictably?

Two price series can move together closely and still have a spread that keeps widening over time. High correlation does not mean the gap between them will eventually return to normal.

A spread is the difference between two prices, adjusted by a hedge ratio if needed. Stationarity means the spread stays around a stable average instead of drifting away. This stable level is called the long-run equilibrium. A cointegrating relationship exists when two prices can drift individually, but their combination forms a stable spread that tends to return to its average.




*Correlation and cointegration answer fundamentally different questions. A stable, mean-reverting spread is a genuinely exploitable pattern; a merely correlated pair offers no such guarantee, two unrelated random walks can drift together for years by pure chance before the illusion collapses.*

## 6. Statistical Distribution of Returns:

A return simply tells us how much an investment gained or lost over a certain period, whether that is a day, a month, or a single trade. If you already know how to calculate a return, you already know the basic building block for everything that follows.

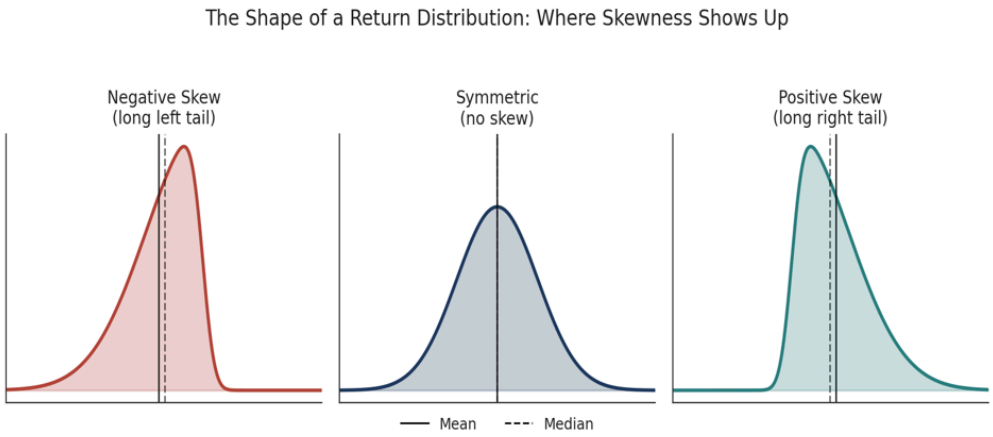
Now, instead of looking at just one return, imagine collecting hundreds or even thousands of them, such as the daily returns of a stock or every return generated by a trading strategy. If we plot these returns based on how frequently they occur, we get a return distribution. In simple terms, it shows us how often different outcomes happen.

Two things are usually the first to look at:

- **Mean return**: The average return. It gives us an idea of the typical or central outcome.
- **Volatility** (standard deviation): How much the returns tend to move around that average. It gives us a sense of how unpredictable the outcomes are.

Once we know the center and spread of a distribution, there are two more questions we can ask about its shape: Which direction do unusual returns tend to lean? This is skewness. How likely are we to see unusually extreme returns? This is kurtosis.

These are called the third and fourth moments of a distribution. The mean is the first moment, and variance is the second. In finance, **skewness and kurtosis** are especially useful because two strategies can have the same average return and volatility but behave very differently during extreme market moves.



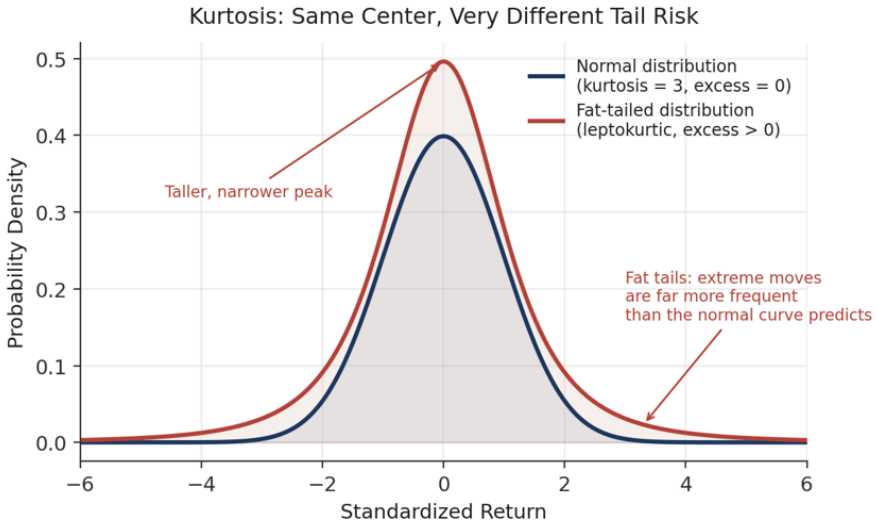
In a symmetric distribution, outcomes are balanced on both sides, so the mean, median, and mode are roughly equal. In a positively skewed distribution, most outcomes are modest losses, but a few large gains create a long right tail, pulling the mean above the median. In a negatively skewed distribution, most outcomes are modest gains, but occasional large losses create a long left tail, pulling the mean below the median.

This distinction matters in practice. Negative skewness is common in option selling and other insurance like strategies, where there are many small gains but occasional large losses. It is often described as “picking up pennies in front of a steamroller.”

Positive skewness works the other way. Strategies like trend following, option buying, and venture investing may have many small losses but a few large gains that make up for them.

Kurtosis tells us how often extreme outcomes occur compared to a normal bell shaped distribution. While volatility shows the typical spread of returns, it does not tell us how likely very large gains or losses are. Kurtosis helps us understand this risk of extreme events.

A distribution can have thin tails (platykurtic, kurtosis<3), where extreme outcomes are less common than in a normal distribution, or fat tails (leptokurtic, kurtosis >3), where extreme outcomes occur more often. Fat tails are especially common in financial markets, where large unexpected gains or losses happen more often than a normal distribution suggests.



Fat tailed returns mean extreme market moves, such as sudden crashes or panic rallies, happen more often than a normal distribution would suggest. This is known as tail risk or crash risk.

Volatility alone may not capture this. Two strategies can have the same volatility but very different levels of crash risk. One may have steady day to day movements, while the other stays calm most of the time but occasionally suffers a huge loss.

| Strategy | Mean | Volatility | Skewness | Kurtosis | Interpretation |
|---|---|---|---|---|---|
| Strategy A | 1.0%/mo | 4.0%/mo | −1.4 | 7.5 | Steady small gains most months; sharp risk of a large loss in a bad month. |
| Strategy B | 1.0%/mo | 4.0%/mo | +0.6 | 4.0 | Closer to a normal shape; occasional strong months without outsized crash risk. |

Strategy A and Strategy B share an identical mean and an identical volatility, which means they also share an identical Sharpe ratio. A quant who evaluates them using only return + volatility would call them interchangeable. Looking at the distribution's shape and tail behaviour reveals they are not: Strategy A is quietly building crash risk, and Strategy B is not.

| Moment number | Name | Measure of | Formula |
|---|---|---|---|
| 1 | Mean | Central tendency | $\bar{X} = \frac{\sum_{i=1}^{N} X_i}{N}$ |
| 2 | Variance (Volatility) | Dispersion | $\sigma^2 = \frac{\sum_{i=1}^{N} (X_i - \bar{X})^2}{N}$ |
| 3 | Skewness | Symmetry (Positive or Negative) | $Skew = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{(X_i - \bar{X})}{\sigma}\right]^3$ |
| 4 | Kurtosis | Shape (Tall or flat) | $Kurt = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{(X_i - \bar{X})}{\sigma}\right]^4$ |

Where X is a random variable having N observations (i = 1,2,…,N).

# 7. Factor Investing and the Fama-French Model

I already know how to do fundamental analysis and technical analysis. Efficient Market Hypothesis raised a real challenge to some of that: markets are efficient enough that these ideas rarely offer effortless outperformance. But can characteristics like "cheap," "small," or "profitable" be measured systematically, across hundreds of stocks at once, rather than one company at a time? That is the idea behind factor investing.

A factor is a measurable characteristic that helps explain differences in return or risk across a large universe of securities, not a story about one company, but a property computed for every stock in the market and used to sort them.

**Three Factor Model:** CAPM says that market risk (beta) explains expected returns. Fama and French found that beta alone could not explain everything: small companies and cheap stocks had historically earned higher returns than CAPM predicted. They added two more factors to explain these differences:

- SMB (Small Minus Big) - The return of small-cap stocks minus large-cap stocks. It measures whether smaller companies earn higher returns than larger companies. Historically, small-cap stocks have often earned a higher return, possibly because they are riskier and less liquid.

- HML (High Minus Low) - The return of value stocks (high book-to-market) minus growth stocks (low book-to-market). It measures whether cheaper stocks earn higher returns than expensive stocks. Historically, value stocks have often outperformed growth stocks, but this advantage has not been consistent in recent years.

**Five Factor Extension: Fama and French** extended the model with two more factors

- RMW (Robust Minus Weak) - The return of more profitable companies minus less profitable companies. It measures whether profitability affects returns. Historically, more profitable companies have tended to earn higher returns than their size and value characteristics alone would suggest.
- CMA (Conservative Minus Aggressive) - The return of companies that invest conservatively minus companies that invest aggressively. It measures whether a company's investment style affects returns. Historically, companies that invest more conservatively have tended to outperform those that expand aggressively.

Mathematically:

$$Ri - Rf = \alpha i + \beta 1(MKT) + \beta 2(SMB) + \beta 3(HML) + \beta 4(RMW) + \beta 5(CMA) + \varepsilon i$$

This is basically the same regression idea as CAPM, but with more factors. Ri − Rf is the stock or portfolio's return above the risk-free rate. Each β shows how strongly the portfolio is exposed to a particular factor, such as value or size. α (alpha) is the return left unexplained after accounting for all the factors. A consistently positive alpha after costs may indicate genuine skill or a missing factor, while an alpha near zero means the returns are largely explained by the factors. ε is the stock-specific noise that the model cannot explain.

*Factor investing turns qualitative investment ideas into measurable, systematic characteristics - the same value and size intuitions from fundamental analysis, now applied across an entire universe of stocks at once.*

## 8. Performance Measurement:

**Total return**: is the overall percentage gain or loss over the entire period. It tells you the final result, but not how you got there or how much risk you took. A 44% return could come from a smooth four year climb or a highly volatile journey. The number alone does not tell us the difference.

**Annualized Return**: Since investments and strategies can run for different lengths of time, total returns are not always easy to compare. Annualized return converts the overall return into an equivalent yearly rate, making it easier to compare strategies over different time periods.

**CAGR** - Compound Annual growth rate

$$\text{CAGR (\%)} = \left(\frac{\text{Ending Value}}{\text{Beginning Value}}\right)^{(1 \div t)} - 1$$

Here, t is the number of years. CAGR answers a simple question: what constant yearly growth rate would have turned the starting value into the ending value, assuming the return compounds each year? It turns a potentially bumpy investment journey into one simple number that is easy to compare across investments.

**Volatility**: Volatility, measured by the standard deviation of returns, shows how much returns typically move around their average. It is widely used as a proxy for risk because it is simple and forms the basis of measures like the Sharpe ratio and Markowitz portfolio theory. Daily volatility is commonly annualized using √252. However, volatility treats gains and losses equally, even though investors usually care much more about downside losses. So, volatility measures variation, but it does not fully capture risk.

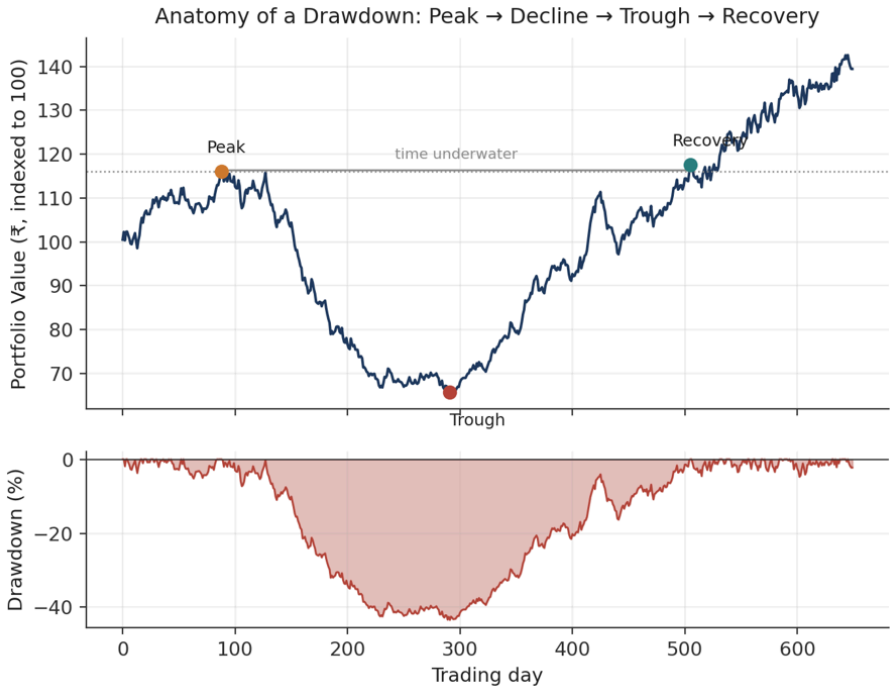
**Sharpe ratio**: Already discussed above - "How much excess return am I generating for the total volatility I am taking?"

**Sortino Ratio:** "How much return am I generating relative to the downside risk I am taking?"

$$S = \frac{R_p - R_f}{\sigma_d}$$

S – Sortino ratio
$R_p$ – Portfolio return
$R_f$ – Risk-free rate
$\sigma_d$ – Standard deviation of negative asset returns (downside deviation)

**Drawdown**: measures the decline in portfolio value from its highest point (peak) before it recovers. Visually, every drawdown episode follows the same anatomy:



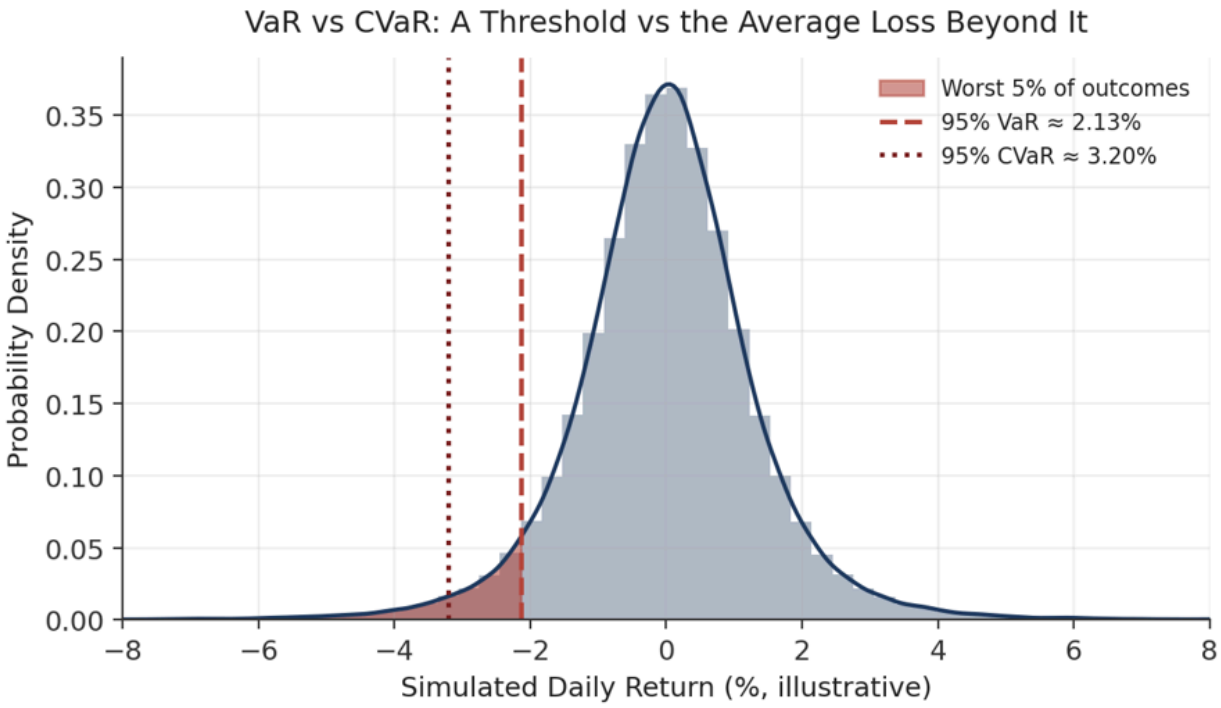
**Maximum Drawdown**: the single largest peak-to-trough decline over the period; the worst-case scenario realized.

$$\text{Maximum Drawdown (\%)} = \frac{\text{Peak Value} - \text{Trough Value}}{\text{Peak Value}}$$

**Calmar Ratio**: "How much annualized return am I generating relative to the worst drawdown I experienced?"

$$Calmar\ Ratio = \frac{CAGR}{Max\ DD}$$

**Value at Risk (VaR)**: asks how much you could lose over a given period at a certain confidence level. For example, a 1 day 95% VaR of 2% means there is a 5% chance of losing more than 2% in one day. Expected Shortfall (ES) goes further by asking how large the loss is likely to be when VaR is exceeded. It measures the average loss in the worst cases, making it more focused on tail risk.



## 9. Conclusion

Every idea in this module showed up because the previous one raised a question it couldn't answer on its own.

I started by picking good stocks, but soon realized that owning good stocks does not automatically mean having a good portfolio. That led me to portfolio management. As I built portfolios, I learned that risk is not just about how risky each stock is on its own, but also about how the stocks move together. That took me to covariance, correlation, and the efficient frontier. I then realized that average return and volatility do not tell the whole story. Skewness and kurtosis help us understand the shape of returns and the possibility of extreme outcomes. From there, I wanted to know what return I should expect for taking systematic risk, which led me to CAPM. I also wondered whether I could consistently beat the market using information that everyone already had, which brought me to EMH. The reality seemed to lie somewhere between “markets are unbeatable” and “anything goes.” When I started looking for pairs of stocks that move together, I learned that correlation and cointegration answer different questions, and cointegration is what can point to a relationship stable enough to trade. Finally, I wanted to understand whether ideas like value and size could be measured systematically across the market, which led me to factor investing and Fama-French. In many ways, Fama-French simply takes the core idea behind CAPM and extends it to multiple sources of risk. I also understand how to evaluate my strategy using performance measurement.

Other resources which you can explore:

- EDHEC Business School - Introduction to Portfolio Construction with python.
- Portfolio Management (Lecture 16) - MIT OCW