

ORIGINAL RESEARCH

# Portfolio Optimization and Risk Management

Diyaa Khurana<sup>1\*</sup>, Pavel Sokolov<sup>2</sup>

<sup>1</sup>*Carlmont High School, 1400 Alameda de las Pulgas, Belmont, CA 94002*

<sup>2</sup>*University of Southern California, 3620 S. Vermont Ave., Los Angeles, CA 90089-2532*

*\*Corresponding author email: [ms.diyaakhurana@gmail.com](mailto:ms.diyaakhurana@gmail.com)*

Accepted: October 21, 2026

## Abstract

This paper focuses on how mathematics supports investment analysis, specifically on portfolio optimization and risk management under Modern Portfolio Theory. We begin with compound interest, then develop the probabilistic foundations of quantitative finance, including expected value, variance, covariance, the Law of Large Numbers, the Central Limit Theorem, and Bayes' Theorem. Building on these, we study diversification, the Sharpe ratio, Markowitz optimization, the efficient frontier, and the Capital Allocation Line. Risk is assessed through Value at Risk and the distinction between systematic and unsystematic risk, illustrated by major historical market crashes. Following the Capital Asset Pricing Model (CAPM), we relate an asset's expected return to the expected market return and beta. We also review common technical indicators, including moving averages, the relative strength index, and the VIX fear index. Finally, we apply linear regression and random forest models to stock return prediction and algorithmic trading, and discuss their advantages and limitations.

**Keywords:** expectation; covariance; volatility; CAPM; Modern Portfolio Theory; return rate; prediction; Bayes' Theorem; random forest; linear regression

---

## 1. Introduction

Mathematics has an essential role in modern finance by allowing people to analyze investments, measure risk, and make informed financial decisions. Mathematical concepts allow investors and financial professionals to make decisions based on data rather than speculation, like calculating future value of investments through compound interest, and constructing diversified portfolios and evaluating market risk. This paper explains the essential mathematical foundations used in finance, including compound interest, probability, expected value, variance, standard deviation, covariance, portfolio optimization, the Capital Asset Pricing Model, Value at Risk, and the Efficient Frontier. It also studies historical market crashes to show how these concepts apply during periods of financial uncertainty and concludes with machine learning techniques used to predict stock returns. By combining probability, statistics, calculus, linear algebra, and optimization methods with real-world financial applications, this paper provides an overview of how mathematics supports investment analysis, risk management, and portfolio construction.

## 2. Compound Interest

Compound interest is the process of earning interest on both an initial principal and the accumulated interest from previous periods. The future value  $A$  of an investment experiencing compound interest is given by:

$$A = P \left( 1 + \frac{r}{n} \right)^{nt}$$

where:

- $P$ : Principal amount (initial investment)
- $r$ : Annual interest rate (as a decimal)
- $n$ : Number of times interest is compounded per year
- $t$ : Number of years the money is invested

**Example 1:** An investor deposits \$5,000 into a Certificate of Deposit (CD) that yields 4.5% annual interest, compounded annually. What is the total balance after 10 years?

$$P = 5000, \quad r = 0.045, \quad n = 1, \quad t = 10$$

$$A = 5000 \left( 1 + \frac{0.045}{1} \right)^{1 \cdot 10} = 5000(1.045)^{10} \approx 5000 \times 1.552969 = 7764.85$$

Thus,

$$A \approx \$7,764.85$$

### 2.1 Applications of Compounding

Compounding is especially effective over long time horizons, underlying several core financial strategies:

- **Retirement accounts** (401(k)s, IRAs)—multi-decade compounding drives exponential growth of invested capital.
- **Dividend reinvestment**—reinvested earnings increase share count, compounding portfolio value over 15–30 years.
- **Education savings**—tax-advantaged accounts started early let returns compound well before funds are needed.

The next section introduces the probability and statistics foundation, expected value, variance, and standard deviation, needed to quantify investment risk formally.

## 3. Probability and Statistics Prerequisites

Building on the concepts introduced above, this section turns to the probabilistic tools needed to measure risk and uncertainty in financial outcomes. Here we introduce and study the fundamental statistical concepts of expected value, variance, and standard deviation. These are basic ideas frequently used in finance.

**Definition:** Let  $X$  be a random variable, and let  $p_i = P(X = a_i)$  for the  $n$  possible values of  $X$ :  $a_1, a_2, \dots, a_n$ . The expected value of  $X$  is

$$E[X] = \sum_{i=1}^n a_i p_i = a_1 p_1 + a_2 p_2 + \dots + a_n p_n.$$

**Note:** Mathematicians commonly use the notation  $\mu = E[X]$ .

**Definition:** With the same setup as in the previous definition, the variance of  $X$  is defined as

$$\text{Var}(X) = E[X^2] - (E[X])^2,$$

or alternatively

$$\text{Var}(X) = E[(X - E[X])^2] = \sum_{i=1}^n (a_i - \mu)^2 p_i.$$

**Definition:** The standard deviation of a random variable  $X$  is defined as

$$\sigma = \sqrt{\text{Var}(X)}.$$

**Example:** Suppose we roll a fair die. Let  $X$  be the random event corresponding to the number rolled. Compute its expected value and standard deviation.

Since it is a fair die, we have  $p_1 = p_2 = \dots = p_6 = \frac{1}{6}$ . Then the expected value of  $X$  is

$$E[X] = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = \frac{7}{2}.$$

Similarly,

$$E[X^2] = \frac{1}{6}(1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2) = \frac{91}{6}.$$

Hence,

$$\text{Var}(X) = \frac{91}{6} - \frac{49}{4} = \frac{182 - 147}{12} = \frac{35}{12}.$$

Therefore, the standard deviation is

$$\sigma = \sqrt{\frac{35}{12}} \approx 1.71.$$

The following section demonstrates how these concepts apply to finance.

#### 4. Expected Returns, Variance, and Standard Deviation

With expected value, variance, and standard deviation now defined in general terms, this section applies these concepts directly to investment. Suppose an investment has possible returns  $R_1, R_2, \dots, R_n$  with probabilities  $p_1, p_2, \dots, p_n$  respectively. The expected return is

$$E[R] = \sum_{i=1}^n p_i R_i.$$

In particular, if two assets have portfolio weights  $w_1, w_2$ . Then  $w_1 + w_2 = 1$ . The expected return of such portfolio is

$$E[R_p] = w_1 E[R_1] + w_2 E[R_2].$$

**Example 1:** Suppose a stock has the following possible annual returns.

**Table 1.** Possible annual returns and their probabilities.

| Return | Probability |
|--------|-------------|
| 5%     | 0.30        |
| 10%    | 0.50        |
| 20%    | 0.20        |

Then

$$E[R] = 0.30(5) + 0.50(10) + 0.20(20) = 10.5\%.$$

Thus, the expected annual return equals 10.5%.

Risk is commonly measured using variance. The variance of returns is

$$\text{Var}(R) = \sum_{i=1}^n p_i (R_i - E[R])^2.$$

This also can be computed using the formula

$$\text{Var}(R) = E[R^2] - (E[R])^2.$$

The standard deviation is

$$\sigma = \sqrt{\text{Var}(R)}.$$

In the finance context, this is called volatility.

**Example 2:** Using the previous example,  $E[R] = 10.5$ . The variance equals

$$0.3(5 - 10.5)^2 + 0.5(10 - 10.5)^2 + 0.2(20 - 10.5)^2.$$

This becomes

$$0.3(30.25) + 0.5(0.25) + 0.2(90.25) = 27.25.$$

Therefore,

$$\sigma = \sqrt{27.25} \approx 5.22\%.$$

This means that the stock's return typically deviated from its 10.5% expected return by about 5.22 percentage points; this is the stock's volatility. The quantities discussed above describe the average behavior of portfolios. In practice, investors need to answer questions such as:

- How likely is a portfolio to lose more than 5%?
- How accurately can we estimate expected returns?
- Why do stock returns often appear approximately normally distributed?

Two fundamental results from probability theory that help to answer these questions are **The Law of Large Numbers (LLN)** and **The Central Limit Theorem (CLT)**.

## 5. Law of Large Numbers and Central Limit Theorem

The next two theorems make precise why sample averages of returns can be trusted as estimates of the true expected return, and how quickly that trust is earned as more data accumulates.

The Law of Large Numbers explains why averages become stable when the number of observations increases.

Suppose  $X_1, X_2, \dots, X_n$  are independent random variables with the same expected value  $E[X_i] = \mu$ . The sample average is

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

The Law of Large Numbers states that

$$\lim_{n \rightarrow \infty} \bar{X}_n = \mu,$$

i.e. the average of many observations becomes close to the true expected value.

The Law of Large Numbers explains why:

- historical averages become better estimates with more data,
- large insurance companies can predict losses,
- diversified portfolios reduce unsystematic risk.

**Example:** Suppose daily returns of a stock have expected return  $\mu = 0.05\%$ . After observing thousands of trading days,  $\bar{R}_n = \frac{R_1 + \dots + R_n}{n}$  should become close to 0.05%.

However, LLN does not tell us how quickly this convergence occurs. For that we use the Central Limit Theorem.

The Central Limit Theorem describes the distribution of sums of random variables. Suppose  $X_1, X_2, \dots, X_n$  are independent and identically distributed random variables with  $E[X_i] = \mu$  and  $\text{Var}(X_i) = \sigma^2$ . Then

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \rightarrow N(0, 1).$$

Equivalently,

$$\bar{X}_n \approx N\left(\mu, \frac{\sigma^2}{n}\right).$$

Here  $N(0, 1)$  denotes that the variable is normally distributed with mean 0 and variance 1, and similarly for  $N(\mu, \frac{\sigma^2}{n})$ .

The remarkable feature of the CLT is that the original distribution does not need to be normal.

**Table 2.** Comparison of the Law of Large Numbers and the Central Limit Theorem.

|                 | LLN                         | CLT                                      |
|-----------------|-----------------------------|------------------------------------------|
| Main question   | Does the average converge?  | What is the distribution of the average? |
| Output          | $\bar{X}_n \rightarrow \mu$ | $\bar{X}_n \approx N(\mu, \sigma^2/n)$   |
| Application     | Estimation                  | Risk measurement                         |
| Finance example | Long-run average return     | Confidence intervals and VaR             |

Both results describe large samples. The difference is:

- LLN gives convergence.
- CLT gives the speed and distribution of fluctuations around the mean.

**Definition:** A confidence interval is an interval  $(a, b)$  such that, at the given confidence level  $\alpha$ , we are certain that the true population mean lies between  $a$  and  $b$ . For example, A 95% confidence interval is a range built from sample data. If we repeated the sampling many times, about 95% of such ranges would contain the true average.

How to calculate  $a$  and  $b$ ? The confidence interval is always of the form

$$(\bar{x} - z_{\frac{\alpha}{2}} SE, \bar{x} + z_{\frac{\alpha}{2}} SE),$$

where  $SE$  is the standard error given by  $\frac{\sigma}{\sqrt{n}}$ , where  $n$  is the data set size.

**Example 1:** Suppose a portfolio has daily returns  $R_1, \dots, R_{252}$  over one trading year. Assume

$$E[R_i] = 0.05\% \text{ and } \sigma = 1\%.$$

The annual average daily return is

$$\bar{R} = \frac{1}{252} \sum_{i=1}^{252} R_i.$$

By the CLT,

$$\bar{R} \approx N\left(0.05\%, \frac{0.01^2}{252}\right).$$

The standard error is  $\frac{1\%}{\sqrt{252}} = 0.063\%$ .

Therefore, approximately 95% of possible sample averages satisfy  $0.05\% \pm 1.96(0.063\%)$ . Hence,

$$-0.073\% < \bar{R} < 0.173\%.$$

The CLT allows investors to estimate uncertainty in average returns.

The next section discusses further applications of probability theory to finance.

## 6. Bayes' Theorem

Another probability topic to be used is Bayes' Theorem. It helps to simplify conditional probability problems. One example of its application is to answer the question: What is the probability that the stock will go up tomorrow given the market news today.

Formula: For any events  $A$  and  $B$   $P(A)$  denotes the probability of event  $A$  and  $P(A|B)$  denotes the probability of event  $A$  if we know that event  $B$  happened.

Before explaining the formula for Bayes' theorem, first, let's study the general probability. The probability of event  $A$  is the number of favorable outcomes (those belonging to the set  $A$ ) over the number of possible outcomes:

$$P(A) = \frac{\#\{\text{favorable outcomes}\}}{\#\{\text{possible outcomes}\}}.$$

This formula also applies to conditional probability of event  $A$  given that event  $B$  occurred. However, the sizes of both sets  $\{\text{favorable outcomes}\}$  and  $\{\text{possible outcomes}\}$  may be affected.

**Example:** Roll a fair die twice. What is the probability that the sum of two rolls is 11? What is the probability that the sum of the rolls is 11 if we saw that the first roll was a 4? Are the probabilities the same? The first probability can be denoted as  $P(S = 11)$ , while the second one can be denoted as  $P(S = 11|X = 4)$ , where  $X$  is the variable representing the outcome of the first roll,  $Y$  is the variable representing the outcome of the second roll, and  $S = X + Y$  is the variable representing the sum of two rolls.

In the first case, we have

$$P(S = 11) = \frac{\#\{\text{favorable outcomes}\}}{\#\{\text{possible outcomes}\}} = \frac{\#\{(5,6), (6,5)\}}{36} = 1/18.$$

In the second case, we have

$$P(S = 11|X = 4) = \frac{\#\{\text{favorable outcomes}\}}{\#\{\text{possible outcomes}\}} = \frac{0}{6} = 0.$$

The total space size is reduced to 6 since the value of  $X$  is known and there are only 6 possible values of  $Y$ . There are no favorable outcomes since it is impossible to have  $Y = 7$ , which is required in order to have  $S = 4 + Y = 11$ .

**Bayes' Theorem:**

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

if  $P(B) > 0$ . Use it when you know  $P(B|A)$  and want  $P(A|B)$ .

**Example:** Suppose that on any given day there is a 60% chance that stock A goes up. When the stock goes up, positive news appears 70% of the time. If the stock goes down, positive news appears 40% of the time. Suppose there is positive news. Calculate the probability that the stock will go up.

Stock goes up:  $P(\text{up}) = 0.60$ . Stock goes down:  $P(\text{down}) = 0.40$ . Positive news given the stock goes up:  $P(\text{News}|\text{Up}) = 0.70$ . Positive news given the stock goes down:  $P(\text{News}|\text{Down}) = 0.40$ .

By the **Law of Total Probability**,

$$\begin{aligned} P(\text{News}) &= P(\text{News}|\text{Up}) \cdot P(\text{Up}) + P(\text{News}|\text{Down}) \cdot P(\text{Down}) \\ &= 0.70 \cdot 0.60 + 0.40 \cdot 0.40 = 0.42 + 0.16 = 0.58 \end{aligned}$$

$$P(\text{Up}|\text{News}) = \frac{P(\text{News}|\text{Up}) \cdot P(\text{Up})}{P(\text{News})} = (0.70 \cdot 0.60) / 0.58 = 0.42 / 0.58 = 0.7241 = 72.41\%$$

Therefore, given the positive news, we have a 72.41% chance that stock A will go up.

The following few sections discuss more probabilistic tools used in finance with examples.

## 7. Covariance and Correlation

When studying multiple assets we must understand how they move together. This section extends the discussion from single random variables to pairs of assets, introducing the statistical tools needed to describe how two investments move together.

The covariance between assets  $X$  and  $Y$  is

$$\text{Cov}(X, Y) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}).$$

Positive covariance means the assets tend to move together. Negative covariance means they tend to move in opposite directions.

**Example:** Suppose  $X = \{4, 6, 8\}$  and  $Y = \{3, 7, 8\}$ . The sample means are  $\bar{X} = 6$ , and  $\bar{Y} = 6$ . The covariance is

$$\frac{(-2)(-3) + 0(1) + 2(2)}{3} = \frac{10}{3} = 3.33.$$

Moreover, for several assets we organize all covariances into a covariance matrix. Before proceeding to the example below, one needs to understand the following definition.

**Definition:** A positive semi-definite matrix  $A$  is an  $n$ -by- $n$  matrix such that  $\vec{x}^T A \vec{x} \geq 0$  for all vectors  $\vec{x} \in \mathbb{R}^n$ .

**Example:** Consider 3 assets for this example  $X_1, X_2$ , and  $X_3$ , so we need to construct a 3-by-3 matrix

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 \end{bmatrix},$$

where each  $\sigma_{ij} = \text{Cov}(X_i, X_j)$ . Note that for any  $i = 1, 2, 3$ ,  $\text{Cov}(X_i, X_i) = \text{Var}(X_i)$ , the variance of  $X_i$ .

Important properties:

- $\Sigma$  is symmetric.
- Diagonal entries are variances.
- Off-diagonal entries are covariances.
- $\Sigma$  is positive semi-definite.

Correlation standardizes covariance. Its formula is

$$\rho = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}.$$

The correlation always satisfies  $-1 \leq \rho \leq 1$ .

Interpretation:

- $\rho = 1$ : perfect positive relationship.
- $\rho = 0$ : no linear relationship.
- $\rho = -1$ : perfect negative relationship.

These measures of covariance and correlation are the essential building blocks for quantifying the risk of a portfolio containing more than one asset, which is the focus of the next section.

## 8. Portfolio Variance

Building directly on the covariance and correlation formulas above, this section shows how those pairwise relationships combine to determine the total risk of a portfolio. Portfolio risk is

$$\sigma_p^2 = w_1^2 \sigma_1^2 + w_2^2 \sigma_2^2 + 2w_1 w_2 \text{Cov}(R_1, R_2).$$

Using correlation,

$$\sigma_p^2 = w_1^2 \sigma_1^2 + w_2^2 \sigma_2^2 + 2w_1 w_2 \rho \sigma_1 \sigma_2.$$

**Example:** Suppose  $w_1 = w_2 = 0.5$ ,

$$\sigma_1 = 15\%, \quad \sigma_2 = 20\%, \quad \text{and} \quad \rho = 0.30.$$

Then  $\sigma_p^2 = 0.25(0.15)^2 + 0.25(0.20)^2 + 2(0.5)(0.5)(0.30)(0.15)(0.20) = 0.015625$ . Therefore,

$$\sigma_p = 12.5\%.$$

Notice that the portfolio standard deviation is smaller than the weighted average volatility because diversification reduces risk. Diversification reduces portfolio risk by combining assets that are not perfectly correlated.

If  $\rho < 1$ , then the risk is reduced.

If  $\rho = -1$ , it is theoretically possible to eliminate all risk with appropriate portfolio weights.

While diversification can reduce unsystematic risk, real financial markets can also have shocks that diversification alone cannot fully protect against. The next section examines several historical market crashes and the distinction between systematic and unsystematic risk that they expose.

## 9. Market Crashes and Types of Risk

History offers several examples of how quickly seemingly diversified portfolios can lose value when systemic shocks strike the entire market. On October 19, 1987 (“Black Monday”), the Dow Jones fell approximately 22.6% in a single day. [5] Many risk models at the time significantly underestimated the probability of such a large decline. [3] Later crashes followed, including the 2008 Financial Crisis, triggered by Lehman Brothers’ collapse in September 2008, and the March 2020 COVID crash, in which the S&P 500 fell over 30% from its peak and volatility reached historically high levels. [6] They reinforced the same set of lessons: correlations across assets can spike sharply during a crisis, undermining diversification exactly when it’s needed most; liquidity can disappear rapidly; and historical volatility often understates future risk, forcing models like VaR to be recalibrated on the fly. [11]

Portfolio risk consists of two components, systematic risk and unsystematic risk.

**Definition:** Systematic risk affects the entire market. Examples include recessions, wars, inflation, interest-rate changes. This risk cannot be diversified away.

**Definition:** Unsystematic risk is the risk associated with any particular company. Examples include management scandals, product recalls, lawsuits, accounting fraud. This risk can largely be reduced through diversification.

To work with risk we can use technical indicators, including price of the **fear index** VIX, relative strength index RSI and the Sharpe ratio.

## 10. Technical Indicators

This section develops three of widely used technical indicators, the simple moving average, the relative strength index, and the VIX, each offering a different perspective on price trend, momentum, and market sentiment.

The simple moving average (SMA) is calculated as follows.

$$SMA = \frac{\text{sum of closing prices in selected time period}}{\text{selected time period length}}.$$

**Example:** To calculate a 50-day SMA, add the 50 closing prices of the respective asset, and then divide the result by 50.

When the price of an asset is above its SMA, it is considered to be in an up trend. When the price of an asset is below its SMA, it is considered to be in a down trend. In short, if price  $>$  SMA, then buy, and if price  $<$  SMA, then sell.

How do we decide whether an asset is overbought or oversold? Answer: Use the relative strength index. Relative strength index (RSI) is calculated as follows.

$$RSI = 100 - \frac{100}{1 + RS}, \quad \text{where } RS = \frac{\text{average gains}}{\text{average losses}}.$$

Usually we use a 14-day period when calculating RSI. RSI values below 30 indicate the oversold values, while RSI values above 70 indicate the overbought values.

**Definition:** Divergence occurs when the stock price and RSI move in opposite directions.

**Definition:** Bullish Divergence occurs when the stock price is falling and RSI is recovering/increasing.

**Definition:** Bearish Divergence occurs when the stock price is increasing and RSI is decreasing after a rise.

**Example:** Suppose the stock price is \$50 and RSI is 28. Then the stock price becomes \$45 and the RSI is now 35.

This is an example of a bullish divergence. Another indicator used in trading is the **fear index**, the VIX. It is designed to measure the market's expectation of 30-day forward-looking volatility of the U.S. equity market, as conveyed by S&P 500 Index option prices. [7] It is calculated from those option prices as the square root of variance estimate, multiplied by 100. [7]

An interesting observation is to see the VIX values during crises and compare them to smoother economic environments. Values of VIX of 15 and below are considered good (low expected volatility). Values of VIX of 30 and above indicate risky market conditions (high expected volatility).

Having covered indicators that describe an asset's own price behavior, the next section turns to a model that relates an asset's expected return to its risk relative to the broader market, the Capital Asset Pricing Model.

## 11. Capital Asset Pricing Model (CAPM)

The Capital Asset Pricing Model (CAPM) relates expected return to systematic risk. [9], [10] The CAPM equation is

$$E[R_i] = R_f + \beta_i(E[R_m] - R_f),$$

where

- $R_f$  is the risk-free rate,
- $E[R_m]$  is the expected market return,
- $\beta_i$  measures systematic risk.

Beta measures how sensitive an asset is to market movements.

$$\beta = \frac{\text{Cov}(R_i, R_m)}{\text{Var}(R_m)}.$$

Interpretation:

- $\beta = 1$  moves with the market.
- $\beta > 1$  is more volatile than the market.
- $\beta < 1$  is less volatile.
- $\beta < 0$  tends to move opposite the market.

**Example:** Suppose  $R_f = 2\%$ ,  $E[R_m] = 10\%$ , and  $\beta = 1.5$ . Then

$$E[R_i] = 2 + 1.5(10 - 2) = 14\%.$$

## 12. Value at Risk

Risk management is one of the central topics in finance. Investors, banks, hedge funds, and asset managers seek to quantify the amount of money they could lose under adverse market conditions.

One of the most widely used risk measures is **Value at Risk (VaR)**. A more detailed description of value-at-risk can be found in [3]. VaR estimates the maximum expected loss over a specified time horizon at a given confidence level. [11]

For example, if a portfolio has a one-day 95% VaR of \$10,000, then there is a 95% probability that the portfolio will not lose more than \$10,000 in a single day.

**Definition:** Let  $L$  denote the portfolio loss over a specified period. The  $\alpha$ -level Value at Risk is defined as

$$\text{VaR}_\alpha = \inf\{x : P(L \leq x) \geq \alpha\}.$$

**Note:** The infimum (inf) of a set denotes the smallest value in the set depending on the conditions/criteria required to belong to the set. It also can denote the hypothetical minimum if the set is infinite. For example, if  $S = \{2, 3, 4, 6, 19\}$ , then  $\inf(S) = 2$ . In practice, the confidence levels commonly used are:

- 95%
- 99%
- 99.9%

For normally distributed returns,

$$\text{VaR} = z_\alpha \sigma V,$$

where

- $V$  = portfolio value,
- $\sigma$  = portfolio volatility,
- $z_\alpha$  = standard normal quantile. Common quantiles are

$$z_{0.95} = 1.645, \quad z_{0.99} = 2.326.$$

**Note:** The formula above assumes the mean is 0, i.e.  $\mu = 0$ . If not, then we can simply use the substitutions  $Y = X - \mu$ , where  $X$  is a random variable with mean  $\mu$ .

**Definition:** Given the data set  $x_1, x_2, \dots, x_n$ , the  $z$ -score of a point  $x_i$  is defined as  $z = \frac{x_i - \mu}{\sigma}$ . For example, the  $z$ -score of  $x_1$  is  $z = \frac{x_1 - \mu}{\sigma}$ .

There are also critical values. Not to be confused with the  $z$ -scores but they are related. In this context,  $z_{0.95}$  represents the value such that 95% of the data has  $z$ -scores below it.

Note that VaR can be estimated using Monte Carlo simulations.

**Example (Ten-Day VaR):** Assume  $V = \$2,000,000$  and  $\sigma = 1.5\%$ . The one-day VaR at 95% confidence is

$$1.645(0.015)(2,000,000) = 49,350.$$

Using the square-root-of-time rule [11]:

$$\text{VaR}_{10} = \text{VaR}_1 \sqrt{10}.$$

Hence,

$$\text{VaR}_{10} = 49,350 \sqrt{10} \approx 156,058.$$

Therefore, the ten-day 95% VaR is approximately \$156,058.

VaR has several limitations: It does not describe losses beyond the VaR threshold. Financial returns are often non-normal. Correlations may increase dramatically during crises. Extreme events occur more frequently than predicted by the normal distribution. [11] These limitations became apparent during major market crashes.

### 13. Sharpe Ratio

The Sharpe ratio is widely used in modern portfolio theory (MPT). Economist William F. Sharpe introduced this measure in 1966 to evaluate mutual funds. [8]

The Sharpe Ratio measures return earned per unit of risk. [4] Its formula is

$$S = \frac{E[R_p] - R_f}{\sigma_p},$$

where  $R_f$  is the risk-free rate, and  $\sigma_p$  is the portfolio volatility. Larger Sharpe ratios indicate better risk-adjusted performance.

**Example:** Suppose  $E[R_p] = 14\%$ ,  $R_f = 3\%$ , and  $\sigma_p = 20\%$ . Then the Sharpe Ratio is

$$S = \frac{E[R_p - R_f]}{\sigma_p} = \frac{14 - 3}{20} = 0.55.$$

How do investors construct portfolios to balance expected return and risk? The following section explains some powerful methods used to resolve this question.

## 14. Modern Portfolio Theory

The central idea of Modern Portfolio Theory, developed by Harry Markowitz in 1952, is that investors should not evaluate investments individually but rather as parts of a portfolio. [1] Diversification allows investors to reduce risk without necessarily sacrificing expected return. [1]

### 14.1 Portfolio Optimization

The Markowitz optimization problem is to solve for [1]

$$\min_{\vec{w}} \vec{w}^T \Sigma \vec{w}$$

subject to

$$\sum_{i=1}^n w_i = 1 \quad \text{and} \quad \vec{w}^T \vec{\mu} = \mu_p.$$

The first constraint requires investing all available capital. The second fixes the desired expected return. Then we solve the system of differential equations

$$\frac{\partial f}{\partial w_i} = \lambda \frac{\partial g}{\partial w_i} + \mu \frac{\partial h}{\partial w_i} \quad \text{for each } 1 \leq i \leq n,$$

where  $g(w_1, w_2, \dots, w_n) = \sum_{i=1}^n w_i$  and  $h(w_1, w_2, \dots, w_n) = \sum_{i=1}^n \mu_i w_i$ . This optimization problem can be solved using the method of **Lagrange Multipliers**.

**Example:** For a two-asset portfolio, the risk function is

$$f(w_1, w_2) = w_1^2 \sigma_1^2 + w_2^2 \sigma_2^2 + 2w_1 w_2 \rho \sigma_1 \sigma_2$$

subject to the constraint  $w_1 + w_2 = 1$ . We name the constraint function, for example,  $g(w_1, w_2) = w_1 + w_2$ . Then we solve the system of differential equations

$$\frac{\partial f}{\partial w_i} = \lambda \frac{\partial g}{\partial w_i} \quad \text{for each } 1 \leq i \leq 2$$

This optimization problem can be solved using the method of **Lagrange Multipliers**. The Lagrangian is

$$L = \vec{w}^T \Sigma \vec{w} - \lambda \left( \begin{bmatrix} 1 \\ 1 \end{bmatrix}^T \vec{w} - 1 \right).$$

Differentiating (and setting = 0) gives

$$2\Sigma \vec{w} - \lambda \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 0.$$

You can think of it as differentiation in the one-variable case with  $\vec{w}^T \Sigma \vec{w}$  behaving like  $w^2$ :

$$\frac{\partial}{\partial \vec{w}} (\vec{w}^T \Sigma \vec{w}) = 2\Sigma \vec{w}.$$

Together with the constraint this forms a system of linear equations.

### 14.2 Efficient Frontier

Not every portfolio is optimal. Among all portfolios with identical risk, some produce larger returns. The Efficient Frontier consists of portfolios satisfying

- maximum return for a fixed risk,
- minimum risk for a fixed return.

Every rational investor should choose a portfolio on this frontier.

**Example 1:** Suppose three portfolios have

**Table 3.** Return and risk of three example portfolios.

| Portfolio | Return | Risk |
|-----------|--------|------|
| A         | 8%     | 10%  |
| B         | 10%    | 10%  |
| C         | 11%    | 15%  |

Portfolio A is dominated by Portfolio B because both have identical risk while B has a larger expected return. Portfolio A therefore cannot lie on the Efficient Frontier.

If no target return is specified, we minimize  $\vec{w}^T \Sigma \vec{w}$  subject to  $\sum_{i=1}^n w_i = 1$ . The resulting portfolio is called the Minimum Variance Portfolio. It has the smallest possible volatility among all fully invested portfolios.

**Example 2:** Suppose two independent assets have

$$\sigma_1 = 10\%, \quad \sigma_2 = 20\%.$$

Since the assets are independent, their covariance is zero. Hence, the portfolio variance is

$$0.01w^2 + 0.04(1-w)^2.$$

Differentiate:

$$0.02w - 0.08(1-w) = 0.$$

Hence,  $0.10w = 0.08$ , and so  $w = 0.80$ . Therefore, our decision is to invest 80% in Asset A, and invest 20% in Asset B.

So far, the discussion has assumed investors can only allocate wealth among risky assets, the next section relaxes this assumption by introducing a risk-free asset and shows how doing so extends the efficient frontier into the Capital Allocation Line.

## 15. The Capital Allocation Line

The Capital Allocation Line is a graph that shows a mix of a safe, risk-free asset and a risky asset or portfolio. The y-axis shows expected return, and the x-axis shows risk measured by standard deviation.

The portfolio return becomes

$$R_p = wR_M + (1-w)R_f.$$

Expected return satisfies

$$E[R_p] = R_f + w(E[R_M] - R_f).$$

The graph of this equation is called the Capital Allocation Line (CAL). Its slope equals the Sharpe Ratio.

**Example 1:** Suppose  $R_f = 3\%$ , and the market portfolio has  $E[R_M] = 11\%$ , with volatility 16%. The Sharpe Ratio is  $\frac{11-3}{16} = 0.50$ . Thus, every additional 1% of volatility increases expected return by approximately 0.5%.

Having established how combining a risk-free asset with the market portfolio expands investors' opportunity set along the Capital Allocation Line, the next section moves from these closed-form analytical models to data-driven methods, exploring how machine learning techniques can be used to predict returns directly from historical market information.

## 16. Machine Learning for Return Prediction

Machine learning models attempt to predict future returns using historical market information. Typical predictor variables include:

- Trading volume
- Volatility
- Momentum indicators
- Interest rates
- Macroeconomic variables
- Sentiment indicators

Common machine learning methods include:

- Linear Regression
- Ridge Regression
- LASSO
- Random Forest
- Neural Networks

In this manuscript, we focus primarily on the linear regression and random forest techniques for the stock return prediction.

## 17. Linear Regression Return Prediction

A linear regression predictive model forecasts a numerical output  $y$  based on input features  $x$  by fitting a straight line that minimizes overall prediction error. This is the best line for the given data. It is widely used for trends like predicting house prices. In this project, we discuss this model and its potential to predict stock prices. We demonstrate its predictive power and evaluate its performance with a small example below.

**Example:** Suppose the actual monthly returns are

$$y = \{5, 7, 9, 10\}$$

(percent). A machine learning model predicts

$$\hat{y} = \{4, 8, 8, 11\}.$$

Compute Residuals:  $e_i = y_i - \hat{y}_i$ .

**Table 4.** Actual returns, predictions, residuals, and squared residuals.

| $y_i$ | $\hat{y}_i$ | $e_i$ | $e_i^2$ |
|-------|-------------|-------|---------|
| 5     | 4           | 1     | 1       |
| 7     | 8           | -1    | 1       |
| 9     | 8           | 1     | 1       |
| 10    | 11          | -1    | 1       |

The sum of squared errors is

$$SSE = 1 + 1 + 1 + 1 = 4.$$

The mean squared error (MSE) is calculated as follows.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2.$$

Thus,

$$MSE = \frac{4}{4} = 1.$$

Now we calculate  $R^2$  as follows. First, compute the sample mean:

$$\bar{y} = \frac{5 + 7 + 9 + 10}{4} = 7.75.$$

The total sum of squares is

$$SST = \sum_{i=1}^n (y_i - \bar{y})^2.$$

Therefore,

$$SST = (5 - 7.75)^2 + (7 - 7.75)^2 + (9 - 7.75)^2 + (10 - 7.75)^2.$$

Hence,

$$SST = 7.5625 + 0.5625 + 1.5625 + 5.0625 = 14.75.$$

The coefficient of determination is

$$R^2 = 1 - \frac{SSE}{SST}.$$

Substituting values yields  $R^2 = 1 - \frac{4}{14.75}$ . Thus,

$$R^2 = 0.7288.$$

Therefore,  $R^2 = 0.73$ : the model explains about 73% of the variation in these four monthly returns.

The mean square error and the coefficient of determination are used to evaluate predictive models' efficiencies. Small MSE means that the predictive model is fairly accurate, while large MSE implies that the model is not a good fit for the data.

$R^2$  is the share of the variation in the data that the model explains. With only four data points, this example shows how  $R^2$  is calculated but tells us very little about how well the model would predict real returns.

## 18. Random Forest Predictive Model

Random Forest is an ensemble learning method that builds a large collection of decision trees and aggregates their individual predictions into a single output. [2] Its popularity in finance comes from three properties in particular: it is robust to noisy data, it captures nonlinear relationships between variables, and it is comparatively resistant to overfitting relative to a single decision tree. These features make it a common tool across a wide range of financial applications, like credit risk modeling, fraud detection, asset allocation, and portfolio management. [2] The section focuses specifically on stock price prediction and algorithmic trading.

A decision tree partitions the feature space into a set of regions by recursively splitting the data according to the values of the input features. [2]

**Definition:** A decision tree is a tree that partitions the feature space into regions by recursively splitting data.

Given a feature vector  $\vec{X} = (x_1, x_2, \dots, x_p)$ , the tree's prediction for an observation falling into leaf node  $R_m$  is the average of the outcomes in that region:

$$\hat{y} = \frac{1}{N_m} \sum_{i \in R_m} y_i,$$

where  $N_m$  is the number of training observations in the region  $R_m$  (leaf node).

### 18.1 Random Forest Algorithm

Random Forest improves decision trees by introducing randomness:

- Bootstrap sampling (bagging): each tree is trained on a random sample of the data, drawn with replacement.

- Random feature subset at each split, only a random subset of the available features is considered, rather than all of them.

The forest's prediction is averaged across  $B$  trees [12]:

$$\hat{f}(x) = \frac{1}{B} \sum_{b=1}^B T_b(x)$$

This combination of bagging and random feature selection reduces the variance of the model relative to any individual tree, allows it to handle high-dimensional data, captures nonlinear relationships in the data and produces feature importance measures that indicate which inputs are driving the model's predictions.

To apply the random forest technique follow the steps below:

1. Collect financial data
2. Decide on features
3. Split into train/test sets
4. Train Random Forest model
5. Evaluate using metrics such as MSE

With the mechanics of the algorithm established, the next section turns to how random forest is actually applied within finance, focusing on stock price prediction and algorithmic trading.

## 19. Applications in Finance

### 19.1 Stock Price Prediction and Algorithmic Trading

Random Forest is well suited to both stock price prediction and algorithmic trading, since returns are driven by relationships between technical indicators, macroeconomic conditions, and sentiment data that are rarely linear. [2] In a typical prediction setup, a model takes inputs such as the past 20 days of returns, trading volume, 5 and 20-day moving averages, and RSI, and outputs a probability that the stock closes higher the next trading day—a signal that can directly support portfolio selection. Algorithmic trading extends this same approach: instead of an up/down probability, the model classifies each day into a “buy,” “sell,” or “hold” decision using similar engineered features (price trends, volatility, order-flow, sector performance), with the resulting signal triggering automated entry or exit rules after backtesting.

### 19.2 Limitations of Random Forests

Despite feature importance scores, Random Forests remain less interpretable than linear models. They can also be computationally expensive with a large number of trees or features, and may still overfit noisy financial data if not properly tuned (e.g., tree depth, number of estimators).

## 20. Conclusion

We started this project by focusing on compound interest and its application to financial planning. Then we discussed expectation and variance of random variables, as well as further classical theories of statistics and probability. Then we introduced the foundations of Modern Portfolio Theory, learning how expected return and variance quantify investment performance, how diversification reduces risk, and how the Sharpe Ratio compares portfolios on a risk-adjusted basis. We introduced CAPM, beta, and the Capital Allocation Line, which connect expected return with systematic market risk and market crashes. These concepts form the mathematical basis for portfolio optimization and quantitative asset management. Finally, we explored machine learning techniques such as linear regression and random forest and applied them to finance.

### Conflicts of Interest

The authors declare that there are no conflicts of interest related to this work.

### Works Cited

1. Markowitz, H. 1952. "Portfolio Selection." *The Journal of Finance* 7 (1): 77–91. March 1, 1952. <https://doi.org/10.2307/2975974>.
2. Zheng, J., Xin, D., Cheng, Q., Tian, M., and Yang, L. 2024. "The Random Forest Model for Analyzing and Forecasting the US Stock Market in the Context of Smart Finance." Preprint, arXiv, February 27, 2024. <https://arxiv.org/abs/2402.17194>.
3. Zhang, H. 2024. "Financial Mathematics in Risk Management: Construction and Optimization of Quantitative Models." *Frontiers in Business Economics and Management* 17 (1): 20–23. November 7, 2024. <http://dx.doi.org/10.54097/hx4n4473>.
4. Cunningham, T. 2024. "Portfolio Optimization with Risk Management through Dividend-Paying Stocks." Master's thesis, St. John's University. [https://scholar.stjohns.edu/theses\\_dissertations/742/](https://scholar.stjohns.edu/theses_dissertations/742/).
5. Bernhardt, D., and Eckblad, M. 2013. "Stock Market Crash of 1987." Federal Reserve History, November 22, 2013. <https://www.federalreservehistory.org/essays/stock-market-crash-of-1987>.
6. Wen, Y., and Arbogast, I. 2021. "How COVID-19 Has Impacted Stock Performance by Industry." *On the Economy* (blog), Federal Reserve Bank of St. Louis, March 21, 2021. <https://www.stlouisfed.org/on-the-economy/2021/march/covid19-impacted-stock-performance-industry>.
7. Cboe Global Indices, LLC. 2026. *Cboe Volatility Index Methodology*. Version 6.0, revised February 26, 2026. [https://cdn.cboe.com/resources/indices/Volatility\\_Index\\_Methodology\\_Cboe\\_Volatility\\_Index.pdf](https://cdn.cboe.com/resources/indices/Volatility_Index_Methodology_Cboe_Volatility_Index.pdf).
8. Sharpe, W. F. 1966. "Mutual Fund Performance." *Journal of Business* 39 (1). January 1966. <https://doi.org/10.1086/294846>.
9. Sharpe, W. F. 1964. "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk." *Journal of Finance* 19 (3). September 1964. <https://doi.org/10.1111/j.1540-6261.1964.tb02865.x>.
10. Lintner, J. 1965. "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets." *Review of Economics and Statistics* 47 (1). <https://www.jstor.org/stable/1924119>.
11. Jorion, P. 2007. *Value at Risk: The New Benchmark for Managing Financial Risk*. 3rd ed. New York: McGraw-Hill.
12. Breiman, L. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32. <https://doi.org/10.1023/A:1010933404324>.