Building Efficient Portfolios *

Aman Rana

Department of Computer Science

University of Toronto

aman.rana@mail.utoronto.ca

25th March, 2025

Abstract

This article provides a brief and accessible overview of three fundamental concepts in modern portfolio theory: the Capital Asset Pricing Model (CAPM), Markowitz Mean-Variance Optimization (MVO), and Arbitrage Pricing Theory (APT). We begin by exploring the CAPM, which describes the relationship between systematic risk and expected return for assets, particularly stocks. Next, we delve into Markowitz MVO, a framework for assembling a portfolio of assets such that the expected return is maximized for a given level of risk, using this to build the efficient frontier. Finally, we examine APT, which extends the ideas of CAPM by considering multiple factors that might affect asset returns. We conclude by discussing the advantages and limitations of these models.

*Code and data supporting this analysis is available at: https://github.com/Aman-Rana-02/Efficient_Market_Portfolios

Contents

1 Introduction

Modern portfolio theory is built on the premise that investors can optimize their returns by balancing systematic risk exposures. Over time, several frameworks have emerged to guide this optimization. The Capital Asset Pricing Model (CAPM) came first, based on the belief that market risk drove expected returns, suggesting that firm-specific risks can be diversified away and markets priced all risk. Markowitz Mean-Variance Optimization (MVO) builds on this notion of diversification, we demonstrate how optimal portfolios can be created to minimize risk for a targeted return. Arbitrage Pricing Theory (APT) extends beyond a single factor by incorporating multiple risk factors (such as size and value) to more accurately capture the sources of systematic risk.

In this article, we use historical price data for S&P 500 constituents gathered from Finaeon's Global Financial Dataset, and offer code and notes. First, we describe how the data is assembled and highlight potential biases such as survivorship bias. Next, we delve into the CAPM, showing how betas are estimated and how they relate to expected returns. We then explore deriving an efficient frontier of optimal portfolios. Finally, we discuss the APT approach, including how Fama-Macbeth regressions help estimate multiple factor betas and risk premia. We provide a brief overview of how these models shape efficient portfolio construction.

The remainder of the article is structured as follows:

- Section 2 provides an overview of the dataset used for analysis.
- Section 3 explores the Capital Asset Pricing Model (CAPM) and its implications.
- Section 4 discusses the importance of diversification and the construction of efficient portfolios.
- Section 5 touches on Arbitrage Pricing Theory (APT) and its applications.
- Section 6 concludes by discussing limitations of the methods discussed.

2 Data

We use Python (Van Rossum and Drake, 2009), Pandas (pandas development team, 2020), NumPy (Harris et al., 2020), Statsmodels (Seabold and Perktold, 2010) to analyze S&P 500 price data from Finaeon's Global Financial Dataset (Finaeon Inc., 2025) and the Fama-French

Company Name	Ticker
Apple Inc.	AAPL
General Electric Co.	GE
International Business Machines Corp.	IBM
The Coca-Cola Company	KO
Microsoft Corp.	MSFT

Table 1: Sample assets and their full names

3-factor data library (French, 2024). We use Matplotlib for plotting (Hunter, 2007). We get the S&P500's current constituents from the Finaeon website, and download a time-series of daily prices for each stock using Finaeon's API. Kenneth french's website (French, 2024) provides the risk-free rate, market excess return, and daily returns of the Size and Value portfolios. The risk-free rate they provide is the 1-month T-bill rate, and the market risk premium is the difference between the total US stock market's return and the risk-free rate. The size and value portfolios are constructed according to Fama and French (1993). By only using the return histories of the current S&P 500 constituents, we introduce a survivorship bias in our analysis. Any risk premia we include will be biased, as we only consider firms that have survived to the present day, and are large. The purpose of this article is to introduce concepts of modern portfolio theory, and not to provide a comprehensive analysis of risk factors and premia, so we note this deficiency and move forward with our analysis.

Throughout this article, we'll use a few sample assets to illustrate concepts. We'll refer to them by ticker symbol, and their full names are shown in table 1. These are randomly chosen from the S&P 500, and are not meant to be representative of the index as a whole.

date	symbol	series_id	close	return	RF	Mkt-RF	SMB	HML
2015-01-02 00:00:00	CSCO	58375	27.61	-0.01	0.00	-0.00	-0.01	0.00
2015-01-05 00:00:00	CSCO	58375	27.06	-0.02	0.00	-0.02	0.00	-0.01
2015-01-06 00:00:00	CSCO	58375	27.05	-0.00	0.00	-0.01	-0.01	-0.00
2015-01-07 00:00:00	CSCO	58375	27.30	0.01	0.00	0.01	0.00	-0.01
2015-01-08 00:00:00	CSCO	58375	27.51	0.01	0.00	0.02	-0.00	-0.00

Table 2: Sample of the dataframe used for analysis, RF, MKT, SMB, HML are the daily risk-free rate, market excess return, size and value portfolio returns respectively.

We clean the data by removing outliers, windsorizing the returns to be within 3 standard deviations of the mean. This helps ensure that our analyses are not influenced by extreme market movements which may be the result of data errors or unusual events.

3 Capital Asset Pricing Model (CAPM)

3.1 Intuition of the CAPM β

$$E[R_i] = R_f + \beta_i(E[R_m] - R_f) \tag{1}$$

The Capital Asset Pricing Model is a model that describes the expected returns of an asset as a function of its exposure to systematic risk. The model is given by Equation 1, where:

- $E[R_i]$ is the expected return of asset i
- R_f is the risk-free rate
- β_i is the asset's beta
- $E[R_m]$ is the expected return of the market

Notice that according to the model, a firm's expected return is only a function of its exposure to the market's risk. This comes from a fundamental belief that in the long-run, any firm-specific/idiosyncratic risk can be diversified away, therefore investors look to be compensated for the risk that cannot be diversified away, which is the market risk.

To estimate the CAPM beta for firms, we regress a firm's historical returns against the market's returns, and take the slope of the line of best fit. This is given by Equation 2:

$$R_i - R_f = \alpha + \beta_i(R_m - R_f) + \epsilon \tag{2}$$

Where:

- R_i is the firm's return
- R_f is the risk-free rate
- R_m is the market return
- α is the intercept of the regression line
- β_i is the slope of the regression line
- ϵ is the error term

Asset	Beta
AAPL	1.10
GE	0.92
IBM	0.75
KO	0.49
MSFT	1.14

Table 3: Sample assets and their CAPM betas

symbol	capm_beta	capm_expected_return	std_dev_ann
AAPL	1.10	0.16	0.25
GE	0.92	0.14	0.29
IBM	0.75	0.12	0.21
KO	0.49	0.08	0.16
MSFT	1.14	0.17	0.24

Table 4: Sample assets and their expected returns using the CAPM

α is also known as Jensen's alpha, it is a reflection of how much a firm under or outperforms the market, given its beta. In a world where CAPM perfectly models a firm's returns, α should be zero.

Table 3 shows the CAPM betas for a few sample assets. Higher beta assets are 'riskier' since they are more exposed to the market's risk, and therefore should have higher expected returns.

We can see this more clearly in figure 1, where we plot the historical average annualized return against a firm's beta. We can see that there is a positive relationship between beta and expected return. In this same plot we add the Security Market Line (SML), which is the line that represents the expected return of an asset given its beta. We notice that not all stocks lie on this line, stocks above this line are out performing their expected return, and stocks below this line are under performing their expected return. But, this evaluation of performance is only valid if we assume that the CAPM is a perfect model of the market, which it may not be. To note is that we already discussed the Fama-French 3 factor model which is already an extension showing that the CAPM model is underspecified.

Working backwards, we can get the CAPM expected return for a firm by multiplying each firm's CAPM beta by the average market return over the period. Table 4 shows the expected returns for a few sample assets, given their CAPM and the average market return.

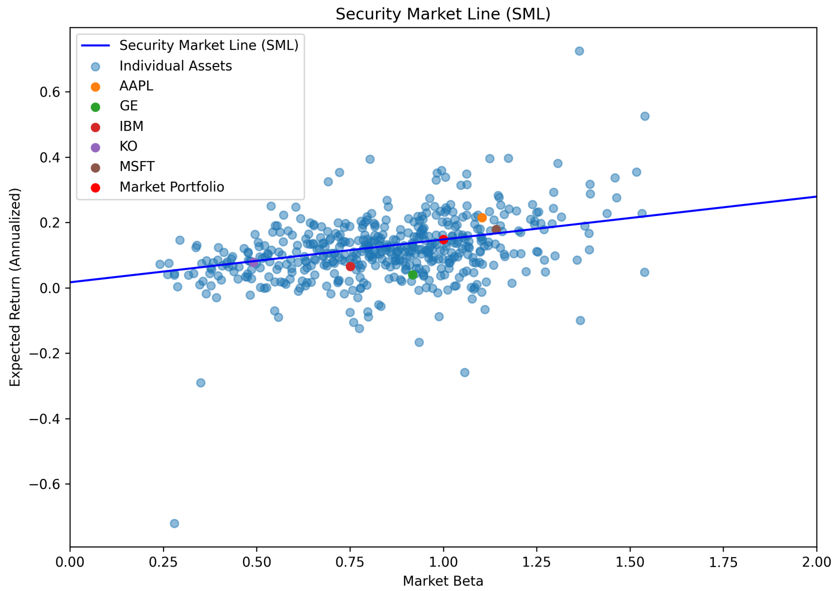

Figure 1: Plot showing the historical average annualized return against a firm's beta

3.2 Cost of Capital - DDM vs CAPM

Recall the definition of cost of capital, and that it is synonymous with expected return. We previously explored using the Dividend Discount Model (DDM) to estimate the cost of capital. The intuition of the DDM is that a stock price represents the present value of all future cashflows, where we proxy all future cashflows using a dividend and expected growth rate. A DDM feels more realistic the better we can forecast a firm's future dividends, as well as their dividend growth rate, given that they pay dividends. If we start to consider that there are firms that don't pay dividends at all, or that the dividend growth rate is hard to forecast, we start to see the limitations of the DDM. Additionally, if we are investors that don't care about dividend payouts, we may not find the DDM useful at all.

The CAPM, on the other hand, is a model that is based on the idea that investors are compensated for the systematic risk they take on. If a risk can be diversified away, then investors shouldn't be compensated for it (as they can get rid of it). If markets are efficient, and investors are rational, then the current stock price should reflect all available information about a firm. There is no need to forecast dividends, or growth rates, or anything else about the firm. We only need to know the firm's exposure to systematic risk.

4 Diversification and Portfolios

By now we have an understanding of how to estimate the expected return of an asset using the Capital Asset Pricing Model (CAPM), and how risk and return should have a proportional relationship. In this section we introduce the concept of diversification, and how it can be used to construct efficient portfolios.

4.1 The Simple Math of Diversification

$$E[R_p] = \sum_{i=1}^{n} w_i E[R_i] \tag{3}$$

Where:

- $E[R_p]$ is the expected return of the portfolio
- w_i is the weight of asset i in the portfolio
- $E[R_i]$ is the expected return of asset i
- n is the number of assets in the portfolio
- w_i is the weight of asset i in the portfolio
- $E[R_i]$ is the expected return of asset i

The expected return of a portfolio is simply the weighted average of the expected returns of the assets in the portfolio.

$$\sigma_p^2 = \sum_{i=1}^{n} w_i^2 \sigma_i^2 + \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j \sigma_{ij} \tag{4}$$

Where:

- σ_p^2 is the variance of the portfolio
- w_i is the weight of asset i in the portfolio
- σ_i^2 is the variance of asset i
- σ_{ij} is the covariance between assets i and j
- n is the number of assets in the portfolio

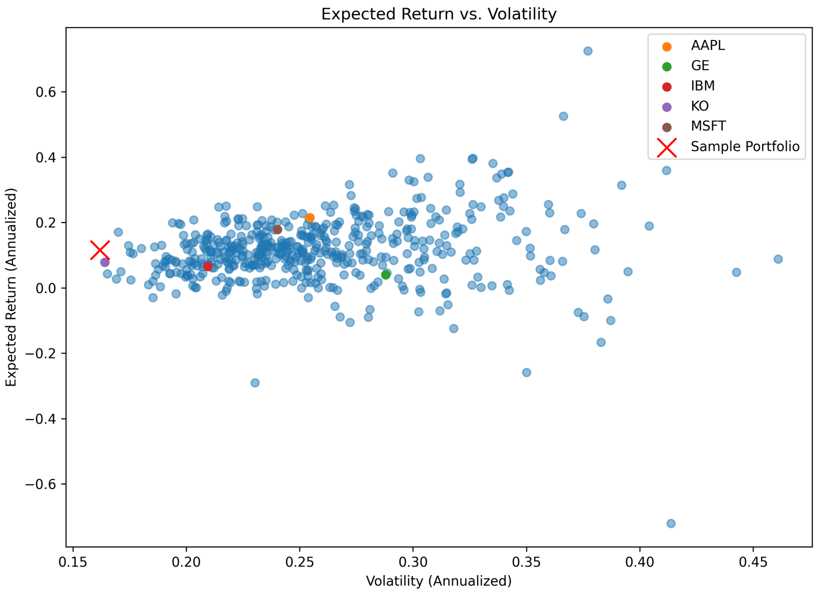

Figure 2: Plot showing the expected return and volatility of our sample assets, as well as the expected return and volatility of an equal weighted portfolio of our sample assets

- w_i is the weight of asset i in the portfolio

The variance of a portfolio is the weighted average of the variances of the assets in the portfolio, plus the covariance between the assets. Since we square the weights, and they are between 0 and 1, the variance of the portfolio is always less than or equal to the weighted average of the variances of the assets in the portfolio. This means that as we add more assets to the portfolio, the variance of the portfolio will decrease. This is the essence of diversification, and it is the reason why we can construct portfolios that have lower expected risk for the same expected return.

If we plot the expected returns and volatilities of our stock universe, as well as the expected return and volatility of an equal weighted portfolio of our sample assets, we can see that the equal weighted portfolio has a higher expected return for a given level of risk than any of the individual assets. This is shown in figure 2, where we plot the expected return and volatility of our sample assets, as well as the expected return and volatility of an equal weighted portfolio of our sample assets.

4.2 Efficient Portfolios

The MVO portfolio is the optimal portfolio for a given level of risk. It is the portfolio that has the highest expected return for a given level of risk, or the lowest risk for a given level of expected return. This is typically done by solving the following optimization problem:

$$\begin{aligned} \text{minimize} \quad & \sigma_p^2 \\ \text{subject to} \quad & E[R_p] = E^* \\ & 1 \geq w_i \geq 0, \forall i \end{aligned} \tag{5}$$

Where:

- $E[R_p]$ is the expected return of the portfolio
- E^* is the target expected return
- σ_p^2 is the variance of the portfolio
- w_i is the weight of asset i in the portfolio
- n is the number of assets in the portfolio

Lets say that we can't short-sell, or take leverage, therefore we have the constraint that $1 \geq w_i \geq 0, \forall i$.

In our implementation, we use the scipy library to solve this optimization problem, minimizing the variance for a range of target returns. This allows us to trace out the efficient frontier and identify two key portfolios: the minimum variance portfolio, which has the lowest possible risk regardless of return, and the maximum Sharpe ratio portfolio, which offers the highest risk-adjusted return.

4.3 The Efficient Frontier

The efficient frontier is just the set of all efficient portfolios. We can see this visually in figure 3, where we plot the expected return of every efficient portfolio at every level of risk.

4.4 The Tangent Portfolio and The Capital Market Line

The capital market line (CML) is the line that connects the risk-free rate to the tangent portfolio. The tangent portfolio is the optimal risky portfolio, and it is the portfolio that has the highest

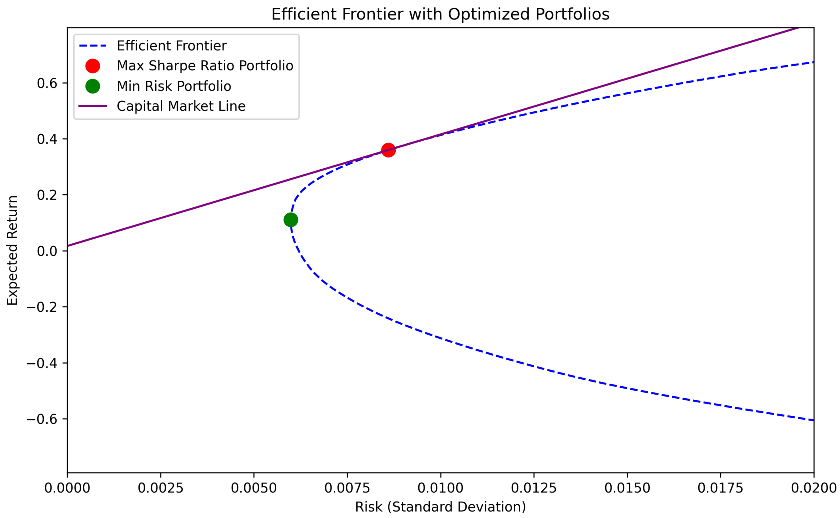

Figure 3: Plot showing the efficient frontier, tangent portfolio, capital market line, and the minimum variance portfolio

Sharpe ratio. The Sharpe ratio is given by:

$$S = \frac{E[R_p] - R_f}{\sigma_p} \tag{6}$$

Where:

- S is the Sharpe ratio
- $E[R_p]$ is the expected return of the portfolio
- R_f is the risk-free rate
- σ_p is the standard deviation of the portfolio

The Sharpe ratio is a measure of the risk-adjusted return of a portfolio, and it is used to compare the performance of different portfolios. The higher the Sharpe ratio, the better the portfolio's return per unit of risk. In Figure 3, the max Sharpe ratio portfolio is marked with a red circle.

Going back to Figure 3, we can see that the CML is the line that connects the risk-free rate to the tangent portfolio. It represents the set of all efficient portfolios that can be constructed by combining the risk-free asset with the tangent portfolio. The line to the left of the tangent portfolio represents having a positive exposure to the risk-free asset, and the line to the right of the tangent portfolio represents having a negative exposure to the risk-free asset. A negative exposure means taking out a loan to invest in the tangent portfolio. The figure also shows the minimum variance portfolio (green circle), which is the portfolio with the lowest possible risk on the efficient frontier.

The formula for the CML is given by some weight on the risk free rate, and some weight on the tangent portfolio:

$$E[R_p] = w_f R_f + w_t E[R_t] \tag{7}$$

Where:

- $E[R_p]$ is the expected return of the portfolio
- w_f is the weight of the risk-free asset in the portfolio
- R_f is the risk-free rate
- w_t is the weight of the tangent portfolio in the portfolio
- $E[R_t]$ is the expected return of the tangent portfolio

5 Arbitrage Pricing Theory

In this section we'll briefly discuss Arbitrage Pricing Theory (APT), and how it generalizes the Capital Asset Pricing Model (CAPM) as well as how Fama-Macbeth regressions can be used to estimate risk premia. The purpose of this section is not to be comprehensive, but to briefly introduce concepts, as well as the code used when implementing portfolios in a systematic risk framework.

5.1 Arbitrage Pricing Theory

We've discussed Fama and French (1993) and the Size and Value risk factors it proposes. Throughout the article so far we've focused on one source of systematic risk, market risk. From now on we will include the Size and Value factors in our equation for expected returns.

This multi-factor model has the form as per Equation 8. This is the model given by Arbitrage Pricing Theory (APT).

$$E[R_i] = R_f + \beta_{i,m}\gamma_m + \beta_{i,s}\gamma_s + \beta_{i,v}\gamma_v \tag{8}$$

Where:

Mkt-RF_loading	SMB_loading	HML_loading	symbol
1.113	-0.251	-0.379	AAPL
0.983	0.140	0.582	GE
0.804	-0.147	0.335	IBM
0.550	-0.333	0.216	KO
1.168	-0.404	-0.449	MSFT

Table 5: Sample assets and their exposures to the market, size and value risk factors

- $E[R_i]$ is the expected return of asset i
- R_f is the risk-free rate
- $\beta_{i,m}$ is the asset's exposure to the market risk factor
- $\beta_{i,s}$ is the asset's exposure to the size risk factor
- $\beta_{i,v}$ is the asset's exposure to the value risk factor
- γ_m is the market risk premium
- γ_s is the size risk premium
- γ_v is the value risk premium

For an asset i, the expected return is a function of its exposure to the market, size, and value risk factors, as represented by the Betas. Risk premia are the expected returns of the systematic risk factors, and they are the same for all assets. Recall in the CAPM model, we were concerned only with the market risk premium, now we have three risk premia.

Table 5 shows the exposures of a few sample assets to the market, size and value risk factors.

We used to get the market risk premia by taking the average market return over our data period, but we will now introduce a way to get the risk premia for the size and value factors.

5.2 Fama-Macbeth Regressions

Fama-Macbeth regressions are a two stage process that allows us to get the risk premia and factor sensitivities for a set of risk factors.

The first stage is a cross-sectional regression of the asset returns against the risk factors. This gives us the factor sensitivities (betas) for each asset. The second stage is a time-series regression of the factor sensitivities against the risk factors. This gives us the risk premia for each factor at every time-step T.

Stage 1 is given by Equation 9:

$$R_i = \alpha + \beta_{i,m} R_m + \beta_{i,s} R_s + \beta_{i,v} R_v + \epsilon \tag{9}$$

Where:

- R_i is the return of asset i
- α is the intercept of the regression line
- $\beta_{i,m}$ is the asset's exposure to the market risk factor
- $\beta_{i,s}$ is the asset's exposure to the size risk factor
- $\beta_{i,v}$ is the asset's exposure to the value risk factor
- R_m is the market return
- R_s is the size return
- R_v is the value return
- ϵ is the error term

The second stage is given by Equation 10, where we're estimating the gammas using the betas from the first stage at every time period t.

$$R_t = \gamma_{0,t} + \gamma_{m,t} \beta_{i,m} + \gamma_{s,t} \beta_{i,s} + \gamma_{v,t} \beta_{i,v} + \epsilon_t \tag{10}$$

Where:

- R_t is the return of the portfolio at time t
- $\gamma_{0,t}$ is the intercept of the regression line at time t
- $\gamma_{m,t}$ is the market risk premium at time t
- $\gamma_{s,t}$ is the size risk premium at time t
- $\gamma_{v,t}$ is the value risk premium at time t
- ϵ_t is the error term at time t

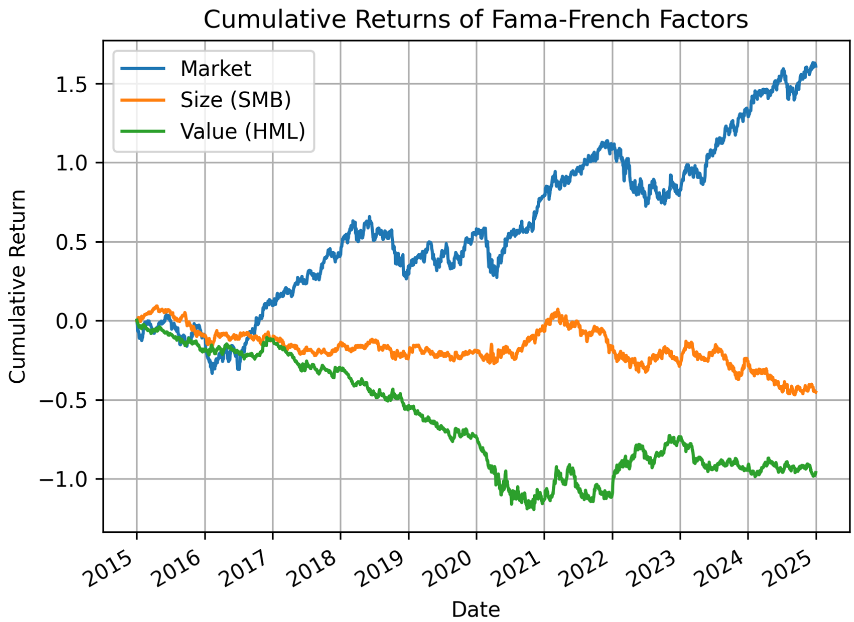

Figure 4: Plot showing the cumulative sum of the risk premia for the market, size and value factors

We then take an average of the timeseries regression coefficients to get the risk premia for each factor. These risk premia are also referred to as factor returns, since they are the expected returns of the risk factors.

If we take the cumulative sum of these risk premia coefficients, we can visually see the performance of the risk factors over time. We can see this in figure 4, where we plot the cumulative sum of the risk premia for the size and value factors.

We can take the averages of these coefficients to get the risk premia for each factor. The results are shown in table 6. The negative premia we see on size and value factors come from a mix of using a small sample of assets, namely the bias from using the current constituents S&P 500, our relatively short lookback period of 10 years, and how value has underperformed in the ast few years.

We need two parts to construct portfolios, expected returns and risk. We can get the expected returns from the Fama-Macbeth regressions, and we now, since we believe that returns are only a function of risk factors, we can use the covariance matrix of the risk factors to get the risk of the portfolio. This greatly reduces the dimensionality of the problem, since we only need to estimate the covariance matrix of the risk factors, and not the covariance matrix of the assets. Expected returns for APT portfolios are still weighted sums of the expected returns of the component

Factors	Expected Factor Returns
Mkt-RF_loading_premium	0.1612
SMB_loading_premium	-0.0453
HML_loading_premium	-0.0962

Table 6: Expected factor returns for the market, size and value factors

symbol	expected_return
A	0.17
AAPL	0.23
ABBV	0.10
ABNB	0.19
ABT	0.16

Table 7: Expected returns according to APT for a few sample assets

assets. Risk is given slightly differently, as we are now using the covariance matrix of the risk factors, and not the covariance matrix of the assets. The expected risk is given by Equation 11:

$$\sigma_p^2 = w^T \Sigma w \tag{11}$$

Where:

- σ_p^2 is the variance of the portfolio
- w is the vector of weights of the assets in the portfolio
- Σ is the covariance matrix of the risk factors

Table 7 shows the expected returns for a few sample assets. Table 8 shows the covariance matrix of the risk factors.

	Mkt-RF_premium	SMB_premium	HML_premium
Mkt-RF_premium	0.06	0.00	-0.00
SMB_premium	0.00	0.02	0.00
HML_premium	-0.00	0.00	0.02

Table 8: Factor covariance matrix for the market, size and value factors

6 Conclusion

6.1 Limitations

In this article and the provided repository we briefly introduce some principles of modern portfolio theory. We showed how APT portfolios could be used to add more informative risks, and provided code with Fama-Macbeth regressions to facilitate future work on exploring different risk factors. This framework also enables factor neutral portfolios, which can be used to create portfolios that are neutral to certain risk factors, but the code provided does not implement this.

The limitations of these models remain that they use historical data, APT attempts to address this by using expected returns generated by risk premia, risk premia exposures, and risk factor covariances, which may be more stable than individual asset returns. The out of sample performance of MVO portfolios is a common criticism, and frameworks such as Hierarchical Risk Parity (HRP) and Black-Litterman have been proposed to address this.

Additionally, in practice there are even more factors added, and even factor covariance matrices can start to suffer from the numerical instability that vanilla MVO suffers from. The code provided in this repository is a starting point for exploring these models, and we encourage readers to explore the literature and implementations of these models to gain a deeper understanding of their limitations and potential improvements.

References

Fama, E. F. and K. R. French (1993, Feb). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics 33*(1), 3–56.

Finaeon Inc. (2025). Global Financial Data. `https://www.finaeon.com/`. Accessed: 2025-02-20.

French, K. R. (2024). Data library. Accessed: 2024-10-16.

Harris, C. R., K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant (2020, September). Array programming with NumPy. *Nature 585*(7825), 357–362.

Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing in Science & Engineering 9*(3), 90–95.

pandas development team, T. (2020, February). pandas-dev/pandas: Pandas.

Seabold, S. and J. Perktold (2010). statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*.

Van Rossum, G. and F. L. Drake (2009). *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace.