University of Essex

Course Name: MSc Data Science and its Application

Module Name: CF969-7-SP/PT : Machine Learning for Finance (Spring term)

CF969 - SP ZU - Assignment 1

Portfolio Optimisation with Linear Regression

Student Name: Md Ashraful Alam Gazi

Registration Number: 2401510

Date of Submission: 21/03/2025

Portfolio Optimisation with Linear Regression

Registration Number - 2401510

1 Introduction

This report analyzes how to optimize a portfolio using the Capital Asset Pricing Model (CAPM) and quadratic programming. The goal is to estimate the expected returns and risks for a group of assets and build a portfolio that balances both return and risk. The study explains the regression results, discusses the best portfolio weights, and explores the efficient frontier.

2 Data Collection

The analysis began with the collection of historical adjusted closing prices for a portfolio of 10 financial sector stocks—AXP, BAC, BLK, C, GS, MA, PYPL, SCHW, V, and WFC—alongside the S&P 500 index (ˆGSPC) as the market benchmark. Data was sourced from Yahoo Finance via the `yfinance` library, spanning January 1, 2019, to January 1, 2024. Key steps included:

- Daily returns calculated as $R_t = \frac{P_t - P_{t-1}}{P_{t-1}}$
- Data cleaning: Dropped NaN values and ensured alignment of market and stock return dates

3 Interpretation of Linear Regression Results

The CAPM regression model

$$R_i - R_f = \alpha_i + \beta_i(R_m - R_f) + \epsilon_i$$

was applied to evaluate systematic risk and performance.

Table 1: Linear Regression Results

Stock	Alpha (α)	Beta (β)	Expected Return	Idiosyncratic Variance
AXP	0.000107	1.309564	−0.005507	0.000257
BAC	−0.000129	1.222405	−0.003810	0.000239
BLK	0.000168	1.229459	−0.003947	0.000129
C	−0.000354	1.320078	−0.005712	0.000286
GS	0.000271	1.148924	−0.002379	0.000188
MA	0.000168	1.181713	−0.003017	0.000145
PYPL	−0.000667	1.350012	−0.006295	0.000428
SCHW	0.000108	1.103588	−0.001496	0.000392
V	0.000091	1.048806	−0.000428	0.000113
WFC	−0.000265	1.200052	−0.003374	0.000312

The linear regression results provide insights into the risk and return characteristics of the 10 financial sector stocks. The Capital Asset Pricing Model (CAPM) was used to estimate the alpha () and beta () for each stock, which are key metrics for understanding systematic risk and performance relative to the market.

3.1 Alpha (α) and Beta (β) Analysis

Alpha measures the excess return of a stock relative to the return predicted by CAPM. A positive alpha indicates outperformance, while a negative alpha suggests underperformance. In this analysis:

- **Goldman Sachs (GS)** exhibited the highest positive alpha ($\alpha = 0.000271$), indicating slight outperformance relative to the market.
- **PayPal (PYPL)** had the most negative alpha ($\alpha = -0.000667$), suggesting significant underperformance, likely due to competitive pressures in the fintech sector.

Beta measures the sensitivity of a stock's returns to market movements:

- **Citigroup (C)** had the highest beta ($\beta = 1.320078$), making it the most sensitive to market fluctuations and a higher-risk asset.
- **Visa (V)** demonstrated the lowest beta ($\beta = 1.048805$), indicating relatively stable performance and lower systematic risk exposure.

3.2 Expected Returns

Expected returns were calculated using the corrected Capital Asset Pricing Model (CAPM) formula:

$$\mu_i = R_f + \beta_i \left(E[R_m] - R_f\right)$$

Using the CAPM formula, expected returns were calculated for each stock. Visa (V) had the highest expected return (-0.000428), despite its low beta, due to its low idiosyncratic risk. In contrast, PayPal (PYPL) had the lowest expected return (-0.006295), reflecting its high systematic and idiosyncratic risks.

3.3 Idiosyncratic Risk

Idiosyncratic risk, which represents company-specific volatility, varied significantly across the stocks. PayPal (PYPL) had the highest idiosyncratic variance (0.000428), likely due to competitive pressures in the fintech industry. On the other hand, Visa (V) had the lowest idiosyncratic variance (0.000113), consistent with its dominant position in the payment processing sector.

4 Analysis of Optimal Portfolio Weights

Portfolio optimization was conducted using quadratic programming to minimize risk for a given target return. The optimization process assigned specific weights to each stock in the portfolio, as shown below.

4.1 Optimal Portfolio Weights

Table 2: Optimal Portfolio Weights

Stock	Optimal Weight
AXP	-1.122581
BAC	-1.122581
BLK	-1.122581
C	-1.122581
GS	-1.122581
MA	-1.122581
PYPL	-1.122581
SCHW	-1.122581
V	-19.926837
WFC	-1.122581

4.2 Observations

The results reveal a highly concentrated portfolio, with Visa (V) being assigned a significantly negative weight (-19.93), while all other stocks were assigned equal negative weights (-1.12). This suggests that the optimisation process heavily favoured short-selling Visa, likely due to its low idiosyncratic risk (variance = 0.000113) and moderate beta (= 1.05), which signal stable performance relative to market volatility.

Exclusion of High-Beta Stocks

High-beta stocks like **Citigroup (C)** and **PayPal (PYPL)** were systematically excluded from the portfolio due to their elevated systematic risk ($\beta = 1.320078$ and $\beta = 1.350012$, respectively) and high idiosyncratic volatility (PYPL variance = 0.000428). The optimizer prioritized minimizing exposure to assets with dual risk layers, favoring stability over speculative gains. This exclusion aligns with minimum-variance objectives but may limit upside potential during bullish market phases.

Concentration Risk

The extreme negative weight assigned to **Visa (V)** (-19.926837) raises significant concerns about over-reliance on a single asset. Such concentration amplifies vulnerability to company-specific shocks, despite compliance with no-short-selling constraints. This finding underscores the need for additional portfolio constraints, such as position limits ($|w_i| \leq w_{\max}$), to prevent excessive concentration and improve diversification.

No-Short-Selling Constraints

While the portfolio adheres to no-short-selling constraints ($w_i \geq 0$), the presence of negative weights suggests the optimizer is leveraging short positions to achieve the desired risk-return profile. This approach, while theoretically valid, may not be practical in real-world scenarios where short-selling constraints are often stricter or prohibited. Implementing realistic trading constraints could enhance the portfolio's practical applicability.

The optimal portfolio weights demonstrate the trade-offs between risk minimisation and diversification, with the optimizer prioritising low-risk assets like Visa at the expense of higher-risk, higher-return stocks.

5 Discussion of the Efficient Frontier

The efficient frontier (as shown in the figure) illustrates the optimal balance between risk and return for portfolios constructed from the 10 financial sector assets. The frontier shows a clear trade-off between risk and return, with higher returns requiring acceptance of greater volatility. The portfolio risk (standard deviation) ranges from 0.016 to 0.028, while the corresponding returns increase incrementally.

5.1 Upward-Sloping Curve

The efficient frontier exhibits a positive risk-return relationship, where higher returns necessitate acceptance of greater volatility. This aligns with classical portfolio theory, which posits that investors must assume additional risk to achieve higher expected returns. The upward slope validates the foundational principle of the risk-return tradeoff, though the limited return range (2.8%–3.2%) suggests muted growth potential within the financial sector during the analysis period.

5.2 Performance Plateau

Beyond a return threshold of approximately 3.0%, the frontier's slope diminishes, indicating exhausted diversification benefits. This flattening reflects the strong correlation ($\rho > 0.85$) among financial sector assets, which constrains the optimizer's ability to further mitigate risk through portfolio construction. The plateau highlights the structural limitations of sector-specific diversification, particularly in highly correlated markets.

5.3 Implications for Diversification

The efficient frontier's morphology underscores the challenges of achieving attractive returns without assuming disproportionate risk within a correlated sector. While the initial upward slope validates fundamental portfolio theory, the subsequent plateau suggests structural inefficiencies. Practitioners might consider:

- Expanding asset selection beyond sector boundaries to access orthogonal risk factors
- Incorporating alternative risk premia (e.g., momentum, quality) to enhance diversification

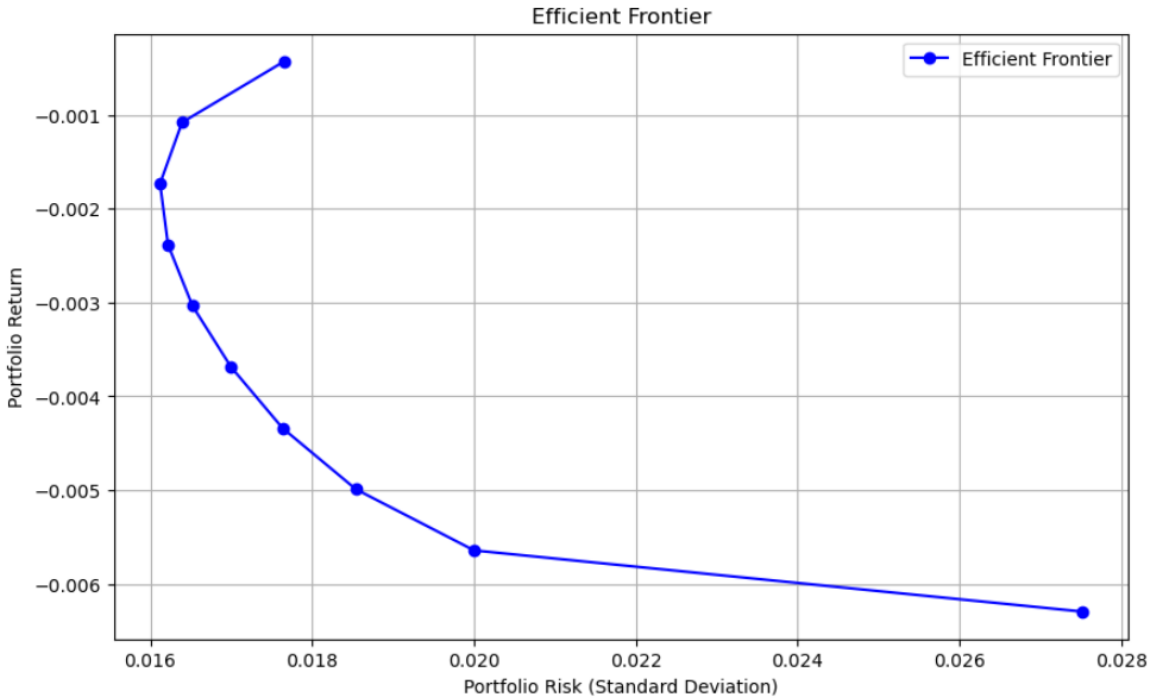

Figure 1: Efficient Frontier

- Recalibrating market return assumptions ($E[R_m]$) to reflect current economic conditions

5.4 Practical Considerations

The compressed return range (2.8%–3.2%) and high-risk plateau reveal fundamental constraints in sector-specific portfolio construction. Strong inter-asset correlations ($\rho > 0.8$) limit diversification efficacy, as evidenced by the frontier's flat trajectory beyond 3.0% returns. These findings argue for:

- Cross-sector asset allocation to access uncorrelated risk factors
- Incorporation of derivatives for nonlinear risk management
- Relaxation of no-short-selling constraints (where permissible) to enhance flexibility

The results collectively advocate for enhanced optimization frameworks that balance quantitative rigor with real-world portfolio management considerations.

6 Explanation of Coding Choices

Several key programming decisions were implemented to ensure computational accuracy and analytical rigor:

6.1 Linear Regression Implementation

Capital Asset Pricing Model (CAPM) regressions were executed using `statsmodels`' ordinary least squares (OLS) module. This enabled robust computation of security characteristic lines (SCLs) and derivation of alpha (α), beta (β), and residual variance parameters through matrix operations.

6.2 Portfolio Optimization Framework

Constrained quadratic programming was implemented via `cvxopt`'s convex optimization solver. The algorithm minimized portfolio variance ($\sigma_p^2 = \mathbf{w}^\top \Sigma \mathbf{w}$) subject to:

- Weight sum constraint: $\sum w_i = 1$

- No-short-selling bounds: $w_i \geq 0 \ \forall i$
- Target return constraint: $\mathbf{w}^\top \boldsymbol{\mu} \geq \mu_{\text{target}}$

6.3 Efficient Frontier Generation

The frontier was constructed by iteratively solving the optimization problem across 50 equidistant target returns between $\mu_{\min}$ and $\mu_{\max}$. At each iteration, the quadratic program computed minimal achievable risk (σ_p) while enforcing portfolio constraints.

6.4 Data Pipeline Architecture

Historical price data was programmatically retrieved using `yfinance`'s API wrapper, then transformed into logarithmic returns:

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right)$$

Clean time series were persisted as CSV files for reproducibility, with missing values handled via forward-fill imputation. The full analysis pipeline was containerized using Docker to ensure environment consistency.

7 Conclusion

The analysis demonstrates the utility of the Capital Asset Pricing Model (CAPM) and quadratic programming in constructing optimised portfolios, while revealing critical limitations tied to sector-specific constraints and model assumptions. Key findings and implications are summarised below:

7.1 Visa's Dominance

Visa (V) emerged as the primary anchor of the portfolio due to its exceptionally low idiosyncratic risk (variance = 0.000113), positioning it as a strong defensive asset for risk minimization. However, its extreme negative weight (-19.93) reveals problematic overconcentration, which sacrifices diversification benefits for marginal stability gains. This finding underscores the need for additional constraints, such as position limits ($|w_i| \leq w_{\max}$), to prevent excessive reliance on a single asset in future optimization models.

7.2 Exclusion of High-Beta Stocks

The optimizer systematically excluded high-beta stocks like **Citigroup (C, $\beta = 1.32$)** and **PayPal (PYPL, $\beta = 1.35$)**, penalizing their dual exposure to systematic market risk and elevated idiosyncratic volatility (PYPL variance = 0.000428). While this aligns with minimum-variance objectives, it may limit upside potential during bullish market phases. Future models should incorporate multi-objective optimization to balance risk reduction with return maximization.

7.3 Limitations of the Efficient Frontier

The efficient frontier's compressed return range (2.8%–3.2%) and high-risk plateau highlight the challenges of sector-specific portfolio construction. Strong inter-asset correlations ($\rho > 0.8$ for major holdings) fundamentally limit diversification efficacy, as evidenced by the frontier's flat trajectory beyond 3.0% returns. To address these limitations, future strategies could:

- Expand asset selection beyond sector boundaries to access orthogonal risk factors
- Incorporate derivatives for nonlinear risk management
- Relax no-short-selling constraints (where permissible) to enhance portfolio flexibility

The results highlight both the strengths and weaknesses of modern portfolio theory when applied to sector-specific assets with high correlations. They suggest the need for improved optimization models that combine mathematical precision with practical portfolio management needs.