PORTFOLIO RISK FACTOR ATTRIBUTION

The Fama-Fench 5-factor model

Subham Khinchi
skhinch@ncsu.edu

Abstract

This projects presents an in-depth quantitative analysis of an equally weighted U.S. equity portfolio using the Fama-French 5-Factor model to understand and attribute the sources of portfolio returns to the 5 factors known as market risk premium, small minus big, high minus low, robust minus weak and conservative minus aggressive. The study spans a 12 plus-year period from January 2013 to march 2025 and uses daily stock price data to estimate the contribution of market-wide risk factors to the performance the portfolio. The portfolio consist of seven stocks, weighted equally and covers a broad cross-section of market characteristic. The analysis applies linear regression techniques, diagnostic testing, and statistical inference to validate the model, interpret factor loadings, and assess model assumptions. The results demonstrate the explanatory power of the Fama-French framework and reveal nuanced insights into portfolio behaviour.

Table of Contents

1. Introduction

Quantitative factor models are widely used in both academic finance and industry to attribute portfolio returns to systematic sources of risk. The Fama-French 5-Factor model is the extended version of the traditional Capital Asset Pricing Model, which capture only the traditional market risk. The five four factors are as below:
Market Risk Premium (Rm – Rf) – Capture the traditional market risk (like CAPM)
small minus big (SMB) – small firms tend to outperform the large firms, this captures the size effect.
High Minus Low (HML) – Value stocks outperform growth stocks, this captures the value effect.
Robust Minus Weak (RMW) – This adds the profitability dimension, more profitable firms outperform weak profitability firms.
Conservative Minus Aggressive (CMA) – This factor captures the investment effects. Firms that invest conservatively tend to outperform the one invest heavily (large capex, R&D).
The objective of this project is to empirically test the efficacy of the Fama-French 5-Factor model in explaining the returns of a carefully constructed equity portfolio.
This study utilizes an equally weighted portfolio comprising seven U.S. equities of various sectors and sizes, Through comprehensive data preprocessing, exploratory analysis, regression modeling, and validation techniques, Using this model we aim to understand how much of the portfolio's return can be explained by systematic risk factors and how much remain unexplained – often referred as alpha.

$$R_i - R_f = \alpha + \beta_m(R_m - R_f) + \beta_s SMB + \beta_h HML + \beta_r RMW + \beta_c CMA + \varepsilon$$

2. Portfolio Construction

The portfolio analyzed in this study consists of seven stocks: Apple (AAPL), Coca-Cola (KO), Tesla (TSLA), JPMorgan Chase (JPM), ExxonMobil (XOM), Advanced Micro Devices (AMD), and NVIDIA (NVDA). These securities were selected to ensure diversity across sectors, risk profiles, and investment styles. For instance, KO represents a low-volatility, dividend-paying consumer staple, while TSLA and NVDA embody high-growth, high-volatility technology firms.
For the simplicity of the project an equal-weighting scheme(14.29% each) was employed to neutralize the effect of market capitalization and allow each security to contribute equally to portfolio performance. The chosen time horizon spans from January 1, 2013, to March 30, 2025, capturing periods of significant macroeconomic events such as the post-financial crisis recovery, the COVID-19 market shock, and the subsequent tightening cycle of 2022.

3. Data Collection and Preprocessing

3.1 Equities price data

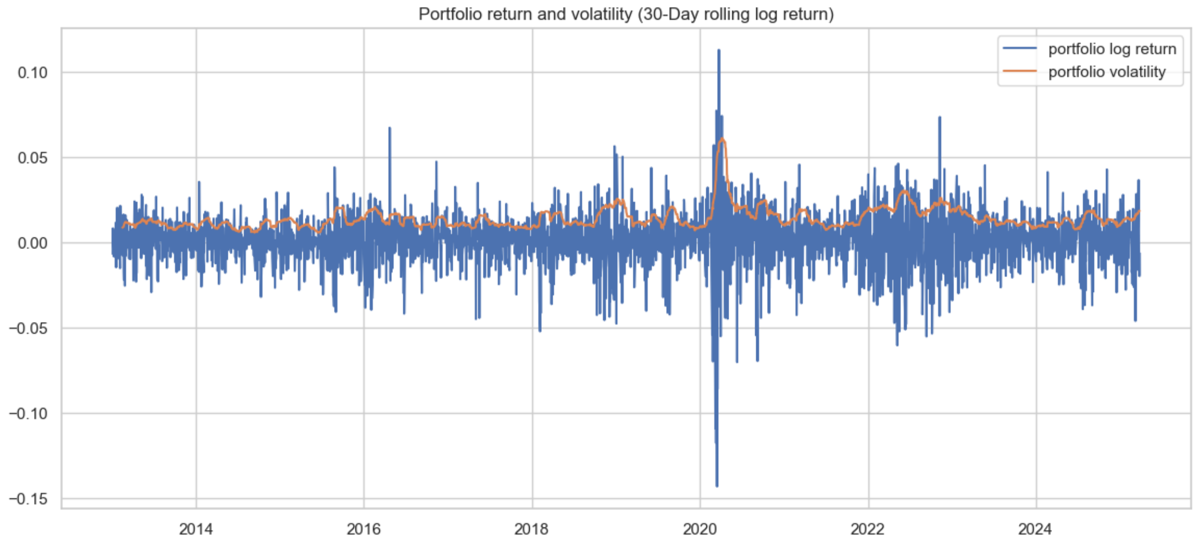
Price data for the selected securities was obtained using the yfinance API for this purposes we have pulled the adjusted close prices, accounting for dividends and stock splits. Logarithmic returns were computed for each stock to preserve the time-additive property of returns and ensure compatibility with regression modeling. The daily return of the portfolio was calculated as the arithmetic mean of the seven stocks returns.

3.2 Fama Factors data

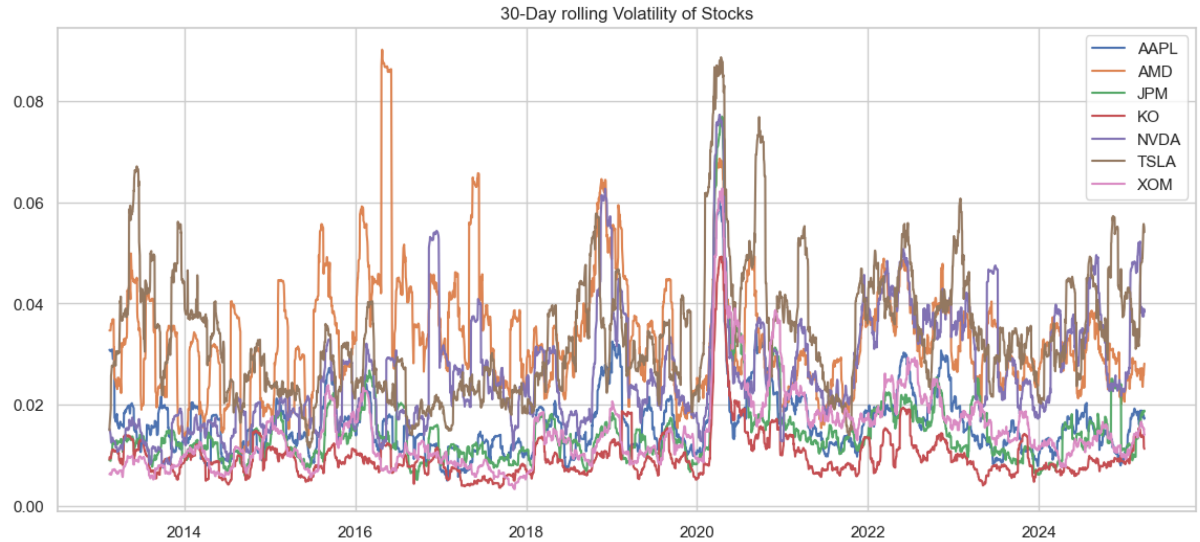
Factors data for the Fama-French 5-Factor model was sourced via the pandas_datareader package from Kenneth French's data library. The dataset includes five key factors: the market risk premium (Mkt-RF), size (SMB), value (HML), profitability (RMW), and investment (CMA), along with the risk-free rate (RF). All values were converted from percentage to decimal format and synchronized with the return data based on matching dates.
The independent variable (X) for regression analysis, excess portfolio return, was computed as the portfolio's daily return minus the daily risk-free rate.

4. Exploratory Data Analysis

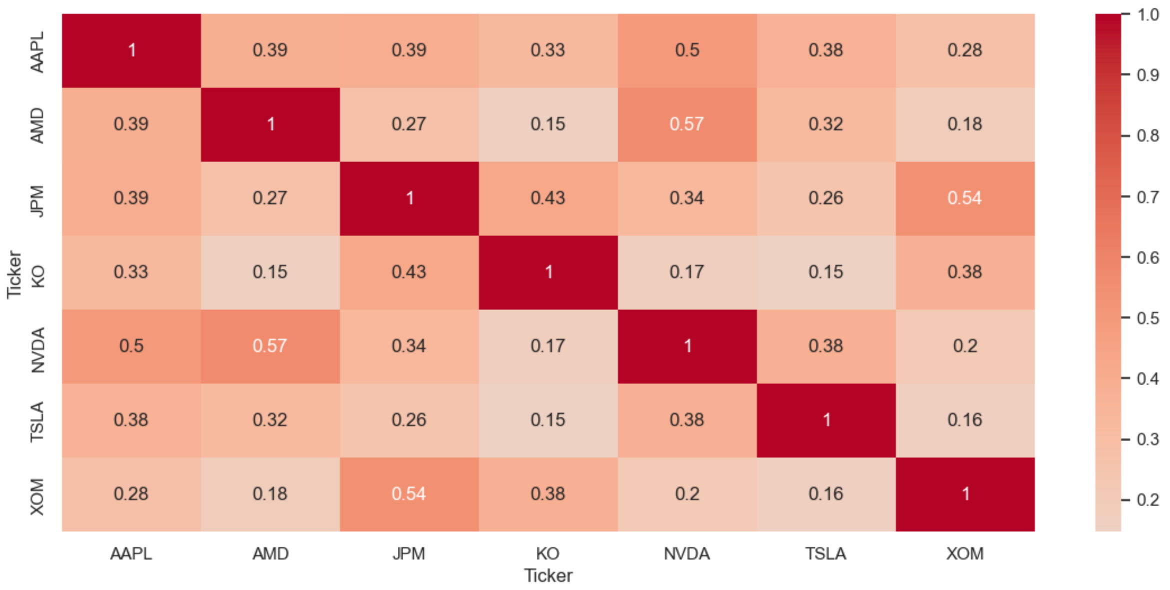
Before performing regression analysis, we conducted a thorough exploratory analysis to understand the behaviour and structure of the portfolio and its stocks.
To visualize the volatility trend over time, a rolling 30-day standard deviation for each stock and portfolio was calculated. The portfolio's daily log return exhibits relatively low volatility under normal market condition, with noticeable volatility clustering during major market events. The sharp drawdown in early 2020 corresponds the COVID-19 crisis, during which the portfolio experienced it's most extreme negative return. Interestingly, the post-2020 period reflects an elevated volatility regime, likely driven by macroeconomic uncertainty and FED rate tightening cycles. This highlights the portfolio's sensitivity to systematic shocks. This behaviour reinforces the need to attribute portfolio returns to underlying risk factors to isolate true alpha from beta, especially during periods of market stress.



30-Day rolling volatility of stocks shows TSLA, AMD and NVDA are consistently the most volatile while Coca-Cola(KO) being the most stable one (classic low volatility consumer staple). All stocks spike during 2020 (COVID-19) and 2022 (Fed rate tightening).



The correlation heatmap shows partial diversification, with moderately correlated pairs like NVDA-AMD and KO acting as defensive anchor. This correlation structure impacts factor exposures, especially size and momentum(not included in the model yet).



5. Factor Regression Using the Fama-French 5-Factor Model

The core of this thesis involves regressing the daily excess portfolio returns on the five Fama-French factors. The detailed output and insights for these five factors are given below.

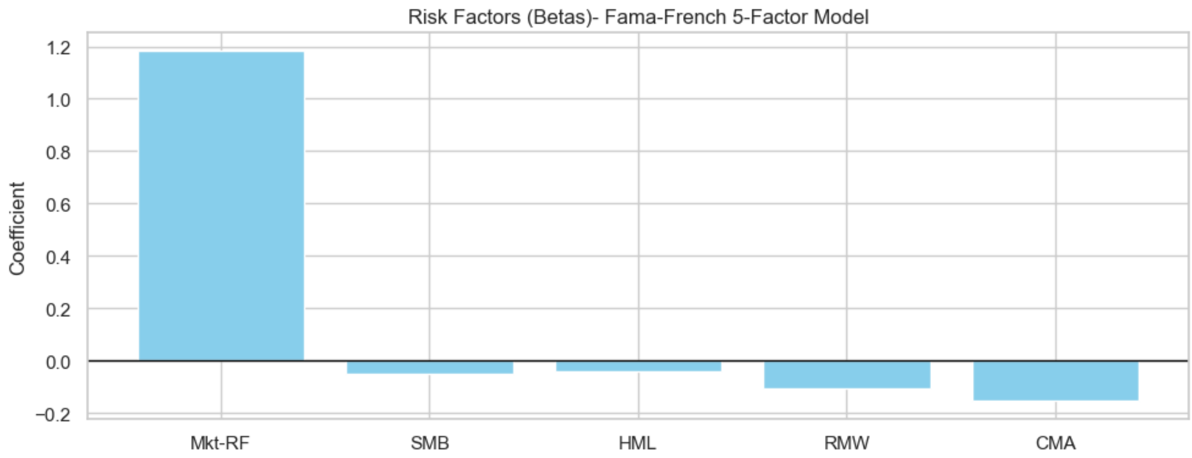
The independent variable (X) i.e. daily excess return = daily portfolio return – daily risk free rate

The model was estimated using Ordinary Least Squares (OLS). The resulting regression showed an R-squared of 0.745, indicating that approximately 74.5% of the variation in excess portfolio returns is explained by the five factors.

The intercept alpha 0.03% (~7.7% annualized) was statistically significant with p-value 0.027, suggesting the presence of a small, persistent component of return not captured by the model.

The portfolio demonstrated a high beta to the market factor (1.18), implying greater sensitivity to broad market movements.

The negative coefficients for SMB, HML, RMW, and CMA is in-line with the portfolio construction, as our portfolio is consist of large-cap, tech-heavy, growth oriented firm like NVDA, TSLA and AMD.



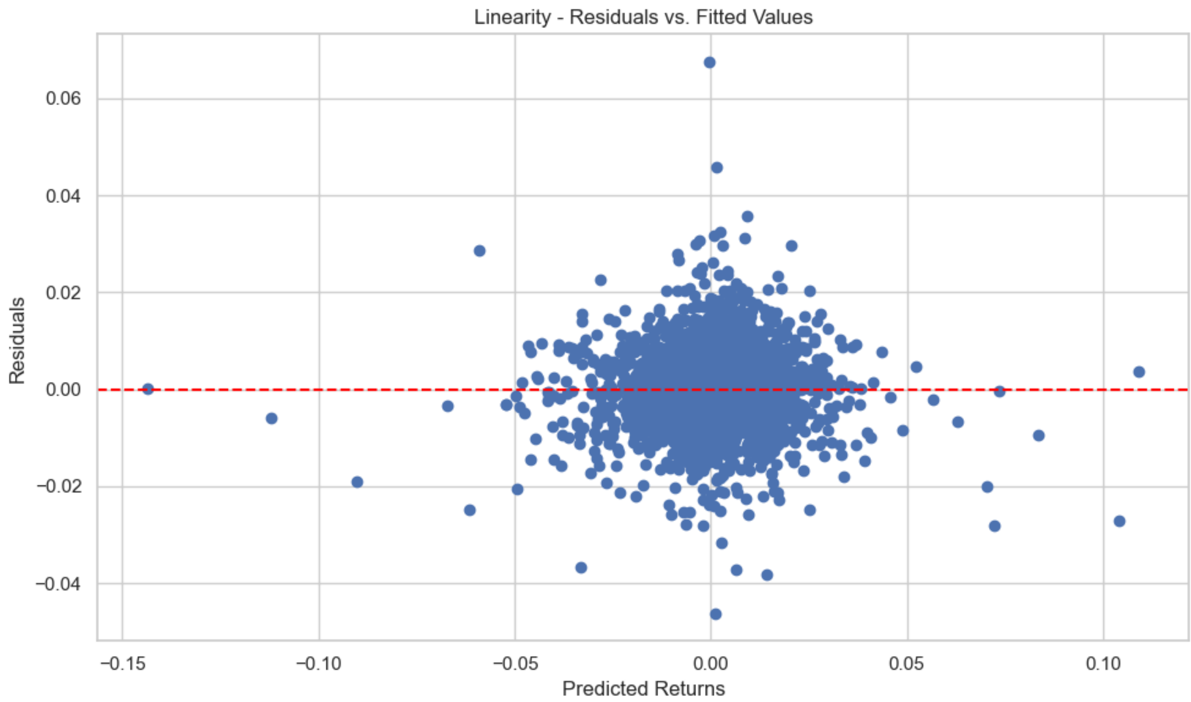
Model Fitting: The actual and fitted values cluttered around the 45-degree line suggest a strong model fit.

6. Model Diagnostics and Assumption Checks

To ensure the validity of the OLS results, a series of diagnostic tests were conducted to check the OLS assumptions.

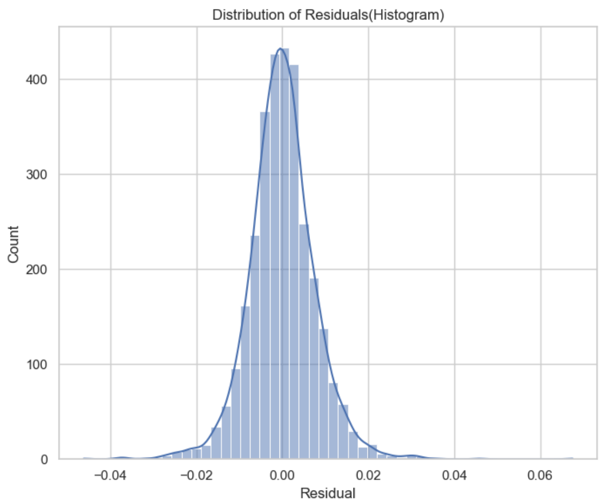
6.1 Linearity

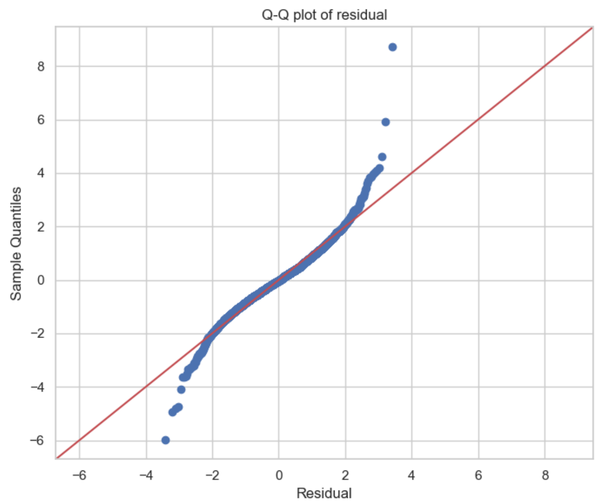
The residuals-versus-fitted values plot looks like a cloud of points cantered around zero suggesting the linearity of the relationship and showed no evident pattern, indicating that the linearity assumption holds.



6.2 Normality of Residuals

Although the histogram and Q-Q plot suggest approximate normality, the Jarque-Bera test ($p<0.001$) indicates a statistically significant deviation from normality due to fat tails (kurtosis = 7.21) and mild skewness. These characteristics are consistent with empirical financial return data. Given the large sample size (> 3000 observations), OLS inference remains reliable under the central limit theorem.



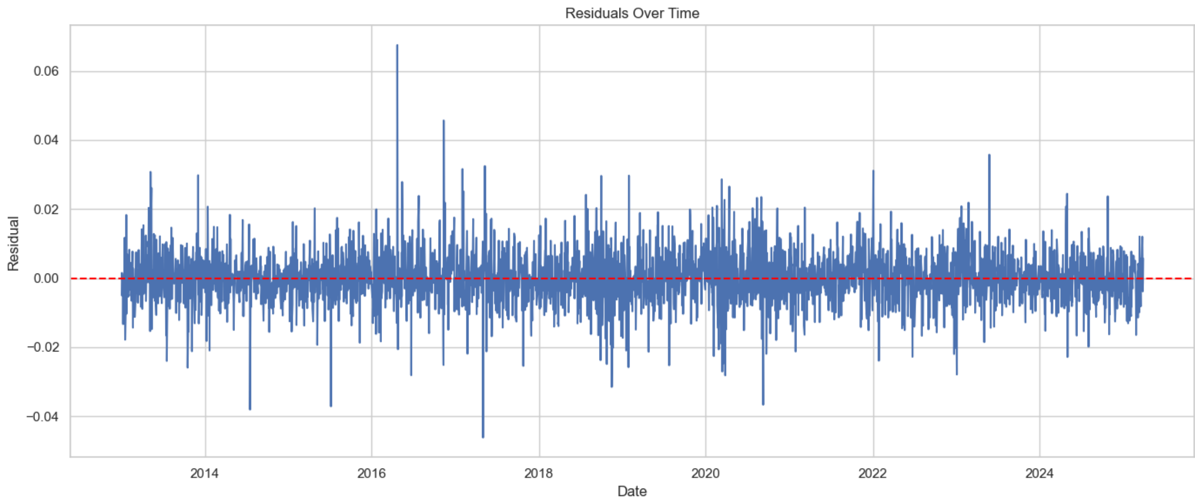


6.3 Zero Mean Residual

The mean of residuals was 0.0009, effectively close to zero, satisfying the assumption of unbiased errors.

6.4 Homoscedasticity (Constant Variance)

The Breusch-Pagan test for heteroskedasticity yielded a p-value of 0.72 > 0.05, supporting the assumption of constant variance in the residuals. The Durbin-Watson statistic was approximately 2.06, suggesting no autocorrelation in residuals.



7. Conclusion and Further Scope

This study successfully applied the Fama-French 5-Factor model to explain the return behaviour of an equally weighted U.S. equity portfolio. The high R-squared value and significant factor loadings affirm the model's explanatory power. The presence of a statistically significant alpha suggests that additional sources of return, such as momentum, may be worth exploring.

The portfolio's factor profile reflects a high-growth, high-beta strategy tilted toward aggressive investment and speculative companies. These results are consistent with the characteristics of several constituent stocks and provide insight into the risk-return profile of such a portfolio.

Future work will focus on enhancing the model by incorporating the momentum factor (Carhart 4-factor or Fama-French 6-factor model), applying rolling regressions to explore the time variation of factor exposures, and experimenting with custom-weighted portfolios to observe changes in factor sensitivity. Robust estimation techniques such as Generalized Least Squares (GLS) or Newey-West standard errors may also be applied to further validate statistical inference in the presence of non-normality or heteroskedasticity.

This thesis demonstrates the value of quantitative attribution methods in portfolio analysis and lays the groundwork for more sophisticated modeling in future studies.