Momentum, Volatility, and Volume Factors in U.S. Stock Returns

Pedro Kawamura, Carson Suriano, Polly Baugh, Chase Renick

Abstract

Technical analysis of stocks have widely been used in portfolio optimization to maximize an investor's returns. This paper investigates the statistical significance of three primary technical factors, Momentum, Volatility, and Volume, as style factors in U.S. stock returns. We analyze a universe of the top 50 S&P 500 stocks (as of January 2021) using weekly data from January 2021 to December 2024.

Our three-part methodology begins by using a weekly cross-sectional Multiple Linear Regression (MLR). This process runs a new regression for each week, using all 50 stocks, to measure the unique "weekly payout" earned by each factor, resulting in three new factor return time-series. We then analyze these new payout series: first, we use Time Series Regression to test if the average weekly payout for any factor is statistically greater than zero. Finally, we apply Simple Exponential Smoothing to forecast future factor returns and evaluate predictability. This analysis allows us to determine if these common technical factors offer a persistent, non-random return premium.

Dataset

Our analysis will use weekly stock market data sourced from the `yfinance` Python library. The dataset spans from January 1, 2021 to December 31, 2024. Our stock selection consists of the top 50 companies in the S&P 500, selected based on their weight as of January 2021. This eliminates survivorship bias by avoiding selection of successful stocks. For each of these 50 stocks, we will use the weekly Adjusted Close price and Volume as the raw data for constructing our factors.

Method

The paper will be constructed as a three-part analysis. We will first use a Multiple Linear Regression to construct our factors and their contribution to stock returns. Then, we apply a Time Series Regression to analyze statistical significance of the factor's average return and then use Simple Exponential Smoothing models to determine predictability of our model.

1. Multiple Linear Regression
   - Defining our Factors: In deciding our factors, it was important that they be distinct, non-overlapping indicators measuring momentum, volatility, and volume. We will use the `pandas_ta` library to calculate factor exposure for each stock.
     - **Momentum (`MOM_Factor`)**: We will use the 26-week Rate of Change (ROC) indicator, lagged by 4 weeks. The 26-week lookback captures intermediate-term price trends, while the 4-week lag avoids the short-term reversal effect, where stocks reverse their performance from the most recent month due to investor overreaction.

- ○ **Volatility (`BBW_Factor`)**: We will use the 10-week Bollinger Band Width (BBW) indicator. Bollinger Bands are lines plotted two standard deviations above and below a standard 10-week moving average, where the width measures the distance between these bands, representing a stock's recent price volatility.
    - ○ **Volume (`VOL_Factor`)**: We will use the 20-week Relative Volume indicator that compares the most recent week's trading volume to its 20 week average. The 20-week average established a robust 'normal', volume size to ensure the factor only captures significant spikes in market attention.
  - For each week `t` in our sample, we will run a separate MLR on our stock universe on each factor lagged by one week to avoid look-ahead bias. The model for each week is given by the following formulation:

$$Return_{i,t} = \alpha + \beta_{MOM}(MOM_{i,t-1}) + \beta_{BBW}(BBW_{i,t-1}) + \beta_{VOL}(VOL_{i,t-1}) + \varepsilon_{i,t}$$

  - We will test for multicollinearity between our factors by calculating the VIF.
  - The coefficients, β_{MOM}, β_{BBW}, β_{VOL}, each representing the return premium of that factor in that week from each of the weekly regressions will be used to create our factor return time series.
2. Time Series Regression for Factor Significance
   - The time series regression will test the properties of the three factor return series built in the MLR.
     - ○ Average Return Test: We will run a regression for each factor: `Factor_Return(t)` $= \alpha + \varepsilon_t$. Performing a two-tailed t-test at significance level $\alpha = 0.05$ on the intercept will determine if the factor's average weekly return is statistically significant.
     - ○ Autocorrelation Test: Testing for each factor's momentum, we will run an AR(1) model to see if performance is predictable from this formulation: `Factor_Return(t)` $= \alpha + \beta_1(Factor\_return(t-1)) + \varepsilon_t$. This model will indicate that past factor returns can predict future factor returns.
3. Simple Exponential Smoothing (SES)
   - We will test for factor predictability and if they can be forecasted.
   - The SES model updates the level estimate at each time period can be modeled with: $I_t = \alpha y_t + (1-\alpha)I_{t-1}$. We will
     - ○ Split the data into weeks into a training set and the test set.
     - ○ Use the statsmodels library in Python to find the optimal α.
     - ○ Produce one-step-ahead forecasts for the test period using the fitted model
     - ○ Calculate forecast performance metrics with MAD, MSE, and MAPE.
   - If the SES results in high forecast accuracy, it can suggest factor returns are predictable and could support a profitable trading strategy.