

Factor Based Investment Strategy

Team 4

2025-12-05

Table of contents I

- 1 Introduction
- 2 Six Factors: Definitions and Data Features
- 3 Portfolio Construction
- 4 Our Own Factor Construction
- 5 Non-Standard Error Analysis
- 6 Closing

Section 1

Introduction

Motivation

- Fama–French factors are a cornerstone of empirical asset pricing.
- Academic factors are widely used in:
 - Explaining the cross-section of expected returns.
 - Designing systematic equity strategies.
- Our project goals:
 - Understand the **six key factors** and their data features.
 - Construct **five portfolios** based on these factors.
 - Replicate Fama–French style factors from CRSP and Compustat.
 - Evaluate factor regressions using **non-standard (robust) errors**.

Section 2

Six Factors: Definitions and Data Features

The Six Factors

We focus on six standard equity factors:

- **Market (MKT)**: excess return on the value-weighted market portfolio.
- **Size (SMB: small minus big)**
 - Long small-cap stocks, short large-cap stocks.
- **Value (HML: high minus low)**
 - Long high book-to-market stocks, short low book-to-market stocks.
- **Profitability (RMW: robust minus weak)**
 - Long firms with high operating profitability, short firms with low profitability.
- **Investment (CMA: conservative minus aggressive)**
 - Long firms with low investment (conservative), short firms with high investment (aggressive).
- **Momentum (MOM: winners minus losers)**
 - Long past winners, short past losers based on prior 11-month returns (skipping the most recent month).

Investment Strategy Based on Factors

Factor Investing Strategies

Momentum Factor
Investing in assets exhibiting strong recent performance

Growth Factor
Investing based on potential earnings growth and future expansion.

Value Factor
Investing based on price-to-earnings (P/E) ratio or price-to-book (P/B) ratio.

Quality Factor
Investing in companies with strong financials, stable earnings and low debt

Multi-Factor
Strategies combining multiple factors to achieve diversification and capture sources of returns

Dividend Yield Factor
Investing based on higher dividend yield



Data Sources

CRSP Monthly Data

- Returns (ret) and excess returns (ret_excess)
- Market equity and lagged market equity (mktcap, mktcap_lag)
- Exchange filter (NYSE used for breakpoint calculations)

Compustat Accounting Data

- Book equity (be)
- Operating profitability (op)
- Investment (inv)

Momentum Data

- Constructed within CRSP using the standard **12–2 momentum rule** (past 12-month return excluding the most recent 1–2 months)

Basic Filtering Rules

- Include U.S. common stocks with standard CRSP share codes (e.g., 10 and 11).
- Exclude extreme penny stocks and obvious data errors.
- Use lagged variables to avoid look-ahead bias:
 - Market cap lagged by one month.
 - Accounting variables lagged in line with Fama–French conventions.
- Winsorize extreme returns to reduce the impact of outliers.

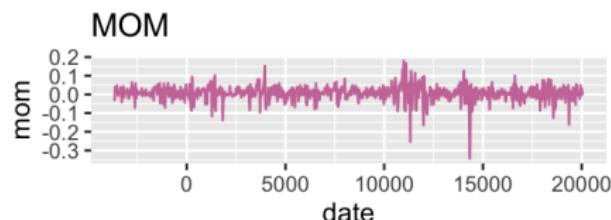
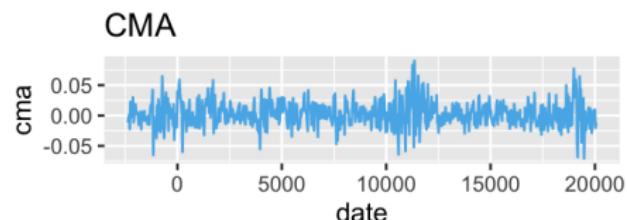
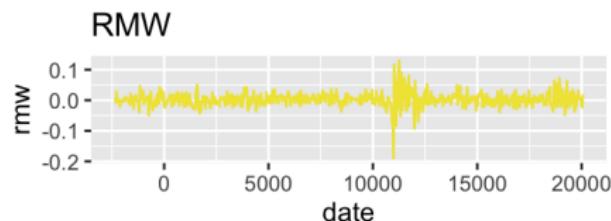
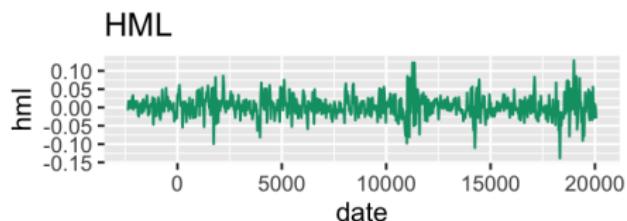
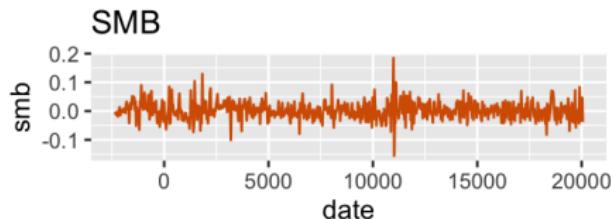
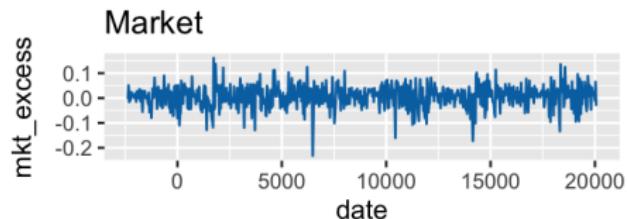
Summary Statistics of Six Factors

Factor	Mean	SD
MKT	0.00587	0.04476
SMB	0.00193	0.03036
HML	0.00282	0.02971
RMW	0.00282	0.02218
CMA	0.00251	0.02057
MOM	0.00623	0.04123

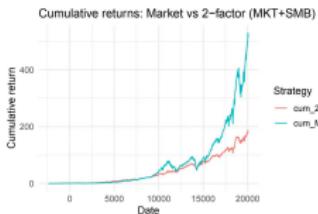
- All six factors have positive average returns.
- The market factor earns about 0.59% per month; the others range from roughly 0.19% to 0.62% per month. So each factor delivers a positive long-run premium.
- The risk profiles are different.
- MKT and MOM have relatively high volatility, around 4–4.5% per month – they are more “aggressive”.
- RMW and CMA have lower standard deviations, about 2–2.2%, which fits their more defensive, quality- and investment-based nature.

Factor Return Characteristics

Monthly Returns of Six Common Factors



CAPM vs. Multi-Factor



Section 3

Portfolio Construction

Overview of Five Portfolios

We build five portfolios using factor information:

- ① **Market portfolio**
- ② **Fixed-weight factor portfolio**
- ③ **Equal-weight factor portfolio**
- ④ **Risk parity factor portfolio**
- ⑤ **Mean–variance factor portfolio**

Each portfolio combines the six factors in a different way, with different risk–return implications.

3.1 Market Portfolio

- Benchmark: value-weighted U.S. stock market.
- Monthly excess return: market return minus risk-free rate.
- Serves as:
 - A baseline for performance comparison.
 - A simple one-factor model (CAPM) in the regressions.

3.2 Fixed-Weight Factor Portfolio

Construction idea:

- Use a balanced allocation across all six factors rather than overweighting any single one.
 - Assign 25% to the Market for core equity exposure.
 - Allocate the remaining 75% evenly (15% each) across the five style factors: Size (SMB), Value (HML), Profitability (RMW), Investment (CMA), and Momentum (MOM).
- The weights are constant over time to reflect a simple, transparent, rules-based multi-factor strategy.

This balanced setup avoids relying too heavily on any single factor and helps capture several different sources of return. Spreading the weights evenly also makes the portfolio more stable than concentrating everything in one factor.

3.3 Equal-Weighted Factor Portfolio

- Straightforward construction: assign the same weight to each factor.
- Monthly return is the simple average of the six factor returns.
- Advantages:
 - No parameter estimation or optimization needed.
 - Naturally diversified and less sensitive to errors in any single factor.

3.4 Risk Parity Factor Portfolio

- Objective: equalize **risk contributions** of each factor.
- Steps:
 - Estimate factor volatilities from historical data.
 - Assign weights inversely proportional to volatility:
 - More weight on lower-volatility factors.
 - Less weight on higher-volatility factors.
- Intuition:
 - Each factor contributes roughly the same amount to total portfolio risk.

3.5 Mean–Variance Factor Portfolio

- Classical Markowitz approach applied to factor returns.
- Steps:
 - Estimate mean vector and covariance matrix of factor returns.
 - Solve for portfolio weights that maximize the Sharpe ratio:
 - Subject to full investment and possibly no-short or other constraints.
- Pros:
 - Theoretically optimal under strong assumptions.
- Cons:
 - Sensitive to estimation error, especially in expected returns.

3.6 Portfolio Performance Comparison

For all five portfolios, we compare the annualized performance:

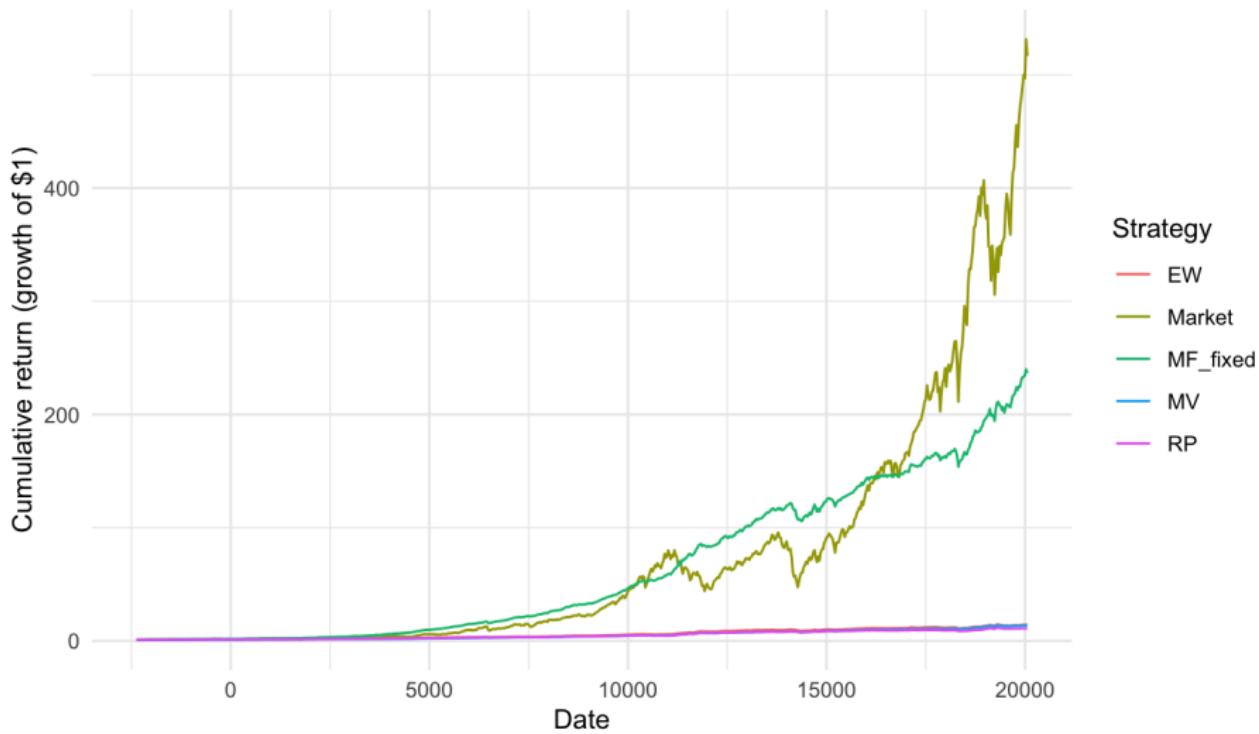
Strategy	Ann. Mean	Ann. SD	Sharpe
Market	0.1140	0.1546	0.7379
Multi-factor (fixed)	0.0902	0.0455	1.9853
Equal-weight	0.0439	0.0401	1.0956
Risk-parity	0.0397	0.0374	1.0611
Mean-variance	0.0432	0.0363	1.1906

Key observations:

- **Multi-factor (fixed)** has the highest Sharpe ratio.
- **Risk-parity** and **mean-variance** have the lowest drawdowns.
- Factor portfolios are **less volatile** and offer **better risk-adjusted returns** than the market.

3.7 Cumulative Return Comparison

Cumulative Returns: Market vs Multi-factor Portfolios (incl. PPP)



Section 4

Our Own Factor Construction

4.1 Constructing Firm Characteristics

Characteristic Definitions

- **Size (ME):** June market capitalization from CRSP
- **Book-to-Market (BM):** Book equity from Compustat paired with December CRSP market equity
- **Profitability (OP):** Operating profitability measure (op)
- **Investment (INV):** Asset growth (inv)
- **Momentum (12–2):** Cumulative return from month t–12 to t–2 (skipping the most recent month)

4.2 Portfolio Formation Using NYSE Breakpoints

Breakpoint Rules

- **Size:** 50% NYSE median → Small vs. Big
- **BM, OP, INV:** 30% / 70% NYSE breakpoints → Low / Neutral / High
- **Momentum:** 30% / 70% (all stocks) → Loser / Neutral / Winner

Timing Rules

- All accounting variables are **lagged** to avoid look-ahead bias.
- Sorting dates are fixed at **July 1** each year, following the Fama–French methodology.

4.3 Factor Construction (SMB, HML, RMW, CMA)

- Form **value-weighted portfolio returns** using lagged market equity.
- Build portfolios sorted on **Size, Book-to-Market, Profitability, and Investment**.
- Each month, compute returns for all characteristic portfolios (e.g., Size \times BM).

Construct long–short factors using portfolio spreads:

- **HML** = High BM – Low BM
- **RMW** = Robust – Weak (high OP minus low OP)
- **CMA** = Conservative – Aggressive (low INV minus high INV)
- **SMB** = Small – Big (averaged across all sorts)

4.4 Factor Construction (Momentum)

- Compute each stock's **12–2 past return** using rolling cumulative returns.
- Form monthly cross-sectional **30/70 breakpoints**.
- Assign stocks into **Winners** (top 30%) and **Loser**s (bottom 30%).
- Compute **value-weighted** returns for Winner and Loser portfolios.
- **MOM factor** = Winner portfolio – Loser portfolio.

Summary of Replicated Factor Statistics

Factor	Mean	SD
SMB (smb_replicated)	0.002196	0.03126
HML (hml_replicated)	0.002277	0.02906
RMW (rmw_replicated)	0.002905	0.02253
CMA (cma_replicated)	0.002097	0.02088
MOM (mom_replicated)	0.003954	0.05795

- All factors show positive mean returns, consistent with FF results.
- Momentum has the highest return and the highest volatility.
- RMW and CMA are the most stable, with the lowest volatility.
- Overall, the replicated factors behave similarly to the official FF factors.

Regression Results (Replication Accuracy)

Factor	Alpha	Beta	R ²	t(Beta)	p(Beta)
SMB	-0.0002	0.9593	0.9868	234.82	0.00e+00
HML	0.0006	0.9804	0.9315	100.05	0.00e+00
RMW	0.0000	0.9495	0.9416	108.94	0.00e+00
CMA	0.0005	0.9551	0.9493	117.33	0.00e+00
MOM	0.0034	0.6615	0.8250	58.46	1.29e-276

- Betas close to 1 show strong replication accuracy for SMB, HML, RMW, CMA.
- MOM has lower beta but still tracks the FF series reasonably well.
- Extremely high t-statistics and near-zero p-values confirm statistical significance.
- High R² values indicate our factors explain most variation in the official FF factors.

Section 5

Non-Standard Error Analysis

5.1 Non-Standard Errors in Financial Time Series

Standard OLS assumption:

Returns satisfy

$$\varepsilon_t \sim \text{i.i.d. } (0, \sigma^2)$$

Why this fails in financial data:

- Returns show **heteroskedasticity** — volatility changes over time.
- Returns show **autocorrelation**, especially at monthly frequencies.
- As a result, OLS **underestimates** the true standard errors.

Non-standard errors are corrected standard errors that **do not rely on the i.i.d. assumption**. They adjust for heteroskedasticity and autocorrelation, providing more reliable inference in financial regressions.

5.2 Why Do We Need Non-Standard Errors in Factor Models?

We run factor replication regressions of the form:

$$\text{FF}_t = \alpha + \beta \cdot \text{OwnFactor}_t + \varepsilon_t$$

OLS standard errors assume:

$$\text{Var}(\varepsilon_t) = \sigma^2, \quad \text{Cov}(\varepsilon_t, \varepsilon_{t-k}) = 0$$

But financial data violates both assumptions:

- Volatility clustering (heteroskedasticity / ARCH effects)
- Serial correlation in factor returns from overlapping signals and persistent economic shocks

Impact on inference leads to **inflated t-statistics** and overstated significance:

$$\text{SE}_{\text{OLS}} < \text{SE}_{\text{true}} \quad (\text{downward biased})$$

5.3 What Newey–West HAC Standard Errors Do

Newey–West estimates the variance as:

$$\widehat{\text{Var}}(\hat{\beta}) = (X'X)^{-1} \left(\sum_{k=-L}^L w_k \Gamma_k \right) (X'X)^{-1}$$

Where:

$$\Gamma_k = \sum_t \varepsilon_t \varepsilon_{t-k} X_t X_{t-k} \quad (\text{lag-}k \text{ autocovariance})$$

$$w_k = 1 - \frac{|k|}{L+1} \quad (\text{Bartlett weights})$$

We use:

$$L = 6 \quad (\text{typical for monthly data})$$

- Includes covariance between residuals at different lags
- Smoothly downweights higher-order autocorrelations
- Produces *larger and more realistic* standard errors
- Valid under both **heteroskedasticity** and **autocorrelation**

5.4 Newey–West Regression Results

Factor	Beta	t_NW	p_NW	R ²
SMB	0.9593	135.3070	0	0.9868
HML	0.9804	62.4367	0	0.9315
RMW	0.9495	71.8690	0	0.9416
CMA	0.9551	94.7485	0	0.9493

- Betas close to 1 confirm accurate replication of Fama–French factors.
- Newey–West t-statistics remain very large even after correcting for autocorrelation.
- p-values are effectively zero for all factors.
- High R² values (>0.93) indicate our replicated factors track the true series extremely well.

Section 6

Closing

Main Takeaways

- **Factor models are replicable**

Our constructed factors closely match the official Fama–French series.

- **Multi-factor portfolios outperform the market**

Higher Sharpe ratios achieved through diversification across multiple factor premiums.

- **Econometric testing confirms robustness**

Newey–West standard errors show the results are not driven by autocorrelation or heteroskedasticity.

- **Data engineering is crucial**

Precise date alignment, portfolio sorting rules, and CRSP–Compustat linking materially affect the results. This is an important lesson for empirical finance.