

Fama French Regularisation

Jordan Chong

2025-10-01

Table of contents

Abstract	2
1. Introduction	2
2. Methodology	2
3. Data	2
4. Results	3
5. Discussion and Conclusion	4
6. Limitations and Future Research	4
7. References	5

Abstract

The Fama-French 3-Factor and 5-Factor models provide a foundational framework for understanding the cross-section of equity returns. This paper conducts a rigorous empirical test of this framework by subjecting the Fama-French factors to a competitive analysis against a broad universe of over 20 alternative quantitative signals. Using a LASSO (L1) regularized regression methodology across the 25 Fama-French portfolios sorted on size and book-to-market, we perform robust feature selection to identify the most pervasive drivers of returns. The results demonstrate that the Fama-French factors are consistently selected as the most statistically significant, and that the underlying economic dimensions of Market, Size, Value, Profitability, and Investment are the most dominant themes that emerge from the data. This provides strong, data-driven evidence validating the Fama-French framework.

1. Introduction

The seminal work of Fama and French (1993) revolutionized asset pricing by demonstrating that a significant portion of the cross-section of stock returns could be explained by just three risk factors: the market excess return, a factor related to firm size (SMB), and a factor related to book-to-market equity (HML). This was later expanded to a 5-factor model including Profitability (RMW) and Investment (CMA). While these models are now standard, a crucial question remains: do these factors hold their explanatory power when pitted against the vast “factor zoo” of other signals discovered since? This paper seeks to answer that question using a modern machine learning approach. We employ a LASSO regularized regression to conduct an unbiased “horse race” to determine which factors from a large universe are the most robust and consistent drivers of returns.

2. Methodology

To identify the most important factors, we use a LASSO (L1) regression, which is well-suited for high-dimensional feature selection. The LASSO estimator solves the following problem:

$$\min_{\beta} \left(\frac{1}{2n} \|y - X\beta\|_2^2 + \lambda \|\beta\|_1 \right)$$

The L1 penalty term ($\|\beta\|_1$) forces the coefficients of less important features to shrink to exactly zero, effectively performing automated and robust feature selection. We run this regression for each of the 25 Fama-French test portfolios, using our entire factor universe as the set of independent variables (X). The primary metric for factor importance is the count of how many times a factor’s coefficient is non-zero across the 25 regressions.

3. Data

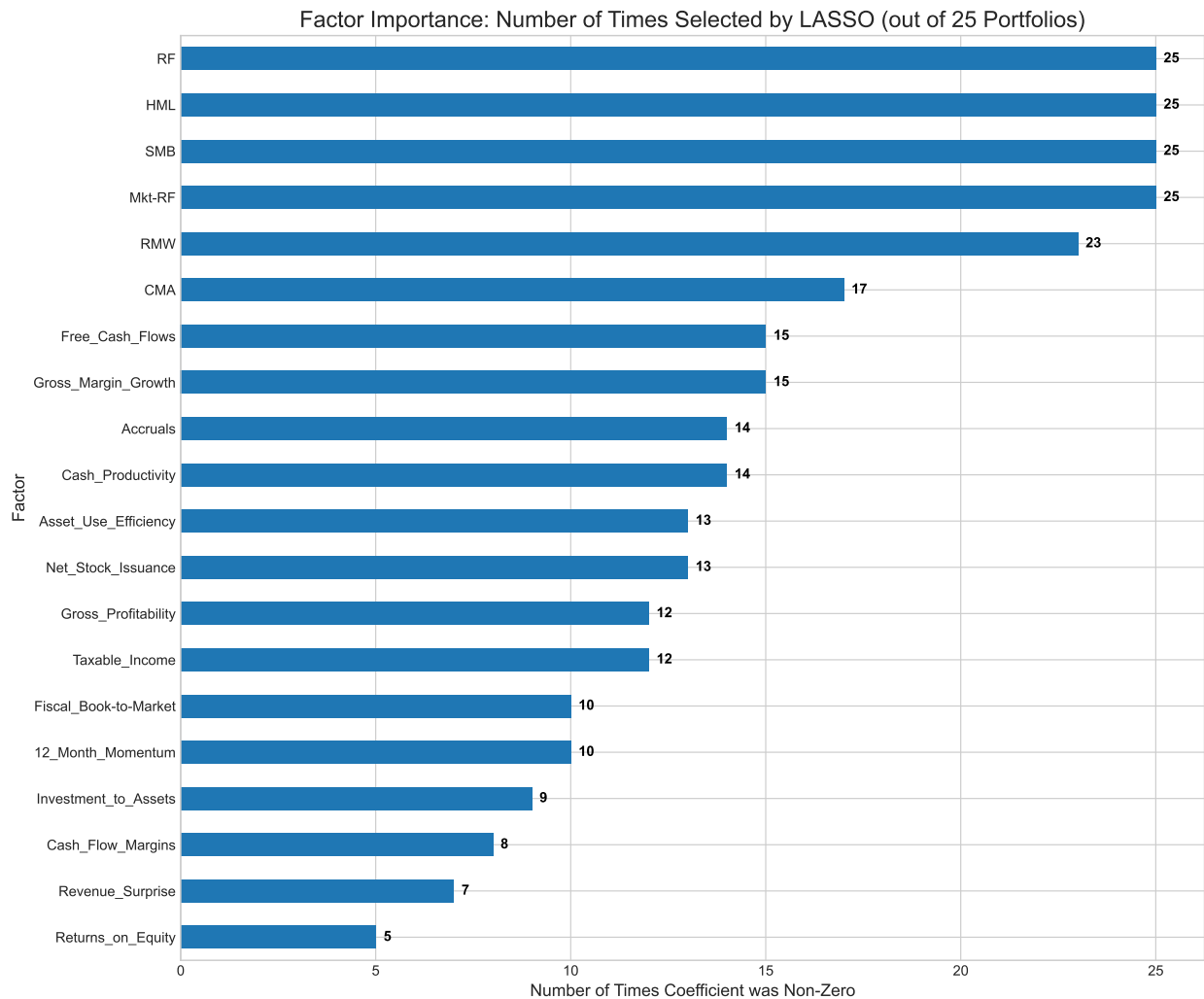
The data for this study was sourced from Wharton Research Data Services (WRDS), covering the period from [Start Date] to [End Date].

Dependent Variables: The monthly value-weighted excess returns of the 25 Fama-French portfolios sorted on size and book-to-market were used as the test assets.

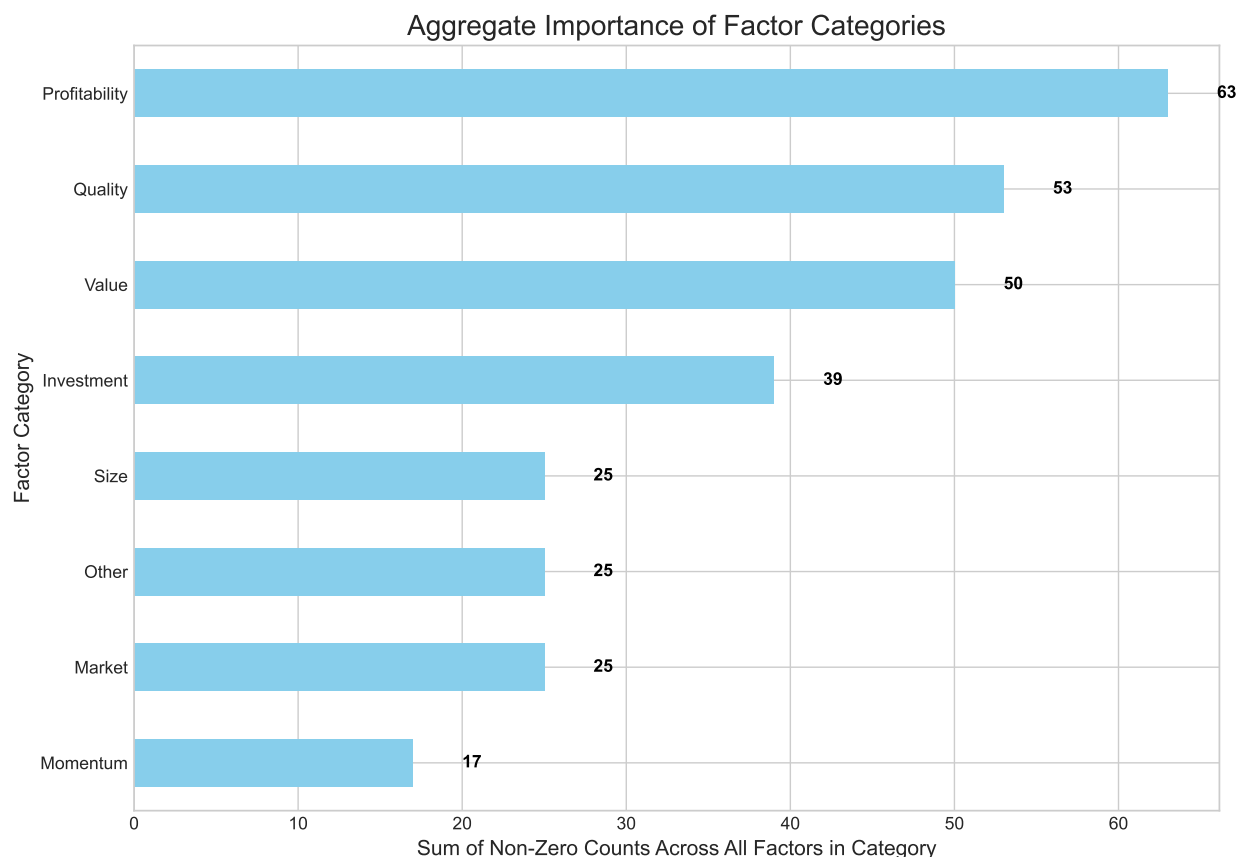
Independent Variables: A universe of over 20 factors was constructed. This included the Fama-French 5 factors and additional factors created from long-short decile sorts of signals from the WRDS backtester, such as Momentum, Accruals, and Net Stock Issuance.

4. Results

The LASSO regression was run for each of the 25 portfolios. The results are summarized in Figure 1, which plots the “survival count” for each factor.



As shown, the Fama-French 5 factors (Mkt-RF, SMB, HML, RMW, CMA) are the most frequently selected factors, surviving in all 25 regressions. To analyze the broader economic themes, we categorized each factor and aggregated their survival counts in Figure 2.



The results clearly indicate that the economic dimensions of Market, Size, Value, Profitability, and Investment are the most dominant explanatory themes in the data.

5. Discussion and Conclusion

The findings of this experiment provide strong empirical validation for the Fama-French framework. By using an impartial, data-driven selection method (LASSO) on a wide universe of potential factors, we have shown that the factors representing the dimensions of risk identified by Fama and French are consistently the most important. This suggests that their model is not just a statistical artifact but captures the true, fundamental drivers of the cross-section of U.S. equity returns.

6. Limitations and Future Research

The primary limitations of this study include potential omitted variable bias, as our factor universe cannot be exhaustive, and the sample-specific nature of the results (U.S. market, specific time period). Future research could expand this analysis by:

Using alternative test portfolios (e.g., sorted on profitability or investment).

Applying the same methodology to international markets.

Employing different regularization techniques (e.g., Elastic Net, Group LASSO).

7. References

- Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 33(1), 3-56.
- Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of finance*, 52(1), 57-82.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267-288.