Quantitative Finance Efficiency, Predictability & Portfolio

December 2, 2024

WARWICK BUSINESS SCHOOL
THE UNIVERSITY OF WARWICK

Submitted in partial fulfilment of the requirements for
Quantitative Finance

Contents

Researchers have continually faced challenges in identifying appropriate proxies for earnings and investments. Part 1 of this report will first provide a summary of the 2015 article "*A five-factor asset pricing model*" by Eugene F. Fama ("Fama") and Kenneth R. French ("French"). Next, we will evaluate the proposed investment strategies in relation to the Efficient Market Hypothesis ("EMH") and assess the discussion regarding the applicability and limitations of the five-factor model. Finally, in part 2, we will consider whether the six-factor model effectively explains time-series and cross-sectional variations in excess stock returns through a series of empirical tests.

Part 1

Summary of the Fama and French (2015) paper

Fama and French extended the three-factor model by introducing two additional factors — profitability and investment — to create the five-factor model. These new factors address the limitations of the three-factor model, particularly its inability to fully explain average returns for small firms with low profitability and high investment. By incorporating these factors, the five-factor model provides a more comprehensive explanation of stock returns.

This paper builds on the Dividend Discount Model, which estimates stock prices based on expectations of future dividends and discount rates. Fama and French explain how profitability and investment factors influence stock returns. Profitability (RMW) captures the return differential between firms with high and low operating profitability, while investment (CMA) measures the spread between firms with conservative and aggressive investment strategies. Empirical evidence shows that firms with higher profitability and lower investment levels tend to exhibit higher average returns, reflecting financial stability and efficient capital allocation.

To test their model, Fama and French employ an extensive dataset covering US stocks listed on NASDAQ, NYSE, and AMEX from July 1963 to December 2013. Their methodology included portfolio construction, regression analysis, and statistical testing:

- Portfolio construction:
  a. Construct 25 portfolios sorted by size, book-to-market ratio, profitability, and investment.
  b. Double and triple sorting was used to systematically isolate the effects of each factor, for instance, size-profitability and size-investment.
- Regression analysis:
  a. Build time-series regression of portfolio returns on the five factors, including market risk, size, value, profitability, and investment.
  b. Analysed intercept (α) to determine the extent of unexplained variation in returns, with smaller intercepts indicating better model performance.
- Statistical testing
  a. The Gibbons Ross Shanken ("GRS") test was used to evaluate whether the model's intercepts are jointly zero, with lower GRS statistics indicating higher explanatory power.

The five-factor model significantly enhances the explanation of firms' average returns in the three-factor model by incorporating the profitability and investment factors. These findings align with Novy-Marx (2013) and other research, which identified profitability as a critical predictor of stock returns. However, the introduction of the profitability and investment factors reduces the independent role of the value factor (HML). High book-to-market stocks often exhibit low profitability and conservative investment patterns, therefore making the value factor redundant in the presence of RMW and CMA.

To summarise, the five-factor model represents a significant advancement in asset pricing, offering a more nuanced framework for understanding stock returns while highlighting areas where market inefficiencies may persist. While the model develops a more comprehensive framework of asset pricing and performs better than the three-factor model, there are limitations to its usage, which will be further discussed below.

The Efficient Market Hypotheses and the Five-Factor Model

The Efficient Market Hypothesis ("EMH") posits that stock prices fully reflect all available information, making it impossible for any chance of excess returns without taking on additional risk and leaving no room for outperformance.

The five-factor model's relationship with EMH can be analysed as follows:

- Alignment with EMH
    - i. Systematic risk explanation

      The five-factor model enhances our ability to explain variations in average returns compared to earlier models like CAPM and the Fama-French three-factor model. This suggests that it captures the additional systematic risks priced by the market, aligning with the EMH idea that returns are compensation for these risks. However, some residual unexplained variations challenge the hypothesis's applicability in its most potent form.

    - ii. Risk-Return trade-offs

      Factors such as profitability and investment reflect rational risk-return trade-offs. For instance, small-sized firms and value stocks carry higher risks due to financial distress or economic sensitivity. In comparison, high-profitability and low-investment firms tend to represent lower risk, reflecting financial stability. This is consistent with EMH's emphasis on compensating investors for bearing systematic risks.

- Challenges against EMH

    i. Unexplained anomalies:

    The five-factor model fails to fully explain the underperformance of small, high-investment, and low-profitability stocks, suggesting that market inefficiencies exist. These persistent anomalies indicate that certain risks associated with these stocks may not be fully priced in by the market, or that behavioural factors, such as investor overreaction or underreaction, could influence asset prices. This implies that market efficiency may have limits, with some pricing anomalies remaining unexplained by traditional models.

    ii. Residual unexplained variations:

    Despite its improvements, the five-factor model does not completely eliminate unexplained variations in returns, especially for specific subsets like small stocks with high investment and low profitability. This indicates that, while robust, the model cannot fully account for all observed return patterns, raising questions about the assumption of complete market efficiency and suggesting opportunities for further refinement.

To address these challenges and further evaluate market efficiency, introducing an additional factor, such as momentum, may provide insights. Momentum, which captures the tendency for past winners to outperform past losers, reflects behavioural tendencies and might align better with persistent market anomalies. This inclusion into a six-factor model could help explain residual variations and provide insights into persistent anomalies.

Possible Trading Strategies according to Fama and French (2015)

Based on the Fama and French (2015) paper and its discussion of the five-factor model, we propose some potential trading strategies that leverage the insights from their findings. These strategies aim to achieve high returns while managing risk.

1. Invest in high-profitability, low-investment stocks

   The five-factor model highlights that firms with robust profitability and conservative investment tend to outperform due to their financial stability and efficient capital allocation. Investors can screen for stocks with high operating profitability and favour companies with low asset or book equity growth rates. Quarterly or semi-annual rebalancing will be needed to maintain exposure to these characteristics.

2. Factor diversification

   To manage risk, investors can construct multi-factor portfolios that combine exposures to market risk, size, value, profitability, and investment. This reduces the reliance on any single factor's performance. Investors can construct a multi-factor portfolio, equally weighting exposure to size (SMB), value (HML), profitability (RMW), and investment (CMA). If individual stock selection is impractical, ETFs and mutual funds are possible options to track these factors.

3. Momentum overlay

   Although the paper does not focus on momentum, adding a momentum strategy could enhance returns as momentum strategies exploit the tendency for stocks with strong past performance to continue performing well in the short term. This could be done by ranking stocks monthly based on 6- or 12-month past returns, excluding the most recent month to avoid short-term reversals. Investors can then go long on the top decile of momentum stocks and short on the bottom decile.

Conclusion

Fama and French's five-factor model represents a significant evolution in asset pricing, providing a more comprehensive understanding of the factors driving average returns. From our discussion, we find that the five-factor model aligns with the weak and semi-strong forms of the EMH by explaining returns through systematic risk factors, but persistent anomalies – such as small, high-investment, and low-profitability firms – challenge the strongest form of EMH and highlight areas where the markets may not be perfectly efficient.

The Fama and French framework provides valuable guidance for potential trading strategies, such as focusing on high-profitability, low-investment stocks for lower risk, or small-cap and value stocks for higher returns, leveraging systematic risk premia while managing exposure to inefficiencies. Overall, while the model is a valuable tool for both academic research and practical investment strategy, it underscores that no model can fully capture the complexities of financial markets, leaving room for further refinement and exploration.

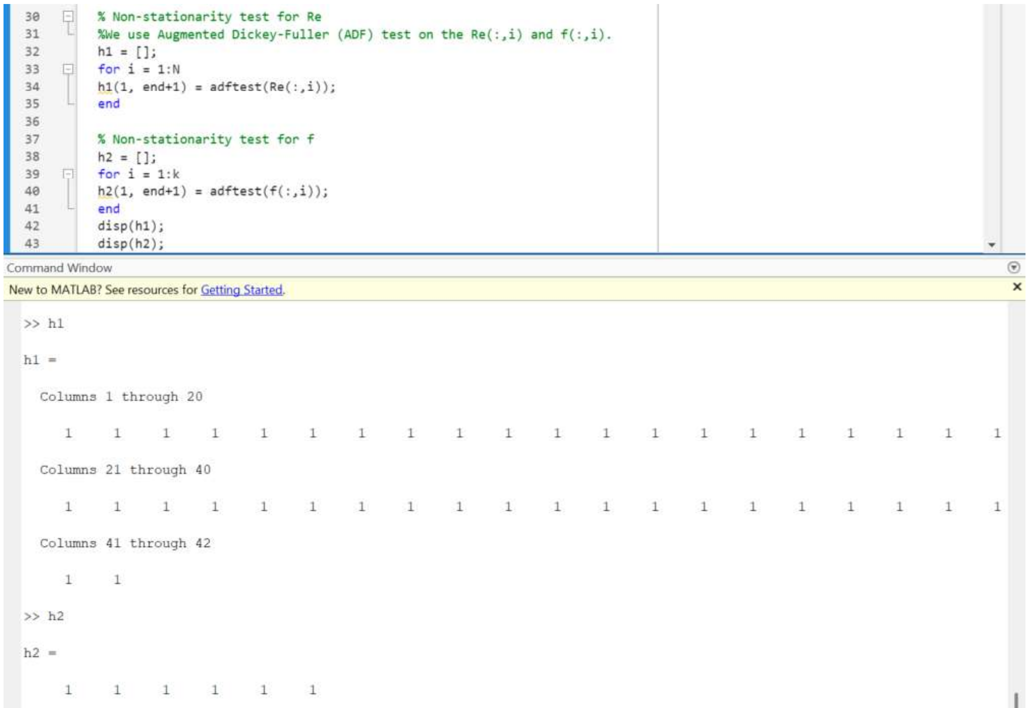
Part 2

The following regression analyses examine the explanatory power of the five-factor model on the excess returns of the 25 Fama-French ("FF") Size-B/M portfolios and 17 industry portfolios, using monthly data from July 1963 until July 2024. (Please refer to **Appendix 7** for the screenshots of all our MATLAB codes)

a) Non-Stationarity Test

In time series analysis, testing for stationarity is to ensure the accuracy and validity of the estimation as non-stationary series can lead to spurious regression results, undermining the reliability and interpretability of the model (Baumohl and Lyocsa, 2009). Therefore, we first introduce the Augmented Dickey-Fuller test ("ADF"), which can deal with the lag of the dependent variable in comparison with Dickey-Fuller, to test for the presence of unit roots on each column of excess return R^e and the factor matrix f. This implies that the statistical properties of the data, such as mean and variance, are constant over time, which is a crucial requirement for reliable econometric modelling.

By rejecting the null hypothesis, the ADF test suggests that the time-series is stationary. As shown in Figure 1 below, the results of h1 and h2 are both 1, implying that the p-value is less than 5% and indicating that we can reject the null hypothesis of non-stationary for each column at a 5% significance level. Therefore, we can conclude that there is strong evidence to support stationarity in both the excess return and the six factors.



```
% Non-stationarity test for Re
%We use Augmented Dickey-Fuller (ADF) test on the Re(:,i) and f(:,i).
h1 = [];
for i = 1:N
h1(1, end+1) = adftest(Re(:,i));
end

% Non-stationarity test for f
h2 = [];
for i = 1:k
h2(1, end+1) = adftest(f(:,i));
end
disp(h1);
disp(h2);
```

Command Window

New to MATLAB? See resources for Getting Started.

```
>> h1

h1 =

  Columns 1 through 20

     1     1     1     1     1     1     1     1     1     1     1     1     1     1     1     1     1     1     1     1

  Columns 21 through 40

     1     1     1     1     1     1     1     1     1     1     1     1     1     1     1     1     1     1     1     1

  Columns 41 through 42

     1     1

>> h2

h2 =

     1     1     1     1     1     1
```

Figure 1: Augmented Dickey-Fuller Test Results

b) Time-series regression R-squared

Using $R^2 = 1 - \frac{RSS}{SSR}$, where RSS is the residual sum of squares, and TSS is the total sum of squares, R^2 explains the portion of the time-series variation in portfolio returns. As calculated and shown in Figure 2 below, an average time-series R^2 of 0.8274 implies that approximately 82.74% of the variation in excess returns across the 42 portfolios is explained by the six-factor model. This result indicates that the model demonstrates substantial explanatory power in capturing the time-series dynamics of portfolio returns, effectively accounting for variance exposure to systematic risk. Moreover, several portfolios exhibit high R^2 values, particularly those exceeding 90%, indicating strong model fit across these portfolios (Please refer to **Appendix 1** for time-series R-squared table), thereby validating its robustness in the time-series analysis of specific portfolios. Nonetheless, the presence of lower R^2 values, such as those around 0.41, may emphasise that there are specific exposures that the model did not capture, potentially omitting non-linear dynamics or tail risk effects. We suggest identifying the specific industries the portfolios belong to, such as high-tech or energy companies, and thereafter conducting a regression on the residuals to observe whether a significant non-linear relationship or omitted variable exists.

```
Average R-squared:
    0.8274
```

Figure 2: Time-series average R-squared

As a further step to investigate the omitted variable, we recommend incorporating external contingencies, such as COVID-19 and human-induced factors, into the analysis. COVID-19 has had varying impacts across industries, inducing panic and creating disparities in asset performance. For instance, the tourism industry in China suffered significant downturns due to travel restrictions, while the pharmaceutical sector performed well due to the experienced increased demand (Wang and Liu, 2022). To further prove that this had a global impact, we introduce the Carnival Cruise Corporation's ("CCL") performance during the pandemic as a representative, as it is the largest global cruise company. CCL's net income was negative from 2020 to 2022 (CCL, 2023) and also experienced a huge drop in prices from USD $50.21 to $8.80 in two months and rebounded to $31.08 in 3 months, and overall, from January 2020 to August 2022, its price decreased of approximately 55% (Yahoo, 2024). Additionally, human behaviours, such as underreaction and overreaction to contingencies (Choi and Hui, 2014), have further shaped the impacts on different industries. These behavioural responses, combined with the unique characteristics of each industry, led to divergent asset return patterns.

To intuitively observe the impacts of contingents or special events, we plot the residuals over time as shown in Figure 3 below, and we observe that around years 1975, 2000, 2015, and 2020, the residuals substantially increased. Apart from that, we also plot the residuals across 42 portfolios (Please refer to **Appendix 2** for the plot of residuals), and we observe that residuals of different portfolios reacted differently to the contingents and special events, suggesting evidence of the presence of other omitted factors. To further test for the impact of special events, we need to conduct significance testing during each specific time window.

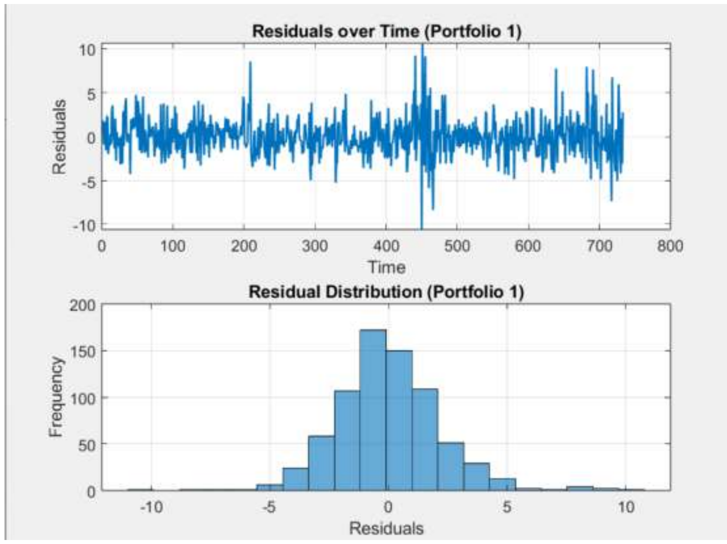


Figure 3: Residuals over time

Moreover, we suggest applying the Principal Component Analysis (“PCA”) to analyse omitted factors more effectively. Under PCA, the original factors are replaced with new estimates under different economic cycles to help evaluate how the six-factor model explains portfolio excess returns in varied periods (Kim, 2008). However, it is challenging to pinpoint when an economic cycle begins or to identify clear indicators that signal the start of a new cycle in advance.

Overall, the six-factor model effectively explains the time-series variations in portfolio returns. Nonetheless, there remains potential for further improvement, particularly in identifying portfolios with lower explanatory power and investigating the return data to understand the occurrence of extreme values and unique features of specific portfolio segments that the current model does not account for.

c) Cross-sectional regression on 25 FF Size-B/M portfolios and 17 industry portfolios

The zero-beta rate refers to the expected return of an asset or portfolio with no systematic risk, meaning its beta is zero. However, unlike the risk-free rate, the zero-beta rate may still be exposed to frictions due to market imperfections. For instance, investors might demand additional compensation for factors such as transaction costs and liquidity premiums (Dickerson et al., 2023). Therefore, the magnitude and direction of the zero-beta rate are consistent with theoretical expectations that it is higher than the average risk-free rate.

According to Fama and French (2015), there are size, value, profitability, and investment effects on the average stock returns and the momentum factor, which only works in portfolios formed on momentum. Given how each factor is constructed, the factor risk premia estimates should also be positive. However, as seen in the numbers shown in Table 1, not all factors meet the expectations. Another point to note, the magnitude of the momentum factor is substantially larger than anticipated, indicating the relatively strong explanatory power in the cross-sectional analysis. Moreover, the negative coefficient of the investment factor suggests that aggressive investment is outperforming, implying that higher investment signals higher future returns, which is a result contrary to theoretical expectations.

Risk Factors	**Gama Estimates (%)**	**t-statistics**
Zero-beta rate	0.6091	2.2028
Market (MKT)	-0.0132	-0.0403
Size (SMB)	0.2039	1.7616
Value (HML)	0.2436	2.1229
Profitability (RMW)	0.1994	1.3083
Investment (CMA)	-0.0377	-0.1830
Momentum (MOM)	0.7271	1.4261

Table 1: Cross-sectional estimates and their t-statistics

(Refer to **Appendix 3** for output screenshot from MATLAB)

While comparing the performances under time-series and cross-section, as Table 2 below shows, the significant performance of the value factor in both tests supports the robustness, reinforcing its role in generating excess returns over the long term. Moreover, the magnitudes of the estimates differ insignificantly except for the MOM and MKT factors in the two tests. The different values of the momentum factor are likely due to the anomalies of short-run momentum and long-run reversal, which suggests the winner portfolios may perform better in the short run but get worse in the long run (Jegadeesh and Titman, 1993).

Six-Factor Model	**Time-series sample means estimates (%)**	**Cross-sectional estimates (%)**
MKT	1.0172	- 0.0132
SMB	0.3885	0.2039
HML	0.1604	0.2436
RMW	0.0927	0.1994
CMA	0.0580	- 0.0377
MOM	- 0.0431	0.7271

Table 2: Comparison between time-series sample mean and cross-sectional estimates

(Refer to **Appendix 4** for output screenshot from MATLAB)

The market factor's insignificance may be due to instability during extreme market conditions, such as financial crises or the COVID-19 pandemic. For the investment factor, we find a smaller difference in magnitudes with a different sign. To address this anomaly, since fiscal policy, such as tax reduction and increase of government spending, can increase the corporate's needs and willingness to invest (Baum and Koester, 2011), we suggest incorporating an interaction term between investment factor and fiscal-related factors:

$$Ri = \alpha + \beta_{MKT} \cdot MKT_i + \beta_{SMB} \cdot SMB_i + \beta_{HML} \cdot HML_i + \beta_{RMW} * RMW_i + \beta_{CMA} * CMA_i + \beta_{MoM} \cdot MoM_i + \beta_{INT} * Interest\ Rate_i + \gamma(Interest\ Rate * CMA_i) + \varepsilon_i$$

The observed discrepancies between time-series and cross-sectional estimates highlight fundamental differences in these approaches. Time-series analysis captures dynamic variations in factor premiums driven by market conditions, whereas cross-sectional methods assume static factor premiums at a given point in time. For example, the time-varying nature of factor premiums introduces dynamics in time-series analysis, while cross-sectional methods cannot fully reflect these changes. According to Shanken (1992), even when the variability of the residuals is unstable, the cluster-robust standard error remains effective for providing consistent estimates of standard residuals. Additionally, the conclusions are not significantly affected by the sample size, further ensuring robustness in the results.

Furthermore, we choose the 5% and 10% significance levels as our standards to judge whether the factors are significant because we are aiming to minimise the Type I error while having enough power to detect true effects. As the t-statistic shown in Table 1, the coefficient of zero-beta rate and HML are statistically and significantly different from zero at the 5% significance level ($c = |-2.0301|$) and the coefficient of the SMB is statistically and significantly different from zero at the 10% significance level ($c = |-1.6896|$). This provides strong cross-sectional evidence that value stocks tend to outperform growth stocks and weak cross-sectional evidence that small firms perform better than large firms. Therefore, we reject the null hypothesis at the 5% and 10% significance level. This may suggest the robustness of value strategies, which can generate excess returns in the long run. However, investors should exercise caution when relying on statistically insignificant factors, such as the MKT and CMA factors, as their coefficients may lack robustness under varying economic conditions. The observed insignificance underscores the need for further investigation into omitted variables or structural changes in market dynamics. (Please refer to **Appendix 5** for critical value screenshots from MATLAB)

d) Cross-sectional regression R-squared

As Figure 4 shows, the cross-sectional R^2 of 0.4903 indicates that the model leaves approximately 51% of the variation in excess returns unexplained, highlighting the limited robustness and explanatory power of the six-factor model. In other words, the model is insufficient in accounting for half of the cross-sectional variation. Additionally, the comparison between the cross-sectional R^2 of 0.4903 and the average time-series R^2 of 0.8274 highlights consistent limitations in the explanatory power of the six-factor model. This disparity implies that while systematic risk factors may effectively capture return dynamics over time, they are less capable of differentiating between unique portfolio characteristics across the cross-section, indicating the potential presence of omitted factors, especially the industry-specific risks or characteristics (e.g. industry affiliation, country of domicile) of the additional 17 industry portfolios as evidenced by the low time-series R^2, and the 42 residual plot graphs (Please refer to **Appendix 2** for residual plots).

```
R-squared in cross-sectional:
    0.4903
```

Figure 4: Cross-sectional R-squared

This suggests that further enhancements to the model are needed, for example, to mitigate issues of redundancy suggested by Fama and French (2015), we can use the PCA to address the interplay among factors, by extracting common factors from the cross-sectional stock returns that explain maximum common variation across stocks and then determining the significance of the factors (Chincarini and Kim, 2022). Furthermore, it is recommended to analyse the model's performance during specific economic cycles, as the leverage cycle theory (Adrian and Boyarchenko, 2012) suggests that the behaviour of financial intermediaries varies across different economic environments, for example, the leverage increases during economic booms and decreases during recession. These variations can result in risk premium changes and, therefore, impact the market factor’s significance in different stages of economic periods, and other factors due to possible correlations. This could potentially influence the model’s explanatory power for asset returns during crisis periods compared to stable periods and, therefore, decrease cross-sectional R-squared. (Please refer to **Appendix 6 Panel C** for factor correlation evidence.)

Addressing these shortcomings through omitted variable inclusion, factor interactions, and dynamic modelling could enhance its ability to explain cross-sectional return variation, strengthening its relevance in asset pricing research.

e) The GRS test

The GRS test evaluates the cross-sectional explanatory power of an asset pricing model by testing whether the intercepts in a regression of excess returns are collectively zero (Gibbons et al., 1989). In this analysis, as Figure 5 shows, the calculated test statistic is 3.1089, accompanied by a p-value of zero. This implies that, under the null hypothesis of zero intercepts, the probability of observing a GRS statistic greater than or equal to the critical value of 3.1089 is effectively negligible. Consequently, we reject the null hypothesis, suggesting that the six-factor model does not fully account for the variations in excess returns, leaving systematic patterns unexplained.

```
GRS test Intercept and its p-value:
    3.1089    0.0000
```

Figure 5: Intercept and p-value through GRS test

However, when the GRS test rejects our model with redundancy factors, it is unable to identify the sources of the problem (Shanken, 1992). In addition, under limited sample sizes, the GRS test is highly sensitive to the assumption of normality and structure of residual covariance (Gibbons et al., 1989), indicating a potential need for a robustness check. Moreover, Roll (1977) highlights that if the chosen proxy does not perfectly represent the true market portfolio, then the tests do not directly evaluate the validity of the theoretical model, underscoring the gap between theoretical assumptions and real-world implementation. The rejection by the GRS test therefore underscores the need for refinement, exploring omitted variables, and addressing time-varying risk premia.

f) Explanatory power of the six-factor model

Fama and French (2015) argue that the five-factor model is an improvement of the three-factor model, and it is useful in explaining variation in excess returns. Whilst we agree with their claim, the rejection of our six-factor model and FF five-factor model by the GRS test (statistic = 3.1089, p-value ≈ 0) suggests that this model may not be able to capture all relevant factors affecting excess stock returns. However, Shanken (1992) argues that the GRS test may reject the model if the model includes redundant factors. Therefore, the failed result may not be due to the model being wrong but due to some other factors.

Overall, while the six-factor model demonstrates strong time-series explanatory power, its rejection by the GRS test highlights limitations in cross-sectional applications. The unexplained 51% cross-sectional variation, along with insignificant residuals, suggests the presence of noise or potential missing factors. By applying PCA to select the top principal components and ignoring components with smaller eigenvalues — typically carrying redundant information or noise — this could improve the model's robustness (Zuur et al., 2007). To address these limitations, future research may explore additional potential variables, such as macroeconomic indicators or dynamic factors.

Conclusion

Incorporating our discussion and findings from Part 1 and Part 2, Fama and French's five- and six-factor models represent major advances in asset pricing, with the five-factor model refining the understanding of systematic risk by incorporating profitability and investment. However, anomalies such as the underperformance of small-cap, high-investment, and low-profitability stocks expose its limitations, challenging the strongest form of the Efficient Market Hypothesis (EMH). The six-factor model, despite strong time-series performance, struggles in cross-sectional analyses, with GRS test results and unexplained variations indicating factor redundancy and omitted drivers of returns.

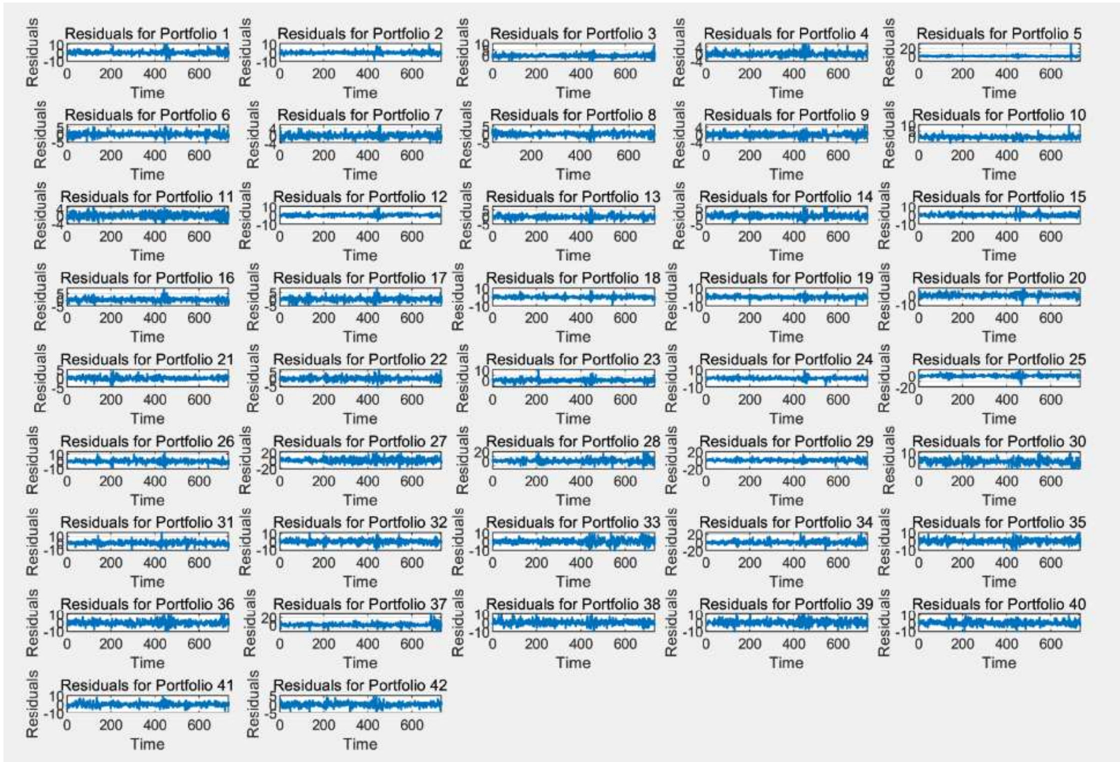
Enhancements such as interaction terms, principal component analysis, and additional macroeconomic or behavioural factors may improve the models' robustness. Analysing performance across economic cycles could also further reveal the strengths and weaknesses of the explanatory power of each factor. While these models have significantly advanced asset pricing and investment strategies, their imperfections call for ongoing refinement to better address market complexities and bridge theory with practice.

Appendix 1: Time Series R-squared

```
R2 =

    0.9244
    0.9391
    0.9515
    0.9517
    0.9056
    0.9540
    0.9555
    0.9442
    0.9486
    0.9532
    0.9502
    0.9319
    0.9190
    0.9247
    0.9090
    0.9333
    0.9088
    0.8968
    0.8893
    0.8832
    0.9531
    0.8993
    0.8534
    0.8917
    0.8223
    0.6542
    0.4107
    0.4723
    0.7291
    0.7644
    0.7440
    0.6475
    0.7978
    0.6510
    0.7571
    0.8159
    0.5840
    0.7861
    0.4311
    0.7217
    0.8485
    0.9399
```

Appendix 2: Plot of residuals of 42 portfolios



Appendix 3: Comparison between cross-sectional estimates and corresponding t-statistics

```
Cross-sectional estimates and their t-statistic:
    0.6091    2.2027
   -0.0132   -0.0403
    0.2039    1.7628
    0.2436    2.1242
    0.1994    1.3085
   -0.0377   -0.1830
    0.7271    1.4261
```

Appendix 4: Time-series sample mean and cross-sectional estimates

```
Time-series sample mean and Cross-sectional estimates:
    1.0172   -0.0132
    0.3885    0.2039
    0.1604    0.2436
    0.0927    0.1994
    0.0580   -0.0377
   -0.0431    0.7271
```

Appendix 5: Critical values at 5% and 10% significance level

```
Critical Value at 5%:
   -2.0301

Critical Value at 10%:
   -1.6896
```

Appendix 6: Correlations between factors

Table 4

Summary statistics for monthly factor percent returns; July 1963–December 2013, 606 months.

$R_M - R_F$ is the value-weight return on the market portfolio of all sample stocks minus the one-month Treasury bill rate. At the end of each June, stocks are assigned to two *Size* groups using the NYSE median market cap as the breakpoint. Stocks are also assigned independently to two or three book-to-market equity (*B/M*), operating profitability (*OP*), and investment (*Inv*) groups, using NYSE medians of *B/M*, *OP*, and *Inv* or the 30th and 70th NYSE percentiles. In the first two blocks of Panel A, the *B/M* factor, *HML*, uses the VW portfolios formed from the intersection of the *Size* and *B/M* sorts ($2 \times 2 = 4$ or $2 \times 3 = 6$ portfolios), and the profitability and investment factors, *RMW* and *CMA*, use four or six VW portfolios from the intersection of the *Size* and *OP* or *Inv* sorts. In the third block, *HML*, *RMW*, and *CMA* use the intersections of the *Size*, *B/M*, *OP*, and *Inv* sorts ($2 \times 2 \times 2 \times 2 = 16$ portfolios). HML_B is the average return on the portfolio(s) of big high *B/M* stocks minus the average return on the portfolio(s) of big low *B/M* stocks, HML_S is the same but for portfolios of small stocks, *HML* is the average of HML_S and HML_B, and $HML_{S\text{-}B}$ is the difference between them. RMW_S, RMW_B, *RMW*, and $RMW_{S\text{-}B}$ and CMA_S, CMA_B, *CMA*, and $CMA_{S\text{-}B}$ are defined in the same way, but using high and low *OP* or *Inv* instead of *B/M*. In the $2 \times 2 \times 2 \times 2$ sorts, *SMB* is the average return on the eight portfolios of small stocks minus the average return on the eight portfolios of big stocks. In the separate 2×3 *Size-B/M*, *Size-OP*, and *Size-Inv* sorts, there are three versions of *SMB*, one for each 2×3 sort, and *SMB* is the average of the three. *SMB* in the separate 2×2 sorts is defined similarly. Panel A of the table shows average monthly returns (Mean), the standard deviations of monthly returns (Std dev.) and the *t*-statistics for the average returns. Panel B shows the correlations of the same factor from different sorts and Panel C shows the correlations for each set of factors.

Panel A: Averages, standard deviations, and t-statistics for monthly returns

	2×3 Factors					2×2 Factors					$2 \times 2 \times 2 \times 2$ Factors				
	$R_M - R_F$	*SMB*	*HML*	*RMW*	*CMA*	$R_M - R_F$	*SMB*	*HML*	*RMW*	*CMA*	$R_M - R_F$	*SMB*	*HML*	*RMW*	*CMA*
Mean	0.50	0.29	0.37	0.25	0.33	0.50	0.30	0.28	0.17	0.22	0.50	0.30	0.30	0.25	0.14
Std dev.	4.49	3.07	2.88	2.14	2.01	4.49	3.13	2.16	1.52	1.48	4.49	2.87	2.13	1.49	1.29
t-Statistic	2.74	2.31	3.20	2.92	4.07	2.74	2.33	3.22	2.79	3.72	2.74	2.60	3.43	4.09	2.71

	HML_S	HML_B	$HML_{S\text{-}B}$	RMW_S	RMW_B	$RMW_{S\text{-}B}$	CMA_S	CMA_B	$CMA_{S\text{-}B}$
2×3 factors									
Mean	0.53	0.21	0.32	0.33	0.17	0.16	0.45	0.22	0.23
Std dev.	3.24	3.11	2.69	2.69	2.35	2.68	2.00	2.66	2.47
t-Statistic	4.05	1.69	2.94	3.06	1.81	1.48	5.49	2.00	2.29
2×2 Factors									
Mean	0.40	0.16	0.24	0.22	0.13	0.09	0.33	0.11	0.22
Std dev	2.39	2.36	1.97	1.93	1.69	2.00	1.53	1.87	1.70
t-Statistic	4.16	1.68	3.05	2.76	1.86	1.09	5.37	1.50	3.17
$2 \times 2 \times 2 \times 2$ Factors									
Mean	0.37	0.22	0.16	0.30	0.21	0.09	0.23	0.07	0.17
Std dev.	2.40	2.36	2.01	2.18	1.53	2.22	1.23	1.58	1.59
t-Statistic	3.83	2.28	1.91	3.41	3.38	1.02	4.64	1.03	2.56

Panel B: Correlations between different versions of the same factor

	SMB			*HML*			*RMW*			*CMA*		
	2×3	2×2	$2 \times 2 \times 2 \times 2$	2×3	2×2	$2 \times 2 \times 2 \times 2$	2×3	2×2	$2 \times 2 \times 2 \times 2$	2×3	2×2	$2 \times 2 \times 2 \times 2$
2×3	1.00	1.00	0.98	1.00	0.97	0.94	1.00	0.96	0.80	1.00	0.95	0.83
2×2	1.00	1.00	0.98	0.97	1.00	0.96	0.96	1.00	0.83	0.95	1.00	0.87
$2 \times 2 \times 2 \times 2$	0.98	0.98	1.00	0.94	0.96	1.00	0.80	0.83	1.00	0.83	0.87	1.00

Panel C: Correlations between different factors

	2×3 Factors					2×2 Factors					$2 \times 2 \times 2 \times 2$ Factors				
	$R_M - R_F$	*SMB*	*HML*	*RMW*	*CMA*	$R_M - R_F$	*SMB*	*HML*	*RMW*	*CMA*	$R_M - R_F$	*SMB*	*HML*	*RMW*	*CMA*
$R_M - R_F$	1.00	0.28	−0.30	−0.21	−0.39	1.00	0.30	−0.34	−0.13	−0.43	1.00	0.25	−0.33	−0.27	−0.42
SMB	0.28	1.00	−0.11	−0.36	−0.11	0.30	1.00	−0.16	−0.32	−0.13	0.25	1.00	−0.21	−0.33	−0.21
HML	−0.30	−0.11	1.00	0.08	0.70	−0.34	−0.16	1.00	0.04	0.71	−0.33	−0.21	1.00	0.63	0.37
RMW	−0.21	−0.36	0.08	1.00	−0.11	−0.13	−0.32	0.04	1.00	−0.19	−0.27	−0.33	0.63	1.00	0.15
CMA	−0.39	−0.11	0.70	−0.11	1.00	−0.43	−0.13	0.71	−0.19	1.00	−0.42	−0.21	0.37	0.15	1.00

Appendix 7: MATLAB Codes

```
%% Part A
%%Build the structure of excess return Re and fatcors f
Rf = RF(:,:); %% Risk free rate for each month
Rm = FF42(:,:); %% Market return on 42 portfolios
Re = Rm - Rf; %% Excess Return (Return on market - Return on risk free rate)

T = size(Re,1); %% Define Number of samples for each portfolios
N = 42; %% Define Number of Portfoliois
k = 6;  %% Define Number of Factors

%%%Factors columns Set Up
MKT = FF6(:,1);
SMB = FF6(:,2);
HML = FF6(:,3);
RMW = FF6(:,4);
CMA = FF6(:,5);
MOM = FF6(:,6);

f = [MKT SMB HML RMW CMA MOM]; %%Six Factors Matrix
Vf = cov(f,1); %%Population covariance matrix Vf
F = [ones(T,1) f];  %Factors; add ones as intercept to capture the Ave return not capturing by factors.

disp(Vf)

%We use Augmented Dickey-Fuller (ADF) test on the Re(:,i) and f(:,i).
%adftest function provided by Matlab
% Non-stationarity test for Re
h1 = []; %%Matrix to store the result of the test for excess return
for i = 1:N
h1(1, end+1) = adftest(Re(:,i));
end

% Non-stationarity test for f
h2 = []; %%Matrix to store the result of the test for factors
for i = 1:k
h2(1, end+1) = adftest(f(:,i));
end

disp(h1);
```

```
%% Part b %%
% OLS estimators %%Then we can generate estimated alpha and beta
betahat = []; %% Store the alpha and beta estimates
for i = 1:N
    betahat(end+1,:) = inv(F' * F) * F' * Re(:,i);  %%Formula: OLS estimated beta= (X'X)^(-1)*X'y (here y=Re;X=F)
end

%uhat is estimated residuals
uhat = []; %%Store the residual estimates
for i = 1:N
y = Re(:,i);
uhat(:, end+1) = y - F * betahat(i,:)';  %%%% Residual = excess return-estimated beta * X; here X=F
end

% R_squared and avgR2 (average R-squared)
R2 = []; %%Store the R-squared estimates
for i = 1:N
y = Re(:,i);
R2(end+1,1) = 1 - uhat(:,i)'*uhat(:,i)/(sum((y - mean(y)).^2)); %% R-squared=1-SSR/SST
end

avgR2 = mean(R2);
disp('R-squared:');
disp(R2);
disp('Average R-squared:');
disp (avgR2);
```

```
%% part c %%
%%Average excess return: Rebar--Given in IB9X60_QMF_Group_Work_2024.pdf--Instruction (2)
Rebar = [];
for i = 1:N
Rebar(end+1,1) = mean(Re(:,i));
end
%Coefficient estimates for the factors from the time-series regression
B = betahat(:,2:end); %% Estimated beta for six factors
Xhat = [ones(N,1) B];   %%% Given in IB9X60_QMF_Group_Work_2024.pdf--Instruction (4)
gamahat = inv(Xhat'*Xhat)*Xhat'*Rebar; %%%Given in IB9X60_QMF_Group_Work_2024.pdf--Instruction (3).
Ahat = inv(Xhat'*Xhat)*Xhat'; %%%Given in IB9X60_QMF_Group_Work_2024.pdf--Instruction (4)

% Covariance of gamahat
gamahat1 = gamahat(2:end,1);  %% Extract estimates without intercept
Ehat = cov(uhat); %%According to shaken(1992).
zerok = zeros(k,1); % k = 6
%%formula below are given in IB9X60_QMF_Group_Work_2024.pdf--Instruction (4)
m = [0 zerok'; zerok Vf];
VS = (1 + gamahat1'*inv(Vf)*gamahat1)*Ahat*Ehat*Ahat' + m;

% t-statistic
segamahat = sqrt(diag(VS)/T); %%%Given in IB9X60_QMF_Group_Work_2024.pdf--Instruction (4)--Standard errors
[gamahat gamahat./(segamahat)]; %%%given in IB9X60_QMF_Group_Work_2024.pdf--Instruction (4)
disp('Cross-sectional estimates and their t-statistic:');
disp([gamahat gamahat./(segamahat)]);
c5 = tinv(0.025,N-k-1); %Critical Value at 5% significance level
c10 = tinv(0.05,N-k-1); %Critical Value at 10% significance level
disp('Critical Value at 5%:');
disp(c5);
disp('Critical Value at 10%:');
disp(c10);
%%%Time Series sample Mean
Bbar = [];
for i=1:k %%k=6
    Bbar(end+1,1) = mean(B(:,i));
end
[Bbar gamahat1]; %% Comparison between time series sample mean and cross section estimates.
disp('Time-series sample mean and Cross-sectional estimates:');
disp([Bbar gamahat1]);
```

```
%% Part d%%
ehat = Rebar - Xhat*gamahat; %%given in IB9X60_QMF_Group_Work_2024.pdf--Instruction (5)
ehat0 = Rebar - ones(N,1)*(ones(N,1)'*Rebar/N); %%given in IB9X60_QMF_Group_Work_2024.pdf--Instruction (6)
R2cs = 1 - ehat'*ehat/(ehat0'*ehat0);  %%given in IB9X60_QMF_Group_Work_2024.pdf--Instruction (6)
disp('R-squared in cross-sectional:');
disp(R2cs);
```

```
%% Part e%%
%%%According to the application of GRS test; According to Gibbons, Ross, and Shanken (GRS, 1989).

r1 = f;
r2 = Re;
[stat,pval] = GRS(r1,r2); %%% We save the grs function code in our file, so can direct apply here.

disp('GRS test Intercept and its p-value:');
disp([stat pval])
```

Reference list

Adrian, T. and Boyarchenko, N. 2012. *Intermediary Leverage Cycles and Financial Stability* [Online]. Chicago: THE UNIVERSITY OF CHICAGO. [Accessed 17 November 2024]. Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2133385.

Baum, A. and Koester, G. 2011. The Impact of Fiscal Policy on Economic Activity Over the Business Cycle - Evidence from a Threshold VAR Analysis. *papers.ssrn.com*. [Online]. Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2785397

Baumohl, E. and Lyocsa, S. 2009. Stationarity of Time Series and the Problem of Spurious Regression. *papers.ssrn.com*. [Online]. Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1480682.

Choi, D. and Hui, S.K. 2014. The role of surprise: Understanding overreaction and underreaction to unanticipated events using in-play soccer betting market. *Journal of Economic Behavior & Organization*. **107**, pp.614–629.

Chincarini, L.B. and Kim, D. 2022. *Quantitative Equity Portfolio Management, Second Edition: An Active Approach to Portfolio Construction and Management* [Online] 2nd ed. McGraw Hill Professional. [Accessed 16 November 2024]. Available from: https://ebookcentral.proquest.com/lib/warw/detail.action?docID=30159254.

Carnival Corporation & Plc (CCL) (2023). *Carnival Corporation & PLC 2022 ANNUAL REPORT*. Miami Florida U.S.A: February 2023. Carnival Corporation & PLC website. Available at: Financial information- https://www.carnivalcorp.com/static-files/dc2f5ce1-72f7-4147-9f28-5f4d15e8ec4d.

Fama, E.F. and French, K.R. 2015. A five-factor asset pricing model. *Journal of Financial Economics*. **116**(1), pp.1–22.

Gibbons, M.R., Ross, S.A. and Shanken, J. 1989. A Test of the Efficiency of a Given Portfolio. *Econometrica*. **57**(5), p.1121.

Jegadeesh, N. and Titman, S. 1993. Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *The Journal of Finance*. **48**(1), pp.65–91.

Roll, R. 1977. A Critique of the Asset Pricing theory's Tests Part I: on past and Potential Testability of the Theory. *Journal of Financial Economics*. **4**(2), pp.129–176.

Shanken, J. 1992. On the Estimation of Beta-Pricing Models. *Review of Financial Studies*. **5**(1), pp.1–33.

Wang, Q., Liu, L. 2022 Pandemic or panic? A firm-level study on the psychological and industrial impacts of COVID-19 on the Chinese stock market. *Finance Innov* **8**, 36 (2022). https://doi.org/10.1186/s40854-022-00335-8.

Yahoo 2024. Carnival Corporation (CCL) Stock Price, Quote, History & News - Yahoo Finance. *finance.yahoo.com*. [Online]. [Accessed 28 November 2024]. Available from: https://finance.yahoo.com/quote/CCL/.

Zuur, A.F., Ieno, E.N. and Smith, G.M. 2007. *Principal component analysis and redundancy analysis* [Online]. New York: Springer, pp.193–224. [Accessed 20 November 2024]. Available from: https://link.springer.com/chapter/10.1007/978-0-387-45972-1_12.