Investment Task 2

Siqi Zhu

May 2025

Contents

1 Introduction

This document presents an analysis of the Fama-French three-factor model's ability to explain the equity premium of ten selected US industry Exchange Traded Funds (ETFs) for the period spanning January 2015 to December 2024. The analysis employs a two-pass regression methodology. The first pass consists of time-series regressions of each ETF's excess returns on the Fama-French factors. The second pass involves a cross-sectional regression of the average excess ETF returns on their estimated factor betas from the first pass.

1.1 Reproduce the Results

I used Python to collect and analyze data for this task. All results presented in this report can be fully reproduced by running the Jupyter Notebook available at this link.

To reproduce the results:

1. First, upload the notebook file to Google Colab (see Image 1a).
2. Once the notebook is opened, upload the required Excel data file from this link to the workspace folder on the left-hand side (see Image 1b).
3. Then, click `Runtime` → `Run all` in the top menu to execute the entire notebook (see Image 1c).

1.2 Data Content

The primary datasets for this analysis include:

- **ETF Historical Data:** Monthly historical price data for ten selected U.S. industry-specific Exchange Traded Funds (ETFs). From this, monthly returns were calculated. The key price series used for return calculation was the 'Adjusted Close' price, which accounts for dividends and stock splits, or 'Close' price where 'Adjusted Close' was not available.
- **Fama-French Factors:** Monthly data for the three Fama-French factors:
  - Mkt-RF: The market risk premium (excess return of the market over the risk-free rate).
  - SMB (Small Minus Big): The return difference between a portfolio of small-cap stocks and a portfolio of large-cap stocks.

(a) Upload ipynb (b) Upload xlsx (c) Run code

Figure 1: Reproduce the results

  - HML (High Minus Low): The return difference between a portfolio of high book-to-market (value) stocks and a portfolio of low book-to-market (growth) stocks.

- **Risk-Free Rate (RF):** Monthly data for the U.S. risk-free rate, typically derived from U.S. Treasury bill rates, obtained alongside the Fama-French factors.

1.3 Data Sources

The financial data were sourced as follows:

- **ETF Historical Data:** Retrieved programmatically from Yahoo Finance using the `yfinance` Python library. Yahoo Finance is a comprehensive source for historical and current financial market data. The data was initially downloaded at a daily frequency and then resampled to obtain monthly closing prices for return calculations. You can download this part of data at link.

- **Fama-French Factors and Risk-Free Rate:** This data was accessed programmatically using the `pandas_datareader` Python library, which provides an interface to various data sources, including the Fama-French factors. The data is provided in percentage terms and was converted to decimal form for calculations. The data is not saved, but you can access it by running the previously mentioned code.

1.4 Industry Selection

The ten U.S. industry included in this study were selected to represent a diverse range of economic sectors. The chosen industries are the same as in Task 1.

1.5 Time Period

The analysis covers the period from **January 1, 2015, to December 31, 2024**. All data (returns, Fama-French factors, and the risk-free rate) were aligned to this monthly frequency within this timeframe.

2 First-Pass Time-Series Regressions

The model for the first-pass regression for each ETF i is:

$$R_{it} - RF_t = \alpha_i + \beta_{i,Mkt}(Mkt - RF)_t + \beta_{i,SMB}SMB_t + \beta_{i,HML}HML_t + e_{it}$$

Where:

- $R_{it} - RF_t$ is the excess return of ETF i in month t.
- α_i is the Jensen's alpha for ETF i.
- $(Mkt - RF)_t$ is the market risk premium in month t.
- SMB_t (Small Minus Big) is the size factor return in month t.
- HML_t (High Minus Low) is the value factor return in month t.
- $\beta_{i,Mkt}$, $\beta_{i,SMB}$, $\beta_{i,HML}$ are the factor sensitivities (betas) for ETF i.
- e_{it} is the error term for ETF i in month t.

2.1 Summary of First-Pass Results

see Figure 2.

--- First-Pass Regression Results Summary ---

ETF	Alpha	t_Alpha	Beta_Mkt-RF	t_Mkt-RF	Beta_SMB	t_SMB	Beta_HML	t_HML	R-Squared	Adj. R-Squared	N_obs
XLV	-0.0011	-0.4592	0.7366	13.3732	-0.1519	-1.6706	-0.0414	-0.6451	0.6164	0.6064	119.0000
DBA	-0.0006	-0.1892	0.1918	2.9496	-0.1498	-1.3948	0.2553	3.3694	0.1477	0.1254	119.0000
XME	-0.0001	-0.0184	1.2953	9.0024	0.6739	2.8366	0.7665	4.5722	0.5510	0.5393	119.0000
XLI	-0.0013	-0.7068	1.0664	25.9737	0.0125	0.1843	0.2817	5.8902	0.8725	0.8692	119.0000
XLU	0.0007	0.1945	0.4984	6.0590	-0.4260	-3.1366	0.1123	1.1717	0.2563	0.2369	119.0000
XLRE	-0.0037	-1.0973	0.8426	11.3355	0.0183	0.1490	-0.0130	-0.1548	0.5778	0.5659	110.0000
XRT	-0.0025	-0.5868	1.1242	11.7446	0.8869	5.6117	0.2265	2.0307	0.6740	0.6655	119.0000
IYT	-0.0050	-1.8478	1.1566	19.2969	0.1319	1.3331	0.2482	3.5549	0.7947	0.7893	119.0000
PEJ	-0.0062	-1.8963	1.1607	16.1067	0.3726	3.1314	0.2719	3.2383	0.7497	0.7432	119.0000
XLF	0.0002	0.1258	1.0426	25.3713	0.0218	0.3209	0.5956	12.4421	0.8852	0.8822	119.0000

Figure 2: first-pass results

2.2 Interpretation of First-Pass Results

- **Alphas (α_i):** The alphas for most ETFs are not statistically significantly different from zero (based on t-statistics generally having absolute values less than 2). This suggests that the Fama-French three-factor model largely explains the average returns of these ETFs, leaving little unexplained systematic outperformance or underperformance. ETFs IYT (Transportation) and PEJ (Leisure & Entertainment) show alphas approaching statistical significance (t-statistics ≈ -1.85 and -1.90 respectively), hinting at slight underperformance not captured by the model.

- **Market Beta (β_{Mkt}):** All ETFs exhibit positive and highly statistically significant market betas, indicating a positive correlation with overall market movements, as expected for equity investments. DBA (Agriculture) and XLU (Utilities) are considerably less sensitive (betas < 0.5).

- **Size Factor Beta (β_{SMB}):**
  - Significant positive exposure (tilt towards small-cap stocks) is observed for XME (Mining), XRT (Retail), and PEJ (Leisure & Entertainment).
  - XLU (Utilities) shows a significant negative exposure, indicating large-cap stock characteristics.
  - Other ETFs do not exhibit a statistically significant size tilt.

- **Value Factor Beta (β_{HML}):**
  - Significant positive exposure (tilt towards value stocks) is observed for DBA (Agriculture), XME (Mining), XLI (Industrials), XRT (Retail), IYT (Transportation), PEJ (Leisure & Entertainment), and XLF (Financials). XLF shows particularly strong value characteristics.
  - XLV (Healthcare), XLU (Utilities), and XLRE (Real Estate) do not show a statistically significant value or growth tilt.

- **R-Squared (R^2):** The model's explanatory power varies across ETFs. It is very high for XLI (87.3%) and XLF (88.5%). It is moderate for most other ETFs. However, the R^2 is relatively low for DBA (14.8%) and XLU (25.6%), suggesting that the Fama-French factors explain a smaller proportion of the return variations for these specific sectors, and other idiosyncratic or sector-specific factors may be more influential.

3 Second-Pass Cross-Sectional Regression

The second-pass regression tests whether the estimated factor betas from the first pass can explain the cross-sectional variation in average ETF excess returns. The model is:

$$\overline{R_i - RF} = \gamma_0 + \gamma_{MktRF}\hat{\beta}_{i,MktRF} + \gamma_{SMB}\hat{\beta}_{i,SMB} + \gamma_{HML}\hat{\beta}_{i,HML} + u_i$$

Where:

- $\overline{R_i - RF}$ is the average excess return for ETF i over the sample period.
- γ_0 is the intercept of the cross-sectional regression.
- γ_{MktRF}, γ_{SMB}, γ_{HML} are the estimated risk premia associated with the market, size, and value factors, respectively.
- $\hat{\beta}_{i,MktRF}$, $\hat{\beta}_{i,SMB}$, $\hat{\beta}_{i,HML}$ are the estimated factor betas for ETF i from the first-pass regressions.
- u_i is the error term.

3.1 Summary of Second-Pass Results

see Figure 3.

```
--- Second-Pass Regression Results ---
          Coefficient  Estimate  Std. Error  t-statistic  p-value
gamma_0 (Intercept)  0.000039    0.002383     0.016367 0.987473
   gamma_beta_MktRF  0.006409    0.002843     2.253826 0.065098
     gamma_beta_SMB -0.002590    0.002445    -1.059233 0.330261
     gamma_beta_HML  0.004466    0.003129     1.427369 0.203382

R-Squared: 0.6609
Adj. R-Squared: 0.4913
Number of observations (ETFs): 10
```

Figure 3: second-pass results

3.2 Interpretation of Second-Pass Results

- **Intercept (γ_0):** The estimate is 0.000039, which is economically and statistically indistinguishable from zero (p-value = 0.987). A non-significant γ_0 suggests that the Fama-French model, on average, does not leave a common unexplained portion of returns across the ETFs after accounting for their factor exposures. This is consistent with the model being well-specified for this cross-section.
- **Market Risk Premium (γ_{MktRF}):** The estimated premium for market risk is 0.006409 (0.64% per month). This premium is positive, as theoretically expected, and is marginally statistically significant (p-value = 0.065). This indicates that, on average, ETFs with higher market betas tended to earn higher returns during the sample period.
- **Size Premium (γ_{SMB}):** The estimated premium for the size factor is -0.002590 (-0.26% per month). This is not statistically significant (p-value = 0.330). The negative sign (though insignificant) suggests that exposure to small-cap characteristics was not, on average, rewarded with higher returns for this set of ETFs over this period.
- **Value Premium (γ_{HML}):** The estimated premium for the value factor is 0.004466 (0.45% per month). While positive, consistent with the notion of a value premium, this estimate is not statistically significant (p-value = 0.203). This implies that the evidence for a value premium among these specific ETFs during this period is not statistically robust.
- **R-Squared (R^2):** The R^2 of the cross-sectional regression is 0.6609. This indicates that approximately 66.1% of the cross-sectional variation in the average excess returns of these ten ETFs can be explained by their differing exposures to the Fama-French factors. The adjusted R^2 is 0.4913. This is a reasonably good fit, especially considering the small sample size of 10 ETFs.

4 Overall Conclusion

The Fama-French three-factor model demonstrates reasonable efficacy in explaining the time-series returns of the selected US industry ETFs for the 2015-2024 period. Most ETFs do not exhibit statistically significant alphas, suggesting the model captures their systematic risk exposures adequately. The explanatory power (R^2) varies, being high for broad sectors like Industrials and Financials, but lower for sectors like Agriculture and Utilities.

In the cross-sectional analysis, the market risk factor (β_{MktRF}) appears to be priced, with a positive and marginally significant risk premium. However, there is no statistically significant evidence from this sample that the size (SMB) or value (HML) factors commanded a risk premium during the period. The intercept (γ_0) of the cross-sectional regression is not significantly different from zero, lending support to the model's overall specification for this set of assets.

The relatively high R^2 in the second-pass regression suggests that the Fama-French factor betas do explain a substantial portion of why average returns differed across these ETFs. However, the limited sample size (10 ETFs) warrants caution in generalizing these findings, particularly concerning the statistical significance of the factor risk premia. Standard errors in the second-pass regression are also subject to the errors-in-variables problem, which could be addressed by more advanced econometric techniques.