Multiple Regression Analysis on Lo 30 Portfolio Returns: A Fama–French 3- and 5-Factor Approach

Wei Chieh Chen
Syracuse University

2025-04-01

Introduction

Overview: This project demonstrates how to perform two multiple linear regression analyses using the "Lo 30" portfolio returns as the response variable. Two sets of factor datasets are used:

- Fama–French 3-Factor Model: Market excess return (Mkt-RF), SMB, and HML.
- Fama–French 5-Factor Model: The above factors plus RMW and CMA.

Both models are estimated for data from July 1, 1963, onward.

Data Import

Summary of Steps:

- Downloaded ZIP archives for three datasets from online sources:
  - Fama-French 3-Factor Data (Weekly)
  - Fama-French 5-Factor Data (Daily)
  - Portfolios Data (Daily)
- Created temporary ZIP files and extracted the first CSV file from each archive.
- Read the CSV files using fread() and previewed the data.
- Renamed the first column to “Date” for consistency.
- Optionally removed temporary ZIP files to clean up.

Data Preparation & Merging the Datasets

Data Preparation Summary:

- Convert "Date" columns to proper Date objects using ymd().
- Filter each dataset to include only records from July 1, 1963 onward.
- Adjust column names in the portfolios dataset with make.names() and rename "Lo 30" (i.e., "Lo.30") to "Lo30".

Merging Summary:

- Merge the portfolios data with the Fama-French 3-Factor data by the "Date" variable.
- Merge the portfolios data with the Fama-French 5-Factor data similarly to ensure matching dates for regression analysis.

Display summary of merged data (example)

```
summary(data_3f[, c(21:23)])
```

```
##       MKTRF                SMB                HML          
##  Min.   :-18.0000   Min.   :-10.0400   Min.   :-8.94000  
##  1st Qu.: -1.0700   1st Qu.: -0.6900   1st Qu.:-0.62000  
##  Median :  0.2800   Median :  0.0500   Median : 0.03000  
##  Mean   :  0.1289   Mean   :  0.0267   Mean   : 0.06521  
##  3rd Qu.:  1.3800   3rd Qu.:  0.7500   3rd Qu.: 0.70000  
##  Max.   : 13.4600   Max.   :  6.9400   Max.   :10.87000
```

Regression Analysis – Fama-French 3-Factor Model

- Fit the 3-Factor regression model and display the summary.

Table 1: Coefficient Table for Fama-French 3-Factor Model

term	estimate	std.error	statistic	p.value
(Intercept)	0.1376482	0.0159182	8.6472433	0.0000000
MKTRF	0.1873448	0.0072505	25.8387630	0.0000000
SMB	0.1319623	0.0127662	10.3368500	0.0000000
HML	0.0011474	0.0116503	0.0984835	0.9215546

Regression Analysis – Fama-French 3-Factor Model

Interpretation:

- The intercept, MKTRF, and SMB are statistically significant.
- HML is not significant.
- Approximately 21% of the variance in Lo30 is explained.

Regression Analysis – Fama-French 5-Factor Model

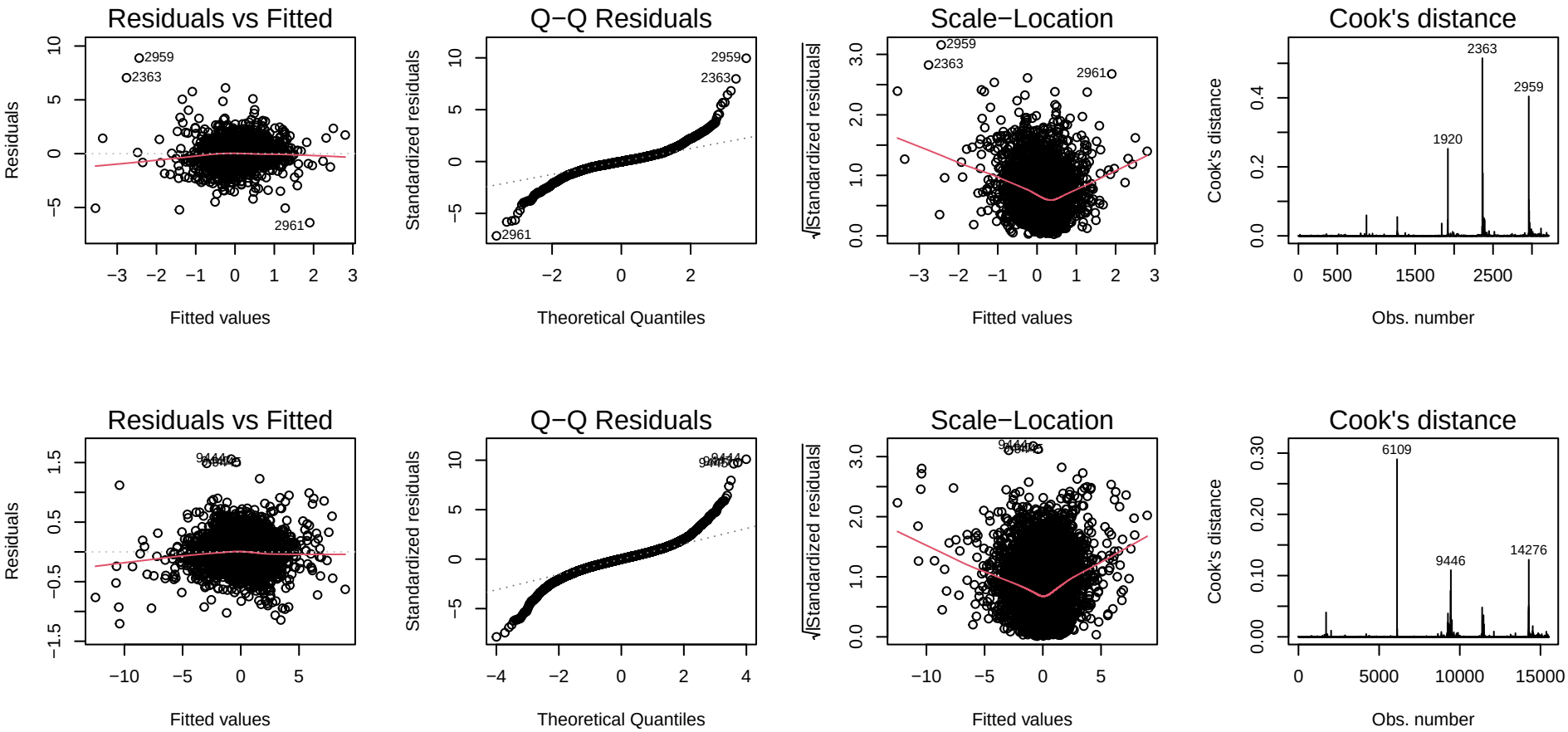
Table 2: Coefficient Table for Fama-French 5-Factor Model

term	estimate	std.error	statistic	p.value
(Intercept)	0.0151992	0.0012436	12.22187	0
MKTRF	0.9703800	0.0013321	728.48422	0
SMB	1.0075361	0.0023986	420.04633	0
HML	0.0926974	0.0025712	36.05276	0
RMW	-0.1022346	0.0033224	-30.77097	0
CMA	0.0622984	0.0041332	15.07278	0

- All predictors are statistically significant.
- The model explains over 98% of the variance in Lo30.

Regression Diagnostics – Fama-French Models

The following plots assess the regression assumptions (linearity, homoscedasticity, normality, and influential observations) for both models. The top row shows the diagnostics for the 3-Factor model, while the bottom row shows those for the 5-Factor model.



Conclusion & Next Steps

Key Findings:

- The 3-Factor model shows moderate explanatory power.
- The 5-Factor model fits exceptionally well but shows some departures from classical assumptions.

Diagnostics Summary:

- Both models have issues with non-normality and heteroscedasticity.
- Future work should consider robust standard errors or alternative estimation techniques.

Next Steps:

- Explore model refinements.
- Validate findings with out-of-sample tests.