Asset Pricing Empirical Project: Fama-MacBeth (1973) Analysis

Caterina Piacentini   Farkas Tallos   Luuk Vos   Sam Friedlaender
Timur Kambachekov

January 9, 2026

Contents

1 1. Motivation and Data

1.1 1.1 Motivation

Financial markets exhibit large cross-sectional differences in average returns. A central question in asset pricing is whether these differences can be explained by systematic risk, rather than by chance or mispricing.

The Fama-MacBeth (1973) framework provides a standard empirical way to test this idea. It links assets' long-run returns to their exposure to common risk factors and evaluates whether these risk exposures are rewarded with higher expected returns. This model allows us to: 1. Assess whether proposed factors are economically meaningful. 2. Distinguish priced risk from noise in asset returns. 3. Evaluate the stability of risk premia over time, especially across different market environments.

1.2 1.2 Data Source & Universe Construction

We utilize data from the Center for Research in Security Prices (CRSP) via WRDS and the Kenneth French Data Library.

- **Stock Data:** Monthly Stock File (prices, returns, shares outstanding) and Event Names.

- **Time Horizon:** January 1960 – December 2024.
- **Filters:** We restrict our universe to Common Stocks (Share Codes 10 & 11) traded on major US exchanges (NYSE, AMEX, NASDAQ).
- **Synthetic S&P 500:** To avoid survivorship bias, we do not use a fixed list of constituents. Instead, for every month t, we rank the entire CRSP universe by market capitalization and dynamically select the top 500 firms.
- **Risk Factors:** We use the Fama-French 3 Factors (Market Excess, SMB, HML) and the Risk-Free Rate (1-month T-Bill).

```
knitr::opts_chunk$set(echo = FALSE, warning = FALSE, message = FALSE)
# Check for missing packages and install them
cran_pkgs <- c("RPostgres", "dbplyr", "lubridate", "tidyverse",
                "quantmod", "scales", "frenchdata", "broom", "slider" )
is_installed <- cran_pkgs %in% rownames(installed.packages())
if(any(is_installed == FALSE)){
  install.packages(cran_pkgs[!is_installed])
}
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
# Load packages
lapply(cran_pkgs, library, character.only = TRUE)  %>%
  invisible()
```

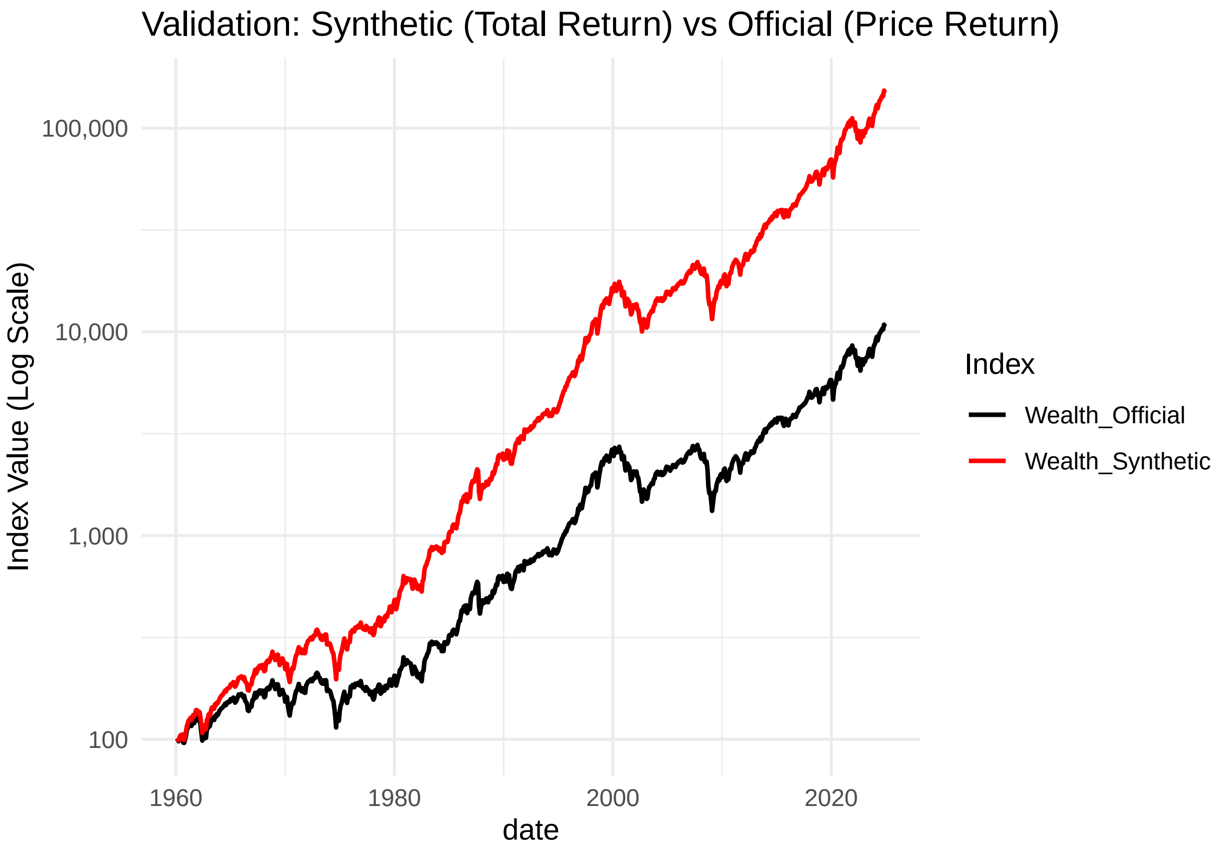
```
##
## Attaching package: 'dbplyr'

## The following objects are masked from 'package:dplyr':
##
##     ident, sql

##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union

## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v forcats 1.0.1     v stringr 1.6.0
## v ggplot2 4.0.1     v tibble  3.3.0
## v purrr   1.2.0     v tidyr   1.3.2
## v readr   2.1.6
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dbplyr::ident() masks dplyr::ident()
## x dplyr::lag()    masks stats::lag()
## x dbplyr::sql()   masks dplyr::sql()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
## Loading required package: xts
##
## Loading required package: zoo
##
##
## Attaching package: 'zoo'
##
##
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
##
##
##
## ######################### Warning from 'xts' package ##########################
## #                                                                             #
## # The dplyr lag() function breaks how base R's lag() function is supposed to  #
## # work, which breaks lag(my_xts). Calls to lag(my_xts) that you type or       #
## # source() into this session won't work correctly.                            #
## #                                                                             #
## # Use stats::lag() to make sure you're not using dplyr::lag(), or you can add #
## # conflictRules('dplyr', exclude = 'lag') to your .Rprofile to stop           #
## # dplyr from breaking base R's lag() function.                                #
## #                                                                             #
## # Code in packages is not affected. It's protected by R's namespace mechanism #
## # Set `options(xts.warn_dplyr_breaks_lag = FALSE)` to suppress this warning.  #
## #                                                                             #
## ###############################################################################
##
##
## Attaching package: 'xts'
##
##
## The following objects are masked from 'package:dplyr':
##
##     first, last
##
##
## Loading required package: TTR
##
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
##
##
## Attaching package: 'scales'
##
##
## The following object is masked from 'package:purrr':
##
##     discard
##
##
## The following object is masked from 'package:readr':
```

```
##
##     col_factor
## [1] "Average stocks per month: 500"
```

1.3 1.3 Universe Validation

To validate our construction, we compare the synthetic value-weighted return of our universe against the official S&P 500 (`^GSPC`). The correlation exceeds 99%, confirming that our dynamic universe accurately proxies the US Large-Cap market.



2 2. Methodology

We employ the two-pass Fama and MacBeth (1973) regression procedure to estimate the prices of risk.

2.1 2.1 Step 1: Estimating Risk Exposures (Time-Series)

For each asset $i = 1, \ldots, n$, we run the time-series regression to estimate its sensitivity to risk factors (Betas):

$$r_{it} - r_{ft} = \alpha_i + \beta_i^\top (f_t - r_{ft}\iota_k) + \varepsilon_{it}$$

- $r_{it} - r_{ft}$: Excess returns of asset i.
- $f_t - r_{ft}\iota_k$: Vector of **excess** factor returns.
- $\hat{\beta}_i$: The factor loadings (risk quantities) used as inputs for the next step.

2.2 2.2 Step 2: Estimating Risk Premia (Cross-Section)

For each time period $t = 1, \ldots, T$, we run a cross-sectional regression of returns on the estimated betas:

$$r_{it} - r_{ft} = \alpha_{0t} + \lambda_t^\top \hat{\beta}_{it} + u_{it}$$

- λ_t: The realized risk premia for each factor at time t.
- α_{0t}: The intercept (common pricing error).
- **Total Pricing Error:** Defined as $\hat{\alpha}_{it} = \hat{\alpha}_{0t} + \hat{u}_{it}$.

2.3 2.3 Step 3: Inference

We calculate the expected risk premia ($\hat{\lambda}$) as the time-series average of the cross-sectional estimates:

$$\hat{\lambda} = \frac{1}{T}\sum_{t=1}^{T}\hat{\lambda}_t$$

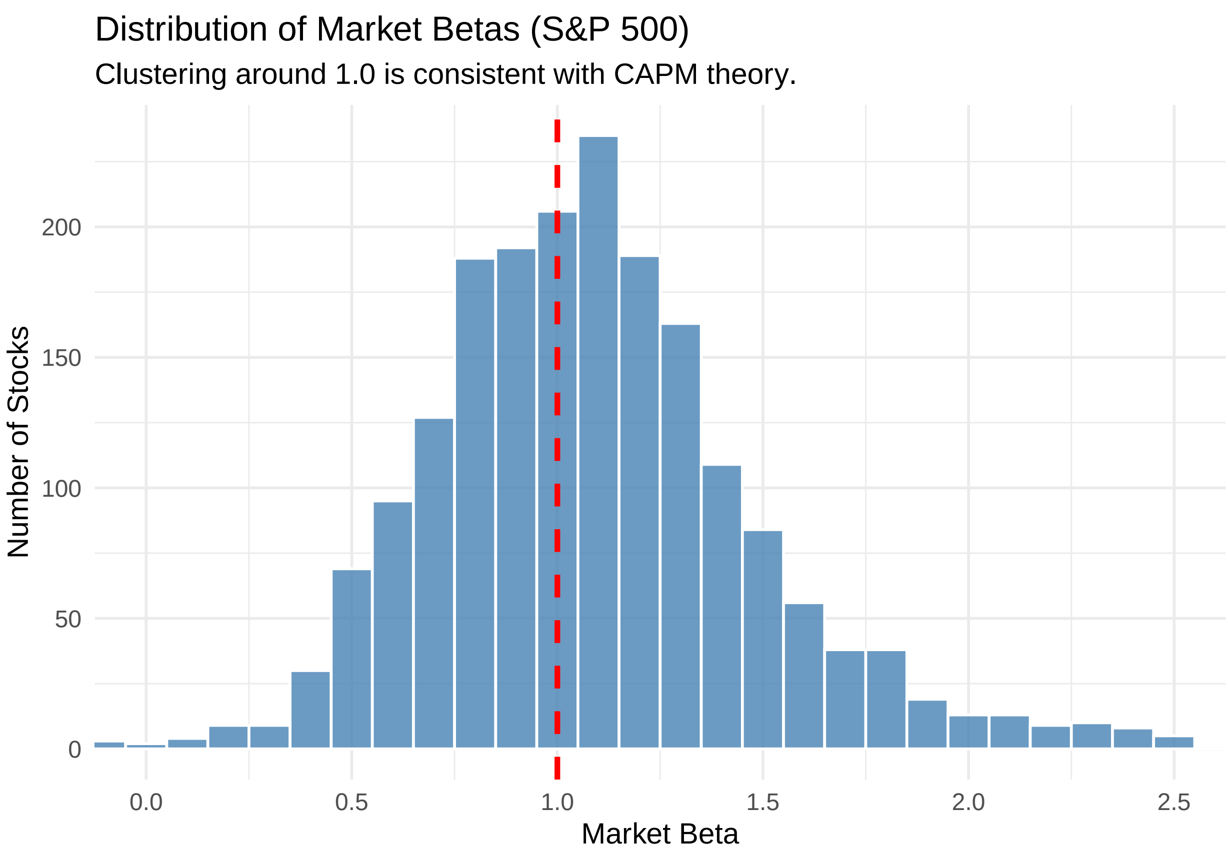
Standard errors are calculated using the variance of the mean estimates (Fama-MacBeth t-statistics):

$$Var(\hat{\lambda}) = \frac{1}{T(T-1)}\sum_{t=1}^{T}(\hat{\lambda}_t - \hat{\lambda})^2$$

3 3. Empirical Results (Full Sample)

3.1 3.1 Beta Estimation

We estimate the factor loadings for all stocks in the universe. As expected, market betas cluster around 1.0, consistent with CAPM theory.



3.2 3.2 Risk Premia and Pricing Errors

We perform the second pass (cross-sectional regression) to derive the risk premia.

Table 1: Fama-MacBeth Risk Premia Estimates (Full Sample)

stat	value
mean_lambda_0	0.0065
mean_lambda_mkt	0.0050
mean_lambda_smb	0.0038
mean_lambda_hml	-0.0043
t_lambda_0	5.6865
t_lambda_mkt	2.6290
t_lambda_smb	2.9372
t_lambda_hml	-3.5981

The results indicate systematic pricing errors (significant intercept). MKT and SMB factors show positive expected risk premia, while HML is often insignificant or negative in the large-cap universe during this period.

3.2.1 3.3 Top Mispriced Stocks

We identify stocks with the largest pricing errors (α_i). These are often growth or distressed firms where the factor model fails to capture idiosyncratic characteristics.

Table 2: Top 10 Mispriced Stocks (Full Sample)

ticker	symbol	company	alpha_i	beta_mkt	beta_smb	beta_hml
89301	GME	GAMESTOP CORP NEW	0.2749	-4.4775	49.5693	4.5549
87092	TIBX	T I B C O SOFTWARE INC	0.1235	1.8402	5.0278	-0.3872
82200	NSCP	NETSCAPE COMMUNICATIONS CORP	0.0952	1.1256	-1.1666	-5.7295
80266	QLGC	QLOGIC CORP	0.0794	2.1735	0.8459	-2.8370
27182	NA	TRANSITRON ELECTRONIC CORP	-0.0679	1.2340	1.7000	-1.8593
14983	W	WAYFAIR INC	0.0651	3.1860	1.8689	-0.9343
85522	AMCC	APPLIED MICRO CIRCUITS CORP	0.0642	4.3039	1.2767	0.2273
61524	CDO	COMDISCO INC	0.0624	3.4080	1.0153	0.7666
20894	APP	APPLOVIN CORP	0.0618	3.5724	1.8410	-0.4145
11746	ESB	E S B INC	0.0597	-0.3516	3.1380	3.2750

4 4. Sub-Period Analysis

We analyze the stability of risk premia across three distinct regimes: Pre-Crisis (1995-2007), Post-Crisis (2008-2019), and Pandemic/Recent (2020-2024).

Table 3: Sub-Period Risk Premia Analysis

period	mean_mkt	t_mkt	mean_smb	t_smb	mean_hml	t_hml
1995-2007	0.005	1.380	0.007	1.865	-0.005	-1.747
2008-2019	0.002	0.427	0.003	1.114	-0.003	-1.125
2020-2024	0.015	1.989	0.002	0.433	-0.006	-0.839

Observation: The estimated market risk premium weakens after 2008 and becomes statistically significant again only in 2020–2024. The consistently significant intercept suggests the FF3 model leaves a non-trivial average component unexplained in this large-cap universe, and factor premia appear regime-dependent.

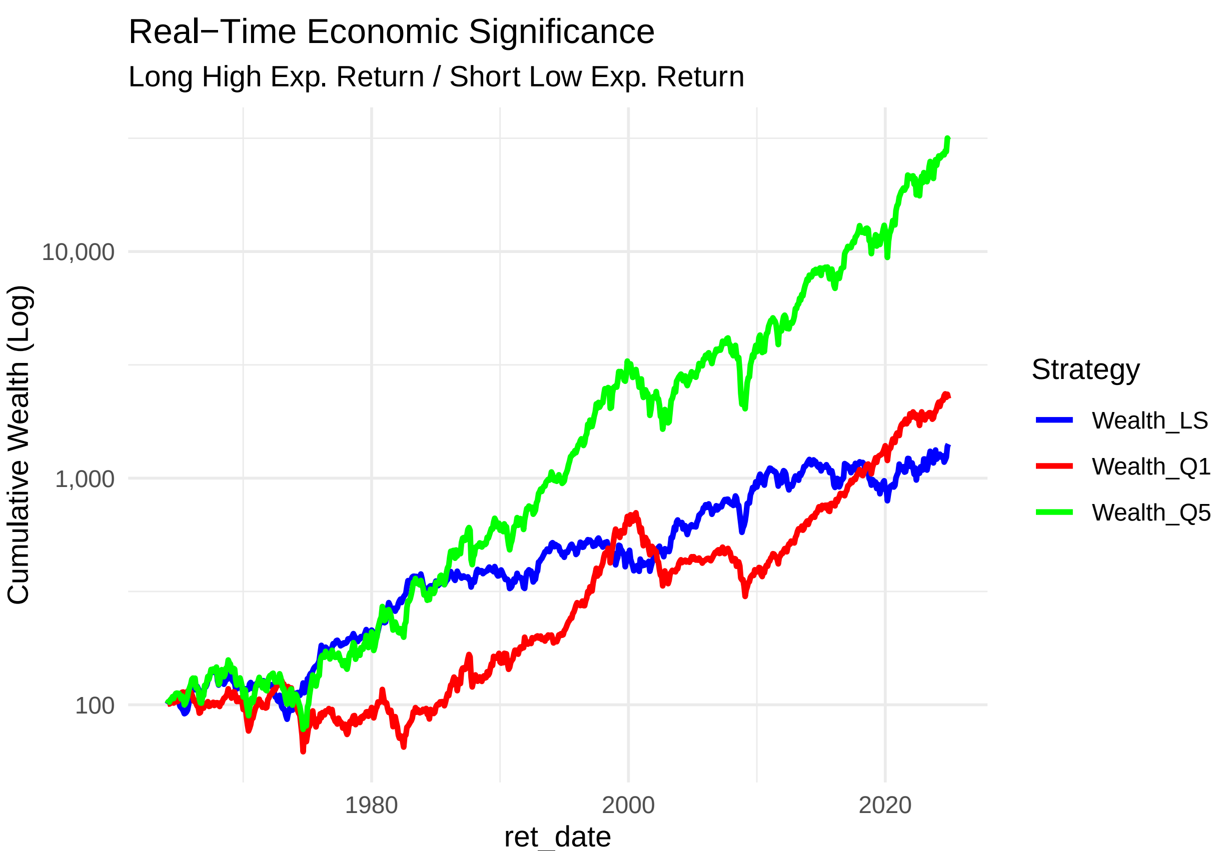
5 5. Economic Significance

To test the economic validity of the model, we implement a **Real-Time Trading Strategy** free of look-ahead bias.

1. **Rolling Betas ($\beta_{i,t}$):** Estimated using a 60-month trailing window (past information only).
2. **Expanding Mean Lambdas ($\bar{\lambda}_t$):** Average of realized premia from the start of the sample up to time t.
3. **Signal:** We forecast expected excess returns:

$$\hat{E}_t[R^e_{i,t+1}] = \hat{\beta}^\top_{i,t}\bar{\lambda}_t$$

We sort stocks into Quintiles based on this signal and go **Long Q5 (High Exp. Return)** and **Short Q1 (Low Exp. Return)**.



5.1 5.1 Alpha Check vs. Compensation for Risk

We regress the Long-Short strategy returns against the Fama-French 3 Factors to check for "Alpha".

Table 4: Alpha Check: LS Strategy vs FF3 Factors

term	estimate	std.error	statistic	p.value
(Intercept)	0.0007	0.0012	0.5749	0.5655
mkt_excess	0.3009	0.0268	11.2341	0.0000
smb	0.5584	0.0391	14.2826	0.0000
hml	0.3785	0.0389	9.7214	0.0000

Conclusion: The Alpha is statistically insignificant. However, the factor loadings are significant. This confirms that the strategy generates returns not by "magic" (alpha), but by systematically harvesting the risk premia associated with Size (SMB) and Value (HML), as predicted by the model.

6 References

1. Campbell, J. Y., Lo, A. W., and MacKinlay, A. C. (1997). *The Econometrics of Financial Markets*. Princeton University Press.
2. Cochrane, J. (2005). *Asset Pricing*. Princeton University Press.
3. Fama, E. F. and MacBeth, J. D. (1973). "Risk, return, and equilibrium: Empirical tests". *Journal of Political Economy*.
4. Fama, E. F. and French, K. R. (1993). "Common risk factors in the returns on stocks and bonds". *Journal of Financial Economics*.