Practice of QCF

Report

Choosing Factors

By Eugene F. Fama, Kenneth R. French

Submitted by:

Rohan Thanki
Mayank Jain
Akarsh Srivastava

Introduction

As part of the Practice of QCF project, we have decided to work on research and analysis of Factors. We are referring to the research paper, ‘Choosing Factors’ by Eugene F. Fama and Kenneth R. French.

The research paper develops insights into the efficacy of using the maximum squared Sharpe ratio for model factors as a metric for ranking asset pricing models. As per the paper, we work on both nested and non-nested models.

We analyze various factor models to try to conclude which model is the “best”. This will be the model that “best” explain the cross-section of stock returns.

We work on multiple nested models, which are mainly:

- The capital asset pricing model
- The three-factor model of Fama and French (1993)
- The five-factor extension in Fama and French (2015)
- A six-factor model that adds a momentum factor

The non-nested models examine three issues about factor choice in the six-factor model:

- cash profitability versus operating profitability as the variable used to construct profitability factors
- long-short spread factors versus excess return factors
- factors that use small or big stocks versus factors that use both

Motivation

Considering the large number of factors that are available to be included in a model, choosing among competing models is a significant challenge.

The paper Choosing Factors by Eugene Fama and Kenneth French discusses about two approaches to tackle the challenge,

- The left-hand-side (LHS) approach judges competing models on the intercepts (unexplained average returns) they leave in time-series regression to explain excess returns on sets of LHS portfolios.
- The right-hand-side (RHS) approach uses spanning regressions to judge whether individual factors contribute to the explanation of average returns provided by a model.

We try to implement the approaches discussed in the paper and analyze the results obtained. In our analysis we take the historical data from July 1963 to June 2020.

Data Sources

COMPUSTAT

Compustat Fundamentals provides standardized North American and global financial statement and market data for over 80,000 active and inactive publicly traded companies as far back as 1950 and point-in-time snapshots beginning in 1987.

CRSP

The Center for Research in Security Prices (CRSP) database contains end-of-day and month-end price on primary listings for the NYSE, NYSE MKT, NASDAQ, and Arca exchanges along with basic market indices.

Risk Free rate

Fama French library was used for 1 month T-bill rate

Data pre-processing

In our computation, we only used CUSIPs data (common to both CRSP and COMPUSTAT) for memory efficient approach

In order to prepare the data for further analysis, we performed data pre-processing on Compustat and CRSP.

COMPUSTAT

- Removing duplicate entries based on Industry formats
- Extracting/Selecting required columns only
- Renaming the CUSIPs - Taking the first 8 characters
- Additional filteration: ‘indfmt’ to ‘INDL’ and ‘STD’; ‘popsrc’ to ‘D’; ‘consol’ to ‘C’; ‘scf’ to less than 4 and equal to 7
- Removing rows with negative book value of equity

CRSP

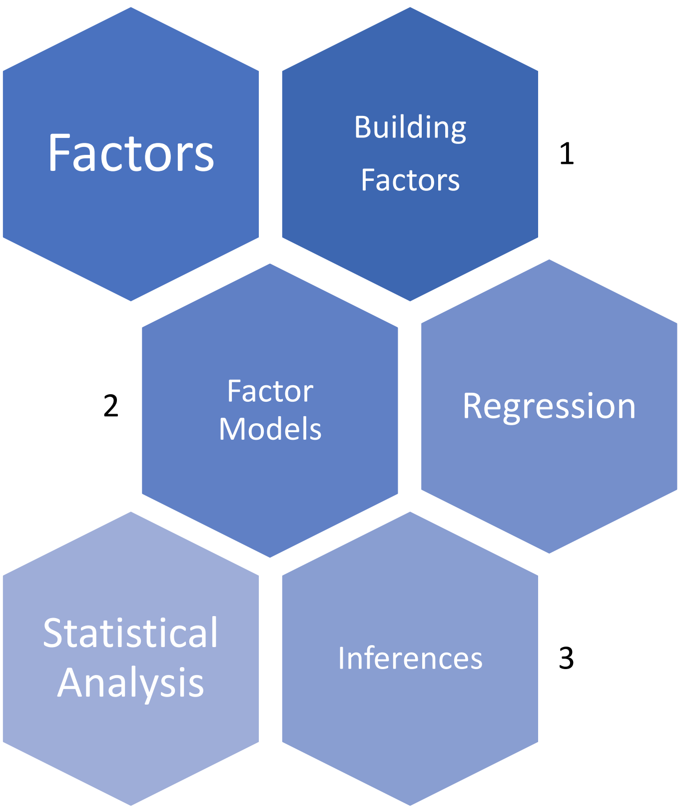
- Share Code as 10 and 11 (ordinary common with no special status found/necessary)
- RET with ‘B’ set to -1 indicating bankruptcy (i.e., 100% drop in share price)
- Converted PRC to numeric absolute (only positive) values
- Converted Date to datetime object
- Computed Market Capitalization (PRC x SHROUT) and converting to absolute value
- Renaming the CUSIPs - Taking the first 8 characters
- Extracting/Selecting required columns only

In order to tackle look-ahead bias, we lagged the fundamental variables by 6 months and price variables by 1 month.

Implementing Choosing Factors

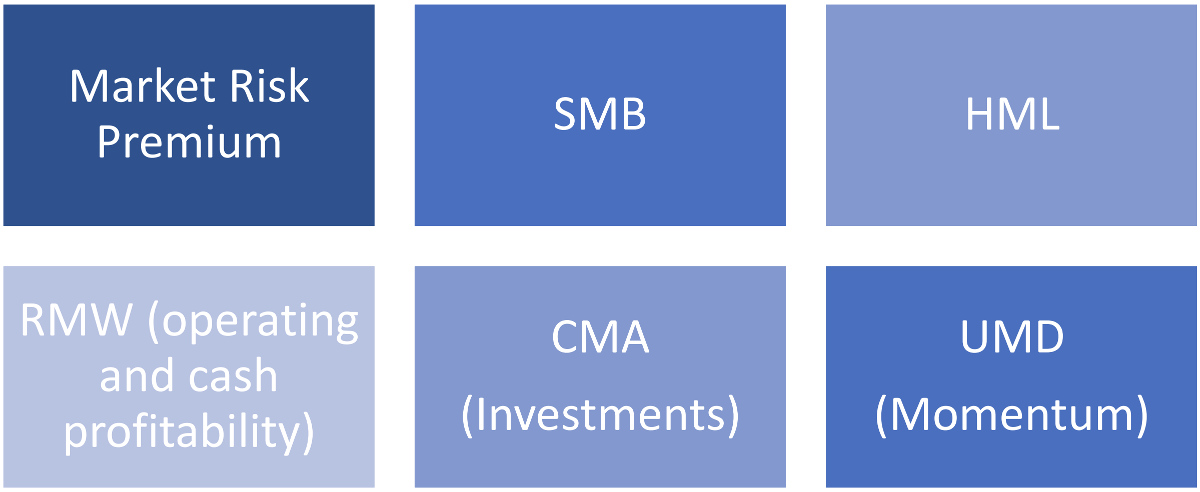
There are multiple parts in the implementation of the paper. We have the divided the implementation into 3 parts.

- In the first part, we create a sound understanding about the different factors used in the analysis and build the factors using the Compustat and CRSP data. We performed an initial experiment to familiarize with the process of factor building (discussed in detail ahead)
- In the second part, we create different factor models and perform regressions on the model as suggested in the paper.
- In the last part, we perform statistical analysis of the results obtained from the regression and derive inferences using the analysis.



Factors

Assessing the paper and the experiments performed, we need below six factors to experiment with different asset-pricing models.



Market Risk Premium:
The market risk premium (MRP) is the difference between the expected return on a market portfolio and the risk-free rate.

SMB:
Small minus big (SMB) is a size premium. SMB factor in the Fama/French stock pricing model that says smaller companies outperform larger ones over the long-term. It represents the spread in returns between smaller companies and larger companies.

HML:
High Minus Low (HML) is a value premium. HML represents the spread in returns between companies with a high book-to-market value ratio and companies with a low book-to-market value ratio

RMW:
RMW (Robust Minus Weak) is the profitability factor. It is the difference between the returns of firms with robust (high) and weak (low) operating profitability.

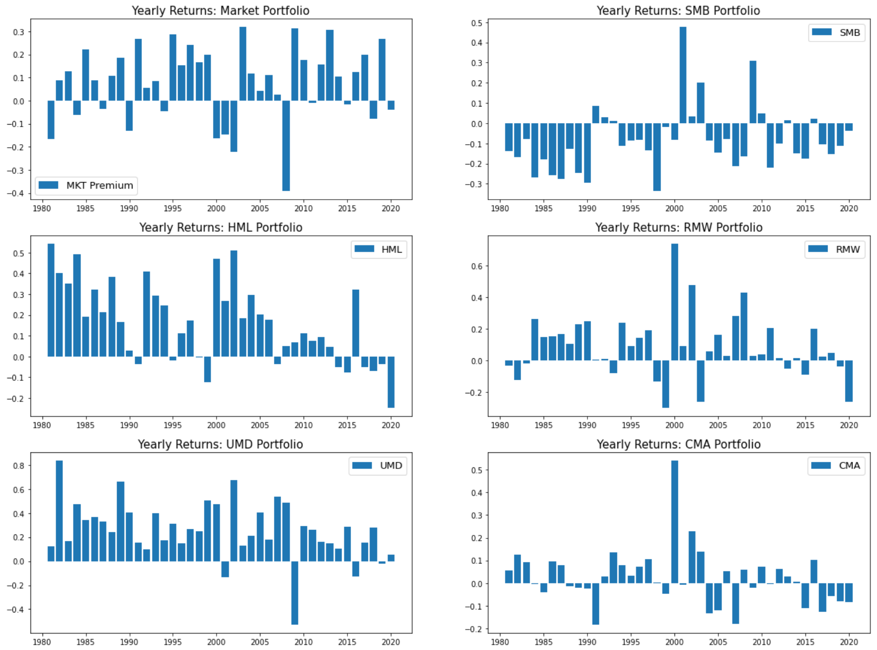
CMA:

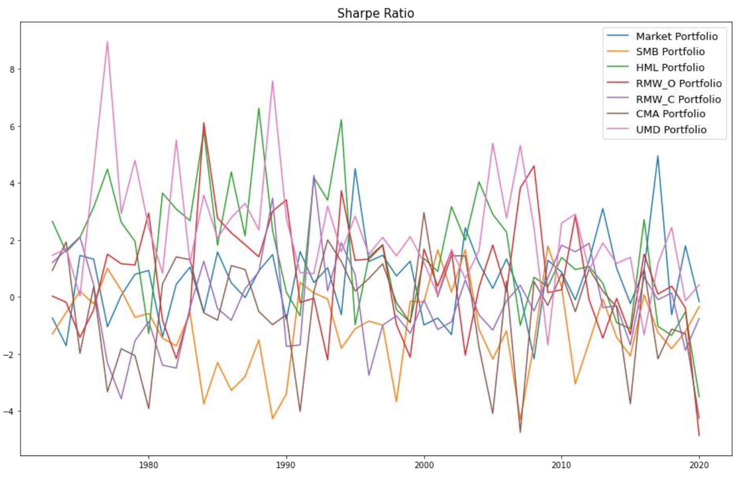
CMA (Conservative Minus Aggressive) is the investment factor. It is the difference between the returns of firms that invest conservatively and firms that invest aggressively.

UMD:

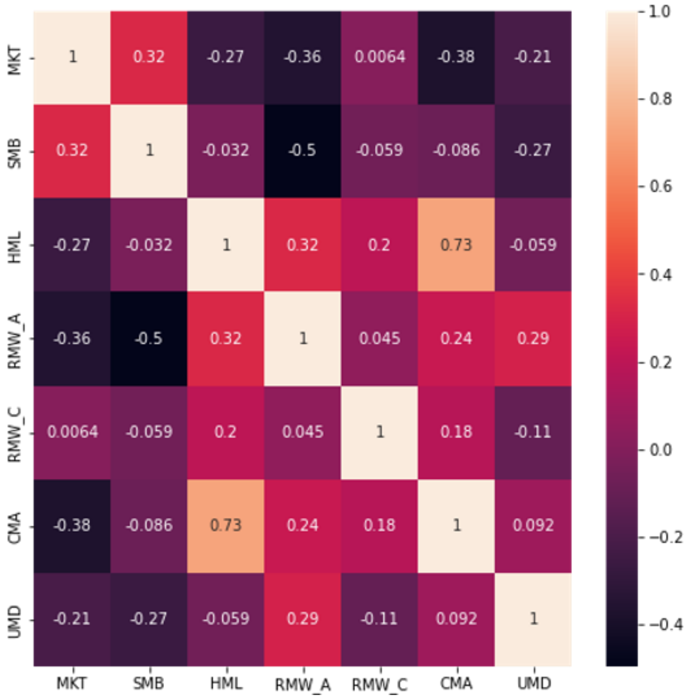
Momentum (UMD) is the momentum factor. It is difference between the return of highest performing stocks minus the return of the lowest performing stocks.

Portfolio Analytics:





Factor return correlations plot:



Building Factors

Exploratory data analysis

As an exploratory analysis and to get acquainted with the process of factor building, we worked on the activity of building factors from scratch. For this exercise, we used month level company return time-series and fundamentals data from CRSP and COMPUSTAT data. In that, we used the top 600 companies by market capitalization each month from Dec 2001 to Dec 2020.

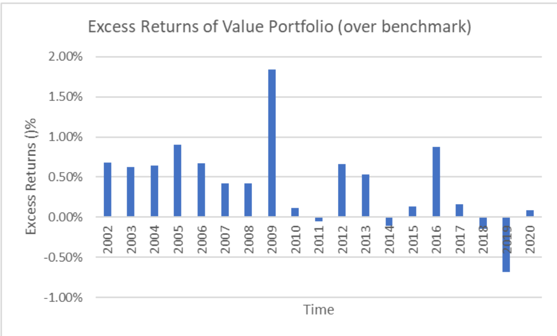
We form six style factor portfolios Profit margin, small cap, large cap, high momentum, value, and growth which are computed based on following methodology:

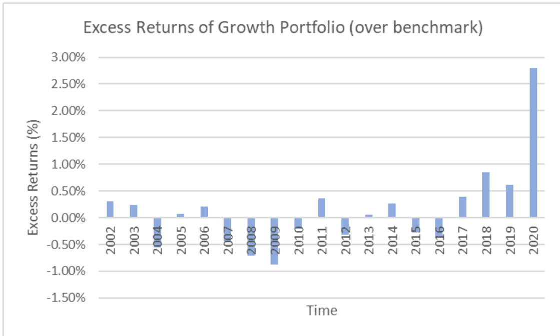
- The Profit margin is calculated as the top decile stocks each year based on the ratio of net income by the sales
- The Small factor is calculated as the bottom decile of stocks every month based on their mkt cap

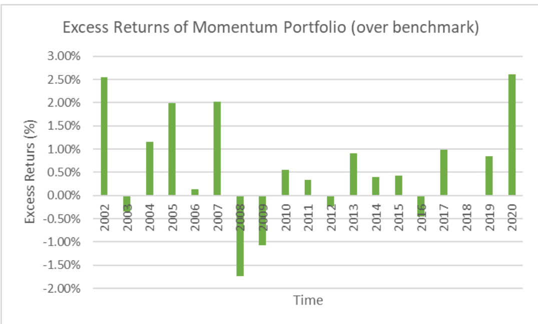
- The Big factor is calculated as the top decile of stocks every month based on their market cap
- The High Momentum factor is calculated as the top decile of stocks every month based on their trailing 3-year cumulative returns
- The Value factor is calculated as the bottom decile of stocks every month based on the ratio of latest market capitalization to the trailing book value
- The Growth factor is calculated as the top decile of stocks every month based on the ratio of latest market capitalization to the trailing book value

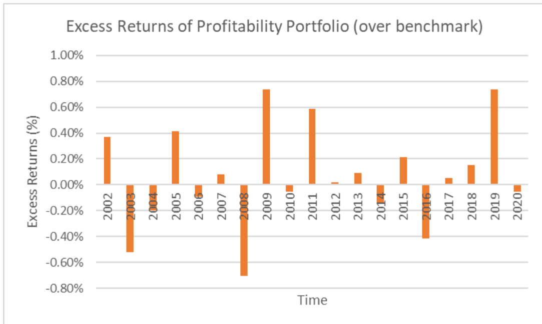
Exploratory data analysis results:

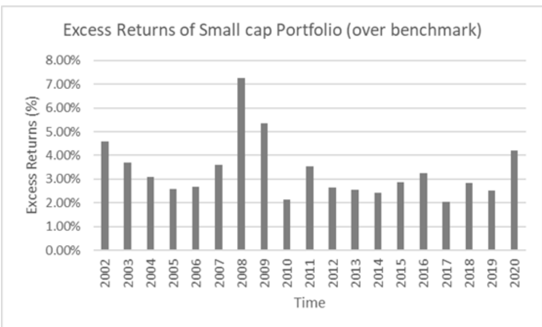
The following are excess return plots of each factor mimicking portfolio (over benchmark):

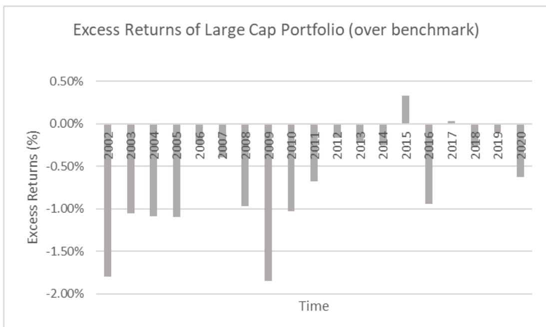












Factor Building Methodology

For implementation of the approaches, we have built multiple factors to use in the asset-pricing models.

6 Portfolios Based on Size (Market Cap) and Value (Book Value to Market Cap Ratio)

	Bottom 30%	Middle 40%	Top 30%
Small Stocks	Small Growth	Small Neutral	Small Value
Big Stocks	Big Growth	Big Neutral	Big Value

6 Portfolios Based on Size (Market Cap) and Profitability (Gross Profit to Book Value)

	Bottom 30%	Middle 40%	Top 30%
Small Stocks	Small Robust	Small Neutral	Small Weak
Big Stocks	Big Robust	Big Neutral	Big Weak

6 Portfolios Based on Size (Market Cap) and Investment (Rate of Change of Assets)

	Bottom 30%	Middle 40%	Top 30%
Small Stocks	Small Conservative	Small Neutral	Small Aggressive
Big Stocks	Big Conservative	Big Neutral	Big Aggressive

6 Portfolios Based on Size (Market Cap) and Momentum (Previous 2 - 12 month Return)

	Bottom 30%	Middle 40%	Top 30%
Small Stocks	Small Up	Small Neutral	Small Down
Big Stocks	Big Up	Big Neutral	Big Down

Portfolios are formed on 30th June at market close each year. The portfolios formed on year t take the information until calendar year t-1 into account.

Profitability is computed in 2 ways:

- Operating profitability – This is revenues less COGS less SG&A less Interest Expense divided by Book Value of Equity
- Cash Profitability – This is operating profit adjusted for non-cash items such as changes in prepaid expense, accounts receivable, accounts payable, inventory, deferred revenue and accrued expenses divided by the Book Value of Equity

The 6 portfolios derived based on momentum are rebalanced at the end of each month at market close. The remaining portfolios are rebalanced annually on 30th June at market close.

Factor Models and Regression

1. Summary Statistics

We first compute the table 1 presented in the Choosing Factors.

In table 1, we have computed the descriptive summary statistics where we have computed average monthly factor returns for combination of factor models (July 1963 - June 2020)

	Mkt	SMB	S-F	B-F
Market and size factors	0.008887	0.00103	0.006799	0.005769

Table-1A: Average Returns

	Mkt	SMB	S-F	B-F
Market and size factors	4.897503	0.836973	2.716503	3.097992

Table-1A: T-statistics

Panel A shows summary statistics for Mkt, SMB, and the excess return factors S–F and B–F constructed from the long and short ends of SMB.

	Value	Prof_O	Prof_C	Inv	Mom
HML, RMW, CMA, UMD	0.002760	0.000058	-0.000665	0.003670	0.004855
HML_S, RMW_S, CMA_S, UMD_S	0.004119	0.000736	-0.001115	0.004903	0.006352
HML_B, RMW_B, CMA_B, UMD_B	0.001401	-0.000619	-0.000215	0.002437	0.003357
H-F, R-F, C-F, U-F	0.007883	0.005828	0.005422	0.008019	0.008673
H_S-F, R_S-F, C_S-F, U_S-F	0.008900	0.006456	0.005924	0.008954	0.009947
H_B-F, R_B-F, C_B-F, U_B-F	0.006866	0.005199	0.004920	0.007084	0.007400
L-F, W-F, A-F, D-F	0.005123	0.005769	0.006087	0.004349	0.003819
L_S-F, W_S-F, A_S-F, D_S-F	0.004780	0.005720	0.007040	0.004051	0.003595
L_B-F, W_B-F, A_B-F, D_B-F	0.005465	0.005818	0.005135	0.004647	0.004042
HML_SMB, RMW_SMB, CMA_SMB, UMD_SMB	0.002719	0.001355	-0.000900	0.002466	0.002995
H_S-H_B, R_S-R_B, C_S-C_B, U_S-U_B	0.002034	0.001257	0.001004	0.001870	0.002547
L_S-L_B, W_S-W_B, A_S-A_B, D_S-D_B	-0.000685	-0.000098	0.001904	-0.000596	-0.000447

Table-1B: Average Returns

	Value	Prof_O	Prof_C	Inv	Mom
HML, RMW, CMA, UMD	2.160028	0.059267	-0.640802	4.238070	2.658809
HML_S, RMW_S, CMA_S, UMD_S	2.901987	0.582172	-0.728592	5.604443	3.523736
HML_B, RMW_B, CMA_B, UMD_B	1.005148	-0.582730	-0.165352	2.223347	1.645696
H-F, R-F, C-F, U-F	3.702452	2.924897	2.151041	3.818448	4.083019
H_S-F, R_S-F, C_S-F, U_S-F	3.719400	2.749436	1.879532	3.512754	4.068506
H_B-F, R_B-F, C_B-F, U_B-F	3.413509	2.823840	2.262780	3.846300	3.785628
L-F, W-F, A-F, D-F	2.199482	2.326413	2.489660	1.804524	1.379459
L_S-F, W_S-F, A_S-F, D_S-F	1.640787	1.957226	2.458630	1.450803	1.167770
L_B-F, W_B-F, A_B-F, D_B-F	2.819400	2.624783	2.154254	2.122108	1.529018
HML_SMB, RMW_SMB, CMA_SMB, UMD_SMB	2.311070	1.080286	-0.464446	2.551668	2.455906
H_S-H_B, R_S-R_B, C_S-C_B, U_S-U_B	1.712686	0.905398	0.507797	1.278967	2.048274
L_S-L_B, W_S-W_B, A_S-A_B, D_S-D_B	-0.409611	-0.064341	0.970972	-0.426826	-0.296327

Table-1B: T-statistics

Panel B shows summary statistics for value (Value), operating profitability (Prof_O), cash profitability (Prof_C), investment (Inv), and momentum (Mom) factors, grouped as they are in the models.

2. <u>Spanning tests for nested models (July 1963–June 2020)</u>

We now compute table 2 presented in the Choosing Factors.
Table 2 tests:

- whether the excess market return, Mkt, spans the size spread factor, SMB, and the value spread factor, HML
- whether Mkt, SMB, and HML span the investment spread factor, CMA, and a profitability spread factor, RMWO or RMWC
- whether Mkt, SMB, HML, CMA, and RMWO or RMWC span the momentum spread factor, UMD.

The tests press on the intercepts from spanning regressions of additional factors on base factors: SMB and HML regressed on Mkt; CMA and RMWO or RMWC on Mkt, SMB, and HML; and UMD on Mkt, SMB, HML, CMA, and RMWO or RMWC.

	Int	Mkt	SMB	HML	RMWO	RMWC	CMA
SMB	-0.000084	0.218492					
HML	0.003733	-0.191030					
RMWO	0.000263	-0.079025	-0.352946	0.203241			
RMWC	-0.001396	0.056962	-0.070795	0.186152			
CMA	0.002887	-0.094021	-0.000613	0.457291			
UMD	0.005016	-0.097562	-0.147251	-0.484568	0.471852		0.489567
UMD	0.004962	-0.126808	-0.327517	-0.338099		-0.193628	0.457803

Table-2A: Coefficients (Spanning Regression)

	Int	Mkt	SMB	HML	RMWO	RMWC	CMA
SMB	-0.071262	8.513467					
HML	3.014806	-7.046405					
RMWO	0.327679	-4.114020	-12.898321	7.845851			
RMWC	-1.360178	2.321710	-2.025577	5.626228			
CMA	4.987410	-6.795210	-0.031122	24.50758			
UMD	2.890386	-2.277524	-2.255618	-6.055111	5.527072		4.130694
UMD	2.803951	-2.939294	-5.538547	-4.290175		-2.849655	3.799677

Table-2B: t-values (Spanning Regression)

	R2	s(e)
SMB	0.104515	2.251060
HML	0.074035	2.251060
RMWO	0.355150	5.067712
RMWC	0.052881	5.067712
CMA	0.567104	5.067712
UMD_O	0.162171	12.655025
UMD_C	0.132112	12.928903

Table-2C: R-squared & S(e)

The table above shows regression coefficients and t-statistics for the coefficients.

3. <u>Comparing six-factor models: Operating profitability or Cash profitability</u>

<u>Squared Sharpe Ratio:</u>

The maximum squared Sharpe Ratio, defined as the closed form solution, is the dot product of transpose of vector of intercept (i.e., unexplained return of the model), inverse of covariance of residuals, and vector of intercepts from the regression of asset returns against the factor mimicking portfolios.

$$\boxed{Sh^2(f) = a^T . \Sigma^{-1} . a}$$

In this part we compute the table 3 presented in the Choosing Factors.

- In table 3 we analyse different factor models that include operating profitability or cash profitability from 1963-2020 (684 months)
- We assess Actual, Average and Median squared Sharpe ratio, $Sh^2(f)$ from Full-sample(FS), In-sample(IS) and Out-sample(OS) simulation runs (10,000)
- FS simulations estimate squared sharpe ratio from random samples (with replacement) of 684 months from July 1963–June 2020
- IS and OS simulations split the 684 sample months into 342 adjacent pairs: months (1, 2), (3, 4), … (341, 342).
- A simulation run draws a random sample with replacement of 342 pairs. The IS simulation run chooses a month randomly from each pair in the run.
- We calculate IS for all models on that sample of months. A model's IS identifies weights for factors in the IS tangency portfolio for the factors.
- These weights and the unused months of the simulation pairs produce an OS estimate of the Sharpe ratio for the IS tangency portfolio.

		Actual	(Full-sample, Average)	(Full-sample, Median)	(In-sample, Average)	(In-sample, Median)	(Out-sample, Average)	(Out-sample, Median)
Six-factor Operating Profitability	**Mkt, SMB, HML, RMW_O , CMA, UMD**	0.134173	0.005367	0.001200	0.008850	0.004031	0.008171	0.003790
	Mkt, SMB, HML_S, RMW_O_S, CMA_S, UMD_S	0.121574	0.004390	0.001789	0.006761	0.002252	0.006391	0.001922
	Mkt, S–F, H–F, R_O–F, C–F, U–F	0.160432	0.004633	0.001114	0.008826	0.003697	0.008186	0.003583
	Mkt, S–F, H_S–F, R_O_S–F, C_S–F, U_S–F	0.161620	0.005852	0.001897	0.008276	0.003336	0.008530	0.002564
	Mkt, S–F, L–F, W_O–F, A–F, D–F	0.144983	0.005590	0.001504	0.006682	0.003289	0.007195	0.002469
	Mkt, S–F, L_S–F, W_O_S–F, A_S–F, D_S–F	0.124277	0.004166	0.000679	0.005709	0.001766	0.005315	0.002093
Six-factor Cash Profitability	**Mkt, SMB, HML, RMW_C , CMA, UMD**	0.133623	0.005281	0.001284	0.008921	0.003870	0.008433	0.003091
	Mkt, SMB, HML_S, RMW_C_S, CMA_S, UMD_S	0.119755	0.004069	0.001416	0.007237	0.002812	0.006935	0.002575
	Mkt, S–F, H–F, R_C–F, C–F, U–F	0.163951	0.004761	0.001003	0.008276	0.003592	0.008231	0.002931
	Mkt, S–F, H_S–F, R_C_S–F, C_S–F, U_S–F	0.160588	0.005933	0.002457	0.008735	0.003197	0.009376	0.003161
	Mkt, S–F, L–F, W_C–F, A–F, D–F	0.145070	0.005549	0.001512	0.007049	0.002850	0.007639	0.003171
	Mkt, S–F, L_S–F, W_C_S–F, A_S–F, D_S–F	0.125043	0.004216	0.000809	0.005451	0.002381	0.005260	0.002149

Table-3A: Levels of Squared Sharpe ratio

- Now we evaluate the Actual $Sh^2(f_C)-Sh^2(f_O)$, the difference between $Sh^2(f)$ when the model uses a CP or OP factor, the means and medians of $Sh^2(f_C)-Sh^2(f_O)$ from 10,000 FS, IS, and OS simulation runs, and the percent of simulation runs in which the model that uses the OP factor beats the model that uses the CP factor

		Actual	(Full-sample, Average)	(Full-sample, Median)	(In-sample, Average)	(In-sample, Median)	(Out-sample, Average)	(Out-sample, Median)
Cash - Operating Profitability	Mkt, SMB, HML, RMW, CMA, UMD	0.000551	0.000086	-0.000084	-0.000071	0.000161	-0.000262	0.000699
	Mkt, SMB, HML_S, RMW_S, CMA_S, UMD_S	0.001818	0.000320	0.000373	-0.000476	-0.000560	-0.000544	-0.000653
	Mkt, S–F, H–F, R–F, C–F, U–F	-0.003520	-0.000128	0.000111	0.000550	0.000106	-0.000045	0.000652
	Mkt, S–F, H_S–F, R_S–F, C_S–F, U_S–F	0.001032	-0.000081	-0.000560	-0.000458	0.000140	-0.000845	-0.000597
	Mkt, S–F, L–F, W–F, A–F, D–F	-0.000086	0.000041	-0.000008	-0.000367	0.000439	-0.000444	-0.000702
	Mkt, S–F, L_S–F, W_S–F, A_S–F, D_S–F	-0.000766	-0.000050	-0.000131	0.000258	-0.000615	0.000055	-0.000056

Table-3B: Differences between $Sh^2(f)$ for cash and operating profitability factors

4. Distributions of differences between $Sh^2(f)$, column model minus row model

For the next part, we understand the purpose of using max squared sharpe ratio.

Max Squared Sharpe Ratio of Intercepts:

- The maximum squared Sharpe ratio of intercepts is a performance metric developed by Barillas Shanken in 2016, which we use heavily in our analysis to rank factor models.
- When we run time-series regressions of test asset returns on factor returns, we seek the model that minimizes the maximum possible squared Sharpe ratio for the intercepts.
- Notations:
  - R: set of excess returns of test assets
  - f_i: factors of model i
  - f_{Ai}: union of factors of model i's competitors
  - Π_i: set of LHS returns for model I
  - Π: union of all factors of all models under consideration
  - a_i: vector of intercepts from regression of Π_i on fi
  - Σ_i: residual covariance matrix
- Formula for max squared Sharpe ratio of intercepts:

$$\boxed{Sh^2(f) = a^T.\Sigma^{-1}.a}$$

- The winner among competing models is the one that produces the smallest $Sh^2(a_i)$.

- GRS (1989) show that $a_i' \sum_i^{-1} a_i$ is the difference between the max squared Sharpe ratio one can construct from f_i and $\prod_i$ together and the max for f_i alone

$$Sh^2(a_i) = Sh^2(\prod_i , f_i) - Sh^2(f_i)$$

- Since $\prod_i$ includes the factors of all model i's competitors, the union of $\prod_i$ and f_i, which we call $\prod$, does not depend on i. Thus, the above equation becomes

$$Sh^2(a_i)=Sh^2(\prod) - Sh^2(f_i)$$

- If the goal is to minimize $Sh^2(a)$, the best model is the one whose factors have the highest $Sh^2(f)$.
- The BS argument that leads to $Sh^2(f)$ as the metric for judging asset pricing models implies that if $\prod$ spans the factors of competing models, inferences do not depend on what is in R, the non-factor assets in $\prod$

Next, in table 4 we analyse the median differences between $Sh^2(f)$ for a column model and a row model from 10,000 full-sample, in-sample, and out-of-sample simulation runs, and the percent of simulation runs in which the row model has higher $Sh^2(f)$ than the column model (% < 0).

The column models are the three that produce the highest sample actual $Sh^2(f)$ in Table 3A (from reference paper).

		Mkt, S–F, H–F, R_C–F, C–F, U–F		Mkt, S–F, H_S–F, R_C_S–F, C_S–F, U_S–F		Mkt, S–F, L–F, W_C–F, A–F, D–F	
		Average	Median	Average	Median	Average	Median
Full-Sample	Mkt, S–F, H–F, R_C–F, C–F, U–F			0.001173	0.001453	0.000788	0.000509
	Mkt, S–F, H_S–F, R_C_S–F, C_S–F, U_S–F	-0.001173	-0.001453			-0.000385	-0.000945
	Mkt, S–F, L–F, W_C–F, A–F, D–F	-0.000788	-0.000509	0.000385	0.000945		
	Mkt, SMB, HML, RMW_O , CMA, UMD	-0.000607	-0.000197	0.000566	0.001257	0.000181	0.000312
	Mkt, SMB, HML, RMW_C , CMA, UMD	-0.000521	-0.000281	0.000652	0.001172	0.000267	0.000228
	Mkt, SMB, HML_S, RMW_C_S, CMA_S, UMD_S	0.000691	-0.000413	0.001864	0.00104	0.001479	0.000096
	Mkt, S–F, L_S–F, W_C_S–F, A_S–F, D_S–F	0.000544	0.000194	0.001717	0.001647	0.001333	0.000703
In-Sample	Mkt, S–F, H–F, R_C–F, C–F, U–F			0.000459	-0.000395	-0.001227	-0.000741
	Mkt, S–F, H_S–F, R_C_S–F, C_S–F, U_S–F	-0.000459	0.000395			-0.001686	-0.000347
	Mkt, S–F, L–F, W_C–F, A–F, D–F	0.001227	0.000741	0.001686	0.000347		
	Mkt, SMB, HML, RMW_O , CMA, UMD	-0.000574	-0.000439	-0.000115	-0.000834	-0.001801	-0.001181
	Mkt, SMB, HML, RMW_C , CMA, UMD	-0.000645	-0.000279	-0.000186	-0.000673	-0.001872	-0.00102
	Mkt, SMB, HML_S, RMW_C_S, CMA_S, UMD_S	0.001039	0.00078	0.001498	0.000385	-0.000188	0.000038
	Mkt, S–F, L_S–F, W_C_S–F, A_S–F, D_S–F	0.002825	0.00121	0.003284	0.000816	0.001598	0.000469
Out-of-Sample	Mkt, S–F, H–F, R_C–F, C–F, U–F			0.001145	0.00023	-0.000592	0.00024
	Mkt, S–F, H_S–F, R_C_S–F, C_S–F, U_S–F	-0.001145	-0.00023			-0.001737	0.000011
	Mkt, S–F, L–F, W_C–F, A–F, D–F	0.000592	-0.00024	0.001737	-0.000011		
	Mkt, SMB, HML, RMW_O , CMA, UMD	0.00006	-0.000859	0.001205	-0.000629	-0.000532	-0.000619
	Mkt, SMB, HML, RMW_C , CMA, UMD	-0.000202	-0.00016	0.000943	0.00007	-0.000794	0.00008
	Mkt, SMB, HML_S, RMW_C_S, CMA_S, UMD_S	0.001296	0.000355	0.002441	0.000585	0.000704	0.000596
	Mkt, S–F, L_S–F, W_C_S–F, A_S–F, D_S–F	0.002971	0.000782	0.004116	0.001011	0.002379	0.001022

Table-4: Distributions of differences between $Sh^2(f)$, column model minus row model

5. Spanning regressions and marginal contributions to $Sh^2(f)$ for three models that produce highest $Sh^2(f)$ in Table 4

Table 5 shows that marginal contributions ($a^2/s^2(e)$) of factors to $Sh^2(f)$ depend on residual variances in spanning regressions as well as on regression intercepts. For example, in the two regressions that use spread factors, the intercepts for RMW_C and RMW_CS are less than those for Mkt, but the residual variances in the regressions for the profitability factors are also less than those in the Mkt regressions. As a result, the marginal contributions of RMW_C and RMW_CS to $Sh^2(f)$ are similar to those of Mkt.

Panel A and B:

		a	Mkt	SMB	HML	RMW_C	CMA	UMD	t(a)	R2	s(e)	Sh2(f)	a2/s2(e)
Panel A: Spanning regressions for combined spread factor model	Mkt	0.008033		0.398327	-0.067877	0.14694	-0.690702	-0.108897	4.963737	0.243905	13.294425	0.133623	3.651119e-07
	SMB	0.000284	0.204803		0.018602	-0.116949	0.074881	-0.144611	0.240229	0.148903	12.439014	0.133623	5.222591e-10
	HML	-0.000573	-0.020029	0.010676		0.076624	1.06036	-0.085675	-0.639895	0.547365	6.898266	0.133623	6.910997e-09
	RMW_C	-0.001401	0.059286	-0.091773	0.104771		0.153786	-0.067089	-1.338099	0.062269	11.550623	0.133623	1.470530e-08
	CMA	0.002628	-0.087745	0.018501	0.456508	0.048421		0.049944	4.541304	0.575683	6.549306	0.133623	1.609822e-07
	UMD	0.004962	-0.126808	-0.327517	-0.338099	-0.193628	0.457803		2.803951	0.125078	12.928903	0.133623	1.472990e-07

Table-5A: Spanning regressions for combined spread factor model

In Table 5A, SMB and HML shows that the intercept (unexplained return) of the model are not statistically significant (or equal to zero) with t statistics less than 95% critical t value. This indicates that returns explained by SMB and HML cannot be explained by other factor portfolios in 6 factor model.

		a	Mkt	SMB	HML_S	RMW_C_S	CMA_S	UMD_S	t(a)	R2	s(e)	Sh2(f)	a2/s2(e)
Panel B: Spanning regressions for small stock spread factor model	Mkt	0.008123		0.349826	-0.313752	0.146392	-0.301149	-0.071903	4.947950	0.242037	11.503892	0.119755	4.986405e-07
	SMB	0.000692	0.168267		-0.210979	-0.041622	0.303816	-0.186844	0.596388	0.201819	10.922635	0.119755	4.016284e-09
	HML_S	0.000300	-0.123004	-0.171959		0.097812	0.996617	-0.024273	0.286144	0.511474	7.074894	0.119755	1.796922e-09
	RMW_C_S	-0.000955	0.12694	-0.075034	0.216344		-0.103418	-0.175443	-0.612585	0.071062	10.614612	0.119755	8.088543e-09
	CMA_S	0.003162	-0.048444	0.101606	0.408932	-0.019185		0.027858	4.797038	0.472260	6.416875	0.119755	2.428723e-07
	UMD_S	0.006293	-0.079498	-0.429481	-0.068454	-0.223699	0.191472		3.612633	0.145643	11.119531	0.119755	3.202735e-07

Table-5B: Spanning regressions for small stock spread factor model

In table 5B, small stock 6 factor model, SMB, HML as well as Cash Profitability shows that the intercept (unexplained return) of the model are not statistically significant (or equal to zero) with t statistics less than 95% critical t value implying that these factors are able to explain the returns of the assets uniquely.

Marginal contributions of factors to $Sh^2(f)$ depend on residual variances in spanning regressions as well as on regression intercepts

- In the two regressions that use spread factors, RMW_C and RMW_C_S have very small intercepts, but the residual variances in the regressions for the profitability factors are also quite small relative to the Mkt factor.
- The same trend is also evident for UMD and UMD_S factors in explaining the marginal contribution to $Sh^2(f)$

The larger marginal contributions of CMA_S and UMD_S, relative to those of CMA and UMD, probably are because investment and momentum premiums are stronger for small stocks, which is less true for the profitability premium.

The intercepts for SMB are similar to those for RMW_C or RMW_C_S (although opposite in sign but close to zero), but residual variances are higher in the SMB regressions, and the marginal contributions of SMB to $Sh^2(f)$ are less than those of RMW_C or RMW_C_S.

The spanning regressions to explain UMD and UMD_S also illustrate the role of residual variances in marginal contributions to $Sh^2(f)$. The intercepts, 0.0049 and 0.0063 (Table 5), are close to the large average returns for UMD and UMD_S, 0.0048 and 0.0063 (Table 1), but the regression R^2 are only 0.12 and 0.14, lower than other factors in the spanning regression.

As a result, the residual variances for UMD and UMD_S are high and their marginal contributions to $Sh^2(f)$ for the spread models are only $1.47e^{-07}$ and $3.20e^{-07}$ (compared to others of the magnitude 10^{-8} to 10^{-10}).

<u>Panel C</u>

		a	Mkt	S_F	L_F	W_C_F	A_F	D_F	t(a)	R2	s(e)	Sh2(f)	a2/s2(e)
Panel C: Spanning regressions for small stock short factor model	Mkt	0.001760		-0.184733	0.228519	0.212168	0.530935	-0.047361	3.302794	0.917804	631.608255	0.119755	7.766889e-12
	S_F	0.002415	-0.281251		0.204204	0.302646	0.498369	0.199258	3.680387	0.933819	612.088026	0.119755	1.556922e-11
	L_F	0.000600	0.081053	0.047573		0.022156	0.882074	-0.048411	1.879855	0.982194	332.462314	0.119755	3.259279e-12
	W_C_F	0.000075	0.430149	0.403019	0.126645		-0.011511	0.126193	0.098425	0.907659	653.819307	0.119755	1.328052e-14
	A_F	-0.001116	0.155209	0.095693	0.726998	-0.00166		0.081115	-3.886673	0.986294	300.699650	0.119755	1.378008e-11
	D_F	-0.001560	-0.165947	0.458578	-0.478233	0.218094	0.972231		-1.552794	0.875496	669.307818	0.119755	5.430230e-12

Table-5C: Spanning regressions for small stock short factor model

In table 5C, again, for excess return spread factor model, Value, Cash Profitability and Investment premiums shows that the intercept (unexplained return) of the model are not statistically significant (or equal to zero) with t statistics less than 95% critical t value implying that these factors are able to explain the returns of the assets uniquely.

Spanning regression results for in Table 5C are different for the model that uses excess returns factors

- A surprising result is that Mkt factor which is a top marginal contributor to $Sh^2(f)$ in spread factor models, is redundant in the model that uses excess returns on the short ends of small stock spread factors.
- S–F, the excess return on the long end of SMB , makes by far the biggest marginal contribution to $Sh^2(f)$
- The marginal contributions of the profitability, investment, and momentum factors, W_S–F, A_S–F, and D_S–F, to $Sh^2(f)$ are moderate

6. Weights (in %) in $Sh^2(f)$ tangency portfolios for the models of Table 4

In table 6, we compute the weights (%) in $Sh^2(f)$ tangency portfolio.

Model	$Sh^2(f)$	Weight in tangency portfolio					
		Mkt	Size	Value	Prof	Inv	Mom
Mkt, SMB, HML, RMW_O, CMA, UMD	0.134	100.0	57.6	-49.5	-55.7	448.3	120.1
Mkt, S-F, H-F, R_C-F,C-F,U-F	0.164	-902.8	-695.9	130.9	-70.9	1000.0	638.8
Mkt, S-F, H_S-F, R_{CS}-F, C_S-F, U_S-F	0.161	-34.2	-1000.0	349.2	-61.4	491.8	354.5
Mkt, S-F, L-F, W_C-F, A-F, D-F	0.145	217.4	429.4	470.4	-13.5	-878.0	-125.7
Mkt, S-F, L_S-F, W_{CS}-F, A_S-F, D_S-F	0.125	-41.2	1000.0	317.3	-51.8	-944.4	-179.8
Mkt, SMB, HML, RMW_C, CMA, UMD	0.134	100.0	48.7	-51.9	-68.7	426.1	90.8
Mkt, SMB, HML_S, RMW_{CS}, CMA_S, UMD_S	0.120	100.0	278.5	-657.6	-132.1	5204.8	1155.5

The weights are percentages of $1 invested in the tangency portfolio. The weights are scaled to make the sum of long and short positions 100%, which implies no net investment in the risk-free asset as it is the tangency portfolio and not a point along the line.

For the spread factor models, this means the weight for the market portfolio is 100% since net investment in each (long - short) spread factor is zero.

For the models that uses excess return factors, the sum of the weights for all factors is 100%.

Conclusion

As we discussed earlier, our goal was to minimize the max squared Sharpe ratio for the intercepts for all assets. So, we should rank models based on the max squared Sharpe ratio for model factors, i.e., $Sh^2(f)$.

Thus, among the six-factor models we consider, the winner is the model that comes Mkt, SMB and the small-stock spread factors HMLS, RMWCS, CMAS, and UMDS.

- However, Fama and French argue that this is not enough evidence to switch to small-stock spreads.
- When the momentum factor is dropped, the base five-factor model performs as good as any other model from the set of models under consideration.

Also, when judging purely based on $Sh^2(f)$, models with cash-profitability factors are much better than their counterparts with operating profitability factors.

- However, when looking at other metrics such spanning-regression intercepts and GRS statistics, it is not so clear.

Fama and French caution that discipline is important in choosing factors when ranking models

- Factors should align with economic theory and must be robust on out-of-sample data
- Otherwise, it becomes a data-mining exercise where long lists of factors that come close to spanning the ex-post mean-variance-efficient (MVE) tangency portfolio of a particular time period.

Another major concern when ranking factor models is the Multiple Comparisons Problem

- For example, if we consider n factors, we will have 2^n-1 possible factor models. In our case with 6 factors, we have a possible 63 models to compare, which would make statistical inference impossible.

Thus, we must have discipline in limiting both the number of models and the number of factors in a given model.

- This would give us a framework of statistically ranking models such as done in this project.

Thank You