MIF
Advanced Asset Management

Factor Models

Janek Masojada-Edwards (15909840), Melhem Charouk El Maalou (15918122) and Konstantin Tsoi (15850803)

Abstract

In this report we opted to investigate the performance of the investment factor - CMA (Conservative Minus Aggressive) introduced by Fama and French (2015) as a part of the five-factor asset pricing model. The CMA factor was proposed by Fama and French to capture the difference between companies which invest aggressively and conservatively. The core idea is that conservative firms earn higher average returns due to lower risk or mispricing. In this report the CMA factor is evaluated, and the Fama-French three-factor model with the additional CMA factor is compared to benchmark models, including CAPM, the Fama-French three-factor model, and Fama-French three-factor with added momentum factor. We use both time-series and Fama-MacBeth (1973) cross-sectional regressions on 25 value-weighted portfolios sorted by size and investment.

Supervisor: Esther Eiling

Date: 22nd of April 2025
Amsterdam Business School (UvA)

Introduction

The three-factor asset pricing model was introduced by Eugene F. Fama and Kenneth R. French more than 30 years ago – the model was an extension to foundational Capital Asset Pricing Model (CAPM). Since then, many researchers including Fama and French proposed their own twist on factor models in attempt to capture the stock returns more efficiently. One such extension was a five-factor model developed by famous duo (Fama and French, 2015), the model added two new factors: profitability (RMW) and investment (CMA), to capture additional variation in expected returns.

In this report we focus on the **investment factor CMA (Conservative Minus Aggressive)**, which captures the trend of firms with more **conservative investment policies** to outperform those with **aggressive investment behaviour**. This factor predicts that high investment firms are more likely to be overvalued due to capital misallocation or over-optimism, leading to lower average future returns. In contrast, conservative firms, which invest less relative to their balance sheet, may illustrate better capital allocation and/or a lack of overly optimistic valuations, resulting in higher average returns.

A key question is why would CMA factor have a positive risk-adjusted return? The economic intuition is that differences in investment behaviour are predictors for differences in expected profitability and risk, while these are not captured by traditional factors like market beta, size (SMB), or value (HML). Thus, the CMA factor is expected to earn a **positive risk-adjusted return** and increase the explanatory power of asset pricing models.

The CMA factor is constructed as a long-short portfolio: long in firms with low asset growth (conservative investors) and short in those with high asset growth (aggressive investors), following the methodology of Fama and French (2015).

Methodology

We adopted a standard empirical approach used in the asset pricing literature: **Fama-MacBeth (1973) cross-sectional regressions**, as well as **time-series regressions**.

Time-series regressions

Factor Return Regression

Using time series regressions, the alpha of the CMA factor was estimated with respect to the three benchmark models. Using the monthly returns on the factors, the following regressions were run.

1. **CAPM**

$$CMA_t = a_i + \beta_{iMKT} * ER\_MKT_t + e_{it}$$

2. **Fama-French 3-factor model**

$$CMA_t = a_i + \beta_{iMKT} * ER\_MKT_t + \beta_{iSMB} * SMB_t + \beta_{iHML} * HML_t + e_{it}$$

3. **Fama-French 3-factor model with momentum**

$$CMA_t = a_i + \beta_{iMKT} * ER\_MKT_t + \beta_{iSMB} * SMB_t + \beta_{iHML} * HML_t + \beta_{iMOM} * MOM_t + e_{it}$$

Where:

- CMA_t: Investment factor (Conservative Minus Aggressive) return in month t
- ER_mkt_t: Excess return market (Market Risk Premium: Market Return – Risk Free Rate) in month t
- SMB_t: Size factor (Small Stocks Return Minus Big Stocks Return) return in month t
- HML_t: Value factor (High Stocks Minus Low B/M (book-to-market equity ratio) Stocks) return in month t
- MOM_t: Momentum factor (Carhart, 1997) return in month t
- a_i: Alpha/Intercept
- B_i: Various Betas attached to factors
- e_it: Error term

Portfolio Return Regression

The excess returns of the 25 test portfolios (sorted by size and investment) were then used in a regression with respect to the base model. The base model is a Fama-French three-factor model with the additional CMA factor. This is how portfolio **betas**, **alphas**, and **t-statistics** were estimated, helping to evaluate whether the CMA factor contributes to explaining return variation beyond traditional factors.

Fama-French 3-factor model with CMA (Base Model)

$$R_{it} - R_{Ft} = a_i + \beta_{iMKT} * ER\_MKT_t + \beta_{iSMB} * SMB_t + \beta_{iHML} * HML_t + \beta_{iCMA} * CMA_t + e_{it}$$

Where:

- R_it: realized return on portfolio i in month t
- R_Ft: risk-free rate in month t

Fama-MacBeth cross-sectional regressions

In the second stage, we test whether the estimated betas from the time-series regressions explain cross-sectional differences in average returns. Each month, we regress the cross-section of portfolio excess returns on their estimated betas:

$$R_{it} - R_{Ft} = \lambda_o + \lambda_{MKT} * \beta_{MKT} + \lambda_{SMB} * \beta_{SMB} + \lambda_{HML} * \beta_{HML} + \lambda_{CMA} * \beta_{CMA} + v_t$$

Where:

- v_t: idiosyncratic pricing error for portfolio i at time t, if the model is good should have mean zero

The average risk premia λ across time and their **statistical significance** are used to determine whether each factor is priced.

Models Tested

We estimate and compare the following models:

1. **CAPM**

$$R_{it} - R_{Ft} = a_i + \beta_{iMKT} * ER\_MKT_t + e_{it}$$

2. **Fama-French 3-factor model**

$$R_{it} - R_{Ft} = a_i + \beta_{iMKT} * ER\_MKT_t + \beta_{iSMB} * SMB_t + \beta_{iHML} * HML_t + e_{it}$$

3. **Fama-French 3-factor model with momentum**

$$R_{it} - R_{Ft} = a_i + \beta_{iMKT} * ER\_MKT_t + \beta_{iSMB} * SMB_t + \beta_{iHML} * HML_t + \beta_{iMOM} * MOM_t + e_{it}$$

4. **Fama-French 3-factor model with CMA (Base Model)**

$$R_{it} - R_{Ft} = a_i + \beta_{iMKT} * ER\_MKT_t + \beta_{iSMB} * SMB_t + \beta_{iHML} * HML_t + \beta_{iCMA} * CMA_t + e_{it}$$

By comparing the performance and explanatory power of these models, we assess whether the investment factor (CMA) improves pricing accuracy and reduces pricing errors.

Data and Summary Statistics

In the analysis in this report, we use the monthly returns for the 25 size-investment sorted portfolios (ME_INV) from July 1963 to December 2024. The data includes market excess return, SMB, HML, MOM and CMA, and the following tables summarise the key characteristics of these factors:

Factor Summary Statistics

	ER_mkt	SMB	HML	MOM	CMA
Count	738	738	738	738	738
Mean	0.0058	0.0020	0.0028	0.0061	0.0026
Std	0.0448	0.0305	0.0299	0.0419	0.0207
Min	-0.2324	-0.1532	-0.1388	-0.3430	-0.0720
25%	-0.0197	-0.0157	-0.0142	-0.0095	-0.0102
50%	0.0097	0.0007	0.0020	0.0072	0.0009
75%	0.0344	0.0200	0.0173	0.0289	0.0147
Max	0.1610	0.1828	0.1280	0.1820	0.0907

Correlation Matrix

	ER_mkt	SMB	HML	MOM	CMA
ER_mkt	1.0000	0.2815	-0.2030	-0.1692	-0.3617
SMB	0.2815	1.0000	0.0048	-0.0787	-0.0864
HML	-0.2030	0.0048	1.0000	-0.1938	**0.6847**
MOM	-0.1692	-0.0787	-0.1938	1.0000	-0.0079
CMA	-0.3617	-0.0864	**0.6847**	-0.0079	1.0000

Average Excess Returns (5x5 Grid for 25 Portfolios)

	INV1	INV2	INV3	INV4	INV5
ME1	0.0098	0.0096	0.0097	0.0086	0.0037
ME2	0.0088	0.0088	0.0090	0.0091	0.0050
ME3	0.0091	0.0089	0.0078	0.0078	0.0056
ME4	0.0079	0.0076	0.0075	0.0078	0.0062
ME5	0.0074	0.0061	0.0056	0.0058	0.0058

Discussion

The correlation matrix shows somewhat weaker correlations between all factors except for the 0.6847 correlation coefficient between HML and CMA, meaning that returns on high value companies may be correlated with returns from low investment companies. In addition, from the average excess returns 5x5 grid, we can observe the changes in excess returns as investment increases from INV1 to INV5 and Market size from ME1 to ME5. This grid shows a trend that as investment increases, excess returns slightly reduce until a big drop in the fifth investment bracket; in line with the understanding that conservative investment companies generate higher returns than aggressive investment companies. When looking at size, the trend also aligns with the theory that small stocks returns are higher than larger ones, although in the high investment category (INV5) this trend is not apparent.

Results

Factor Return Regression

The regression results for the **CMA (Investment)** factor are as follows:

CAPM: CMA ~ ER_MKT

alpha = 0.0035, t-stat = 4.9365, p-value = 0.0000

Despite the alpha having a very low value, the t-stat being higher than 3 and p-value lower than 5% shows strong statistical significance, which suggests that the CMA has an effect on excess returns after accounting for market risk.

FF3 Regression: CMA ~ ER_MKT + SMB + HML

alpha = 0.0020, t-stat = 3.6855, p-value = 0.0002

Like the CAPM regression, the CMA regressed over the Fama-French 3-factor model shows statistical significance and suggests that ER_MKT, SMB and HML do not fully explain the excess returns. This also suggests that CMA may have an additional explanatory effect above these factors.

FF3_MOM Regression: CMA ~ ER_MKT + SMB + HML + MOM

alpha = 0.0016, t-stat = 3.0162, p-value = 0.0026

The results of this regression show similar results as the above 2 regressions, also suggesting that the CMA factor does have an effect in explaining stock returns above the momentum factor in addition to the 3-factor model.

Overall, these results show that CMA has independent pricing power beyond other factors.

Portfolio Return Regression

The base model is a Fama-French three-factor model with the additional CMA factor. The results of the regression on the 25 size-investment portfolios are shown below.

Alphas

ME/INV	INV1	INV2	INV3	INV4	INV5
ME1	-0.0002	0.0013	0.0017	0.0008	-0.0041
ME2	-0.0011	0.0006	0.0012	0.0014	-0.0021
ME3	0.0001	0.0009	0.0004	0.0008	-0.0010
ME4	-0.0010	-0.0000	0.0004	0.0012	0.0001
ME5	-0.0002	-0.0003	-0.0001	0.0010	0.0019

Alpha t-stats

ME/INV	INV1	INV2	INV3	INV4	INV5
ME1	-0.2315	2.6208	3.1322	1.4734	-6.2208
ME2	-2.0669	1.1064	2.4542	2.9599	-4.2668
ME3	0.2131	1.7665	0.7962	1.4775	-1.7414
ME4	-1.6542	-0.0783	0.8142	2.1455	0.1855
ME5	-0.2413	-0.6213	-0.2563	2.1676	3.5732

R squared

ME/INV	INV1	INV2	INV3	INV4	INV5
ME1	0.8893	0.9484	0.9349	0.9423	0.9424
ME2	0.9520	0.9313	0.9411	0.9505	0.9651
ME3	0.9054	0.9220	0.9205	0.9320	0.9477
ME4	0.9066	0.9035	0.9089	0.9078	0.9210
ME5	0.8705	0.9192	0.9281	0.9311	0.9371

CMA Betas

ME/INV	INV1	INV2	INV3	INV4	INV5
ME1	0.6253	0.1523	0.0358	-0.0185	-0.2234
ME2	0.5226	0.1077	0.0989	-0.2186	-0.3550
ME3	0.3249	0.2236	0.0569	-0.1942	-0.4313
ME4	0.5078	0.1876	0.0376	-0.0952	-0.3205
ME5	0.7904	0.4600	0.0538	-0.1673	-0.6708

CMA Beta t-stats

ME/INV	INV1	INV2	INV3	INV4	INV5
ME1	9.7552	4.4675	0.9401	-0.4956	-5.0387
ME2	14.3145	2.9112	2.9125	-6.5714	-10.5547
ME3	6.9576	6.1450	1.5825	-5.2935	-11.1719
ME4	11.8466	4.7912	1.0116	-2.4228	-6.9822
ME5	18.4986	15.4389	1.8549	-5.5117	-18.6622

ER_mkt Betas

ME/INV	INV1	INV2	INV3	INV4	INV5
ME1	1.0528	0.9023	0.8834	0.9106	1.0189
ME2	1.1365	0.9643	0.9156	0.9803	1.0976
ME3	1.1004	0.9615	0.9478	0.9916	1.1006
ME4	1.1570	1.0357	0.9936	0.9990	1.1119
ME5	1.0505	0.9553	0.9550	0.9714	1.0656

ER_mkt Beta t-stats

ME/INV	INV1	INV2	INV3	INV4	INV5
ME1	46.0329	74.1932	65.1009	68.4898	64.4163
ME2	87.2442	73.0626	75.6017	82.5768	91.4610
ME3	66.0407	74.0592	73.9040	75.7642	79.8971
ME4	75.6587	74.1298	74.8905	71.2710	67.9019
ME5	68.9058	89.8635	92.2368	89.7153	83.0859

SMB Betas

ME/INV	INV1	INV2	INV3	INV4	INV5
ME1	1.4336	1.0356	1.0129	1.0639	1.2741
ME2	0.9733	0.7347	0.8197	0.8133	1.0126
ME3	0.6586	0.5358	0.4712	0.5825	0.7385
ME4	0.2697	0.1776	0.2045	0.2802	0.5206
ME5	-0.1274	-0.2077	-0.2014	-0.2290	-0.1805

SMB Beta t-stats

ME/INV	INV1	INV2	INV3	INV4	INV5
ME1	45.5252	61.8460	54.2082	58.1144	58.5033
ME2	54.2607	40.4291	49.1562	49.7551	61.2820
ME3	28.7057	29.9726	26.6838	32.3271	38.9347
ME4	12.8105	9.2300	11.1931	14.5199	23.0911
ME5	-6.0716	-14.1934	-14.1252	-15.3620	-10.2215

HML Betas

ME/INV	INV1	INV2	INV3	INV4	INV5
ME1	-0.2171	0.1985	0.2395	0.1708	-0.0577
ME2	0.0138	0.3003	0.2199	0.3149	-0.1342
ME3	0.1512	0.2488	0.2660	0.2201	-0.0723
ME4	0.1317	0.2876	0.2771	0.1487	-0.2059
ME5	-0.1211	0.0249	0.1436	0.0201	-0.0659

HML Beta t-stats

ME/INV	INV1	INV2	INV3	INV4	INV5
ME1	-5.1306	8.8213	9.5418	6.9433	-1.9722
ME2	0.5713	12.3023	9.8155	14.3383	-6.0461
ME3	4.9046	10.3599	11.2117	9.0904	-2.8383
ME4	4.6563	11.1278	11.2919	5.7340	-6.7986
ME5	-4.2935	1.2660	7.4998	1.0039	-2.7786

Discussion

Alphas and Alpha t-stats

The alphas are generally small, indicating that the model captures most of the systematic return variation. The t-stats for the alphas vary, with some showing high significance and others low significance. However, in some corners of the grid, like the small-conservative and big-aggressive portfolios, the alphas are higher and are statistically significant. For instance, ME1 INV5 has a strongly negative alpha (-0.0041) with a t-stat of -6.22, suggesting the model underpredicts returns for highly aggressive small firms.

R-squared Values

The high R-squared values show the base model explains the majority of variation in portfolio returns. This shows the high explanatory power of the model.

Factor Betas

Market Betas (ER_mkt)

These are close to 1 across the board and highly significant, reflecting broad market exposure in all portfolios.

SMB Betas

Show clear size pattern with small firms having high SMB betas while large firms (ME5) have negative betas, which is consistent with what we expect. All betas are strongly significant.

HML Betas

Show expected patterns, with more conservative firms often exhibiting higher value exposure. Almost all betas are highly significant

CMA Betas

Shows a decreasing trend across the investment scale. Conservative portfolios (INV1) have strongly positive CMA betas, while aggressive ones (INV5) are negative. These betas are also highly significant in most cases. This confirms our interpretation of the CMA factor.

Fama-MacBeth Regression Results

CAPM Model

Risk Premiums (λ):	
ER_mkt	-0.0024 (t-stat: -1.0090)
R-squared	0.0425

The market return factor shows a small negative lambda/risk premium with a t-stat between -1.96 and +1.96, suggesting that is insignificant and the market risk premiums does not fully explain the returns in the model. This is further supported by the R-squared being too low at 4.25%.

FF3 Model

Risk Premiums (λ):	
ER_mkt	-0.0004 (t-stat: -0.1151)
SMB	0.0013 (t-stat: 2.7605)
HML	0.0049 (t-stat: 3.8157)
R-squared	0.6173

The Lambdas for SMB and HML in this regression are significant and the R-squared is high, suggesting that size and value factors do explain variation in returns.

FF3_MOM Model

Risk Premiums (λ):	
ER_mkt	0.0060 (t-stat: 1.1796)
SMB	0.0013 (t-stat: 2.6486)
HML	0.0058 (t-stat: 3.9915)
MOM	0.0204 (t-stat: 1.7571)
R-squared	0.6953

Adding to the above 3 factors the momentum factor does increase the R-squared by 8%, which in turn increases the explanatory power of this factor when combined with the 3-factor moment, despite it being insignificant on its own based on the t-stat.

Base Model

Risk Premiums (λ):	
ER_mkt	-0.0005 (t-stat: -0.1268)
SMB	0.0013 (t-stat: 2.6928)
HML	0.0048 (t-stat: 2.8063)
CMA	0.0023 (t-stat: 3.3926)
R-squared	0.6176

After including the CMA factor with the 3-factor model, the R-square does not change much from FF3 but the CMA is a strong explanatory factor due to its high t-stat of the lambda, suggesting that conservative minus aggressive investment spending factor does explain higher returns when controlling for market, size, and value factors.

Robustness Test

One test for robustness is to split the data into time periods and test if the performance is consistent over time. To perform this the base model time series regression was reperformed on data split into two periods. The results from this are shown below.

Early Period - Alphas

	INV1	INV2	INV3	INV4	INV5
ME1	-0.0013	0.0006	0.0003	0.0005	-0.0047
ME2	-0.0005	0.0014	0.0012	0.0012	-0.0017
ME3	0.0003	0.0019	0.0006	0.0014	-0.0006
ME4	-0.001	-0.0002	0.0002	0.0014	0.0009
ME5	0.0	-0.0	-0.0001	0.0008	0.0017

Early Period - Alpha t-stats

	INV1	INV2	INV3	INV4	INV5
ME1	-1.7006	0.9766	0.4008	0.8254	-6.1263
ME2	-0.7689	1.8886	1.8225	1.9671	-2.6304
ME3	0.3333	2.8016	0.9392	2.0132	-0.7973
ME4	-1.1772	-0.2801	0.3071	1.8955	1.0745
ME5	0.0131	-0.0723	-0.2385	1.442	2.2021

Early Period - R-squareds

	INV1	INV2	INV3	INV4	INV5
ME1	0.9557	0.9579	0.9606	0.9638	0.9634
ME2	0.9591	0.9403	0.9442	0.9614	0.9728
ME3	0.9114	0.9309	0.944	0.9438	0.9595
ME4	0.9141	0.9201	0.9294	0.9327	0.9419
ME5	0.866	0.9308	0.9428	0.944	0.9362

Early Period - CMA Betas

	INV1	INV2	INV3	INV4	INV5
ME1	0.4169	0.132	0.1333	0.0057	-0.1043
ME2	0.5074	0.1443	0.0311	-0.1025	-0.2709
ME3	0.4316	0.2561	-0.0061	-0.3331	-0.3971
ME4	0.6625	0.2272	-0.0391	-0.1789	-0.4224
ME5	0.8939	0.5492	-0.0319	-0.3467	-0.672

Early Period - CMA Beta t-stats

	INV1	INV2	INV3	INV4	INV5
ME1	6.7794	2.566	2.6002	0.1086	-1.7227
ME2	9.4865	2.5182	0.5712	-2.0574	-5.4056
ME3	5.8495	4.6177	-0.1203	-5.9943	-6.943
ME4	9.5176	3.9043	-0.7038	-3.1256	-6.6088
ME5	11.9139	11.7328	-0.7347	-7.4672	-11.199

Late Period - Alphas

	INV1	INV2	INV3	INV4	INV5
ME1	0.0003	0.0023	0.0036	0.0013	-0.004
ME2	-0.0017	0.0002	0.0017	0.0023	-0.0024
ME3	-0.0002	0.0002	0.0007	0.0006	-0.0013
ME4	-0.0008	0.0004	0.001	0.0019	-0.0006
ME5	-0.0003	-0.0	0.0005	0.0016	0.0019

Late Period - Alpha t-stats

	INV1	INV2	INV3	INV4	INV5
ME1	0.1554	2.9708	4.1595	1.6493	-3.8056
ME2	-2.0915	0.2651	2.3961	3.1932	-3.2581
ME3	-0.1522	0.2435	0.8078	0.7902	-1.5248
ME4	-0.8594	0.4196	1.1584	2.2071	-0.6074
ME5	-0.3333	-0.0792	0.734	2.4506	2.5298

Late Period - R-squareds

	INV1	INV2	INV3	INV4	INV5
ME1	0.8564	0.9407	0.9162	0.9288	0.9238
ME2	0.9506	0.9265	0.9424	0.9466	0.9599
ME3	0.9019	0.916	0.9031	0.9255	0.9371
ME4	0.9036	0.8905	0.8954	0.8943	0.9101
ME5	0.8829	0.9164	0.9218	0.9273	0.9385

Late Period - CMA Betas

	INV1	INV2	INV3	INV4	INV5
ME1	0.723	0.1702	-0.0195	-0.0093	-0.2915
ME2	0.538	0.1116	0.1297	-0.2813	-0.4292
ME3	0.2745	0.213	0.1017	-0.129	-0.4662
ME4	0.4331	0.1712	0.0924	-0.0633	-0.2944
ME5	0.7473	0.3925	0.0963	-0.0528	-0.664

Late Period - CMA Beta t-stats

	INV1	INV2	INV3	INV4	INV5
ME1	7.1666	3.6454	-0.3666	-0.1887	-4.5835
ME2	10.859	2.2717	2.9841	-6.4609	-9.3997
ME3	4.4197	4.2938	2.0181	-2.6209	-8.6774
ME4	7.8865	3.1447	1.8418	-1.2029	-4.5862
ME5	14.8674	10.2873	2.5051	-1.3499	-14.5324

It can be seen from the data that the investment factor remained consistent across both periods, maintaining pricing power and a similar pattern of returns. There is some evidence of slight weakening of the effect in the later period in the form of slighly lower R-squared values, more variation in the middle investment grouping, and generally smaller alphas. However, the factor remains significant in both periods, with the same strong positive betas for low invesment firms and negative betas for high investment firms, with continual statistical significance throughout both periods. This suggests that the investment factor is stable and has persisted through different market conditions.

Conclusion

The analysis of the regression results in this report leads us to conclude that CMA is a strong explanatory factor in asset pricing, and performs well when tested against models like the CAPM and Fama French 3 factor model, and in the context of French-Macbeth cross sectional regression where it provides statistically significant alphas and risk premiums. This is in line with the expectation that companies with conservative investment spending generate higher expected return due to the proper allocation of funds into profitable investment opportunities as opposed to companies with aggressive spending behaviour. Given its significance in both time-series and cross-sectional tests, and its consistent performance in split period robustness tests, CMA should be included in smart-beta strategies focused on long-term asset selection.

References

1. Fama, Eugene F., and Kenneth R. French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.
2. Sharpe, William F., 1964. *Capital asset prices: A theory of market equilibrium under conditions of risk*, *Journal of Finance* 19, 425–442.
3. Fama, Eugene F., and Kenneth R. French, 1993. *Common risk factors in the returns on stocks and bonds*, *Journal of Financial Economics* 33, 3–56.
4. Carhart, Mark M., 1997. *On persistence in mutual fund performance*, *Journal of Finance* 52, 57–82.

Appendix

Code:

```
import pandas as pd
import numpy as np
import statsmodels.api as sm
import matplotlib.pyplot as plt
import seaborn as sns
from pathlib import Path

# DATA HANDLING
DATA_DIR = Path("//Users//janekedwards//Desktop//Advanced Asset Management")
START_DATE = "1963-07"
END_DATE = "2024-12"

portfolios = pd.read_excel(DATA_DIR / "25_Portfolios_ME_INV_5x5_Excel.xlsx", sheet_name="Sheet1", index_col=0)
portfolios.index = pd.to_datetime(portfolios.index, format="%Y%m").to_period("M")
portfolios = portfolios.loc[START_DATE:END_DATE]
portfolios = portfolios / 100

factor_df = pd.read_excel(DATA_DIR / "Data_assignmentAAM2025.xlsx", sheet_name="Sheet1", index_col=0)
factor_df.index = pd.to_datetime(factor_df.index, format="%Y%m").to_period("M")
factor_df = factor_df.loc[START_DATE:END_DATE]

# CALCULATE EXCESS RETURNS
excess_returns = portfolios.subtract(factor_df["RF"], axis=0)

def print_summary_stats():
    print("\nFactor Summary Statistics:")
    print(factor_df[["ER_mkt", "SMB", "HML", "MOM", "CMA"]].describe().round(4))

    print("\nCorrelation Matrix:")
    print(factor_df[["ER_mkt", "SMB", "HML", "MOM", "CMA"]].corr().round(4))
```

```
    print("\nAverage Excess Returns (5x5):")
    reshaped = excess_returns.mean().values.reshape(5, 5)
    avg_returns_df = pd.DataFrame(reshaped, index=["ME1", "ME2", "ME3", "ME4", "ME5"],
columns=["INV1", "INV2", "INV3", "INV4", "INV5"])
    print(avg_returns_df.round(4))

    #other summary stats required?

def factor_alpha_test(model_type):

    if model_type == 'capm':
        factors = ['ER_mkt']
    elif model_type == 'ff3':
        factors = ['ER_mkt', 'SMB', 'HML']
    elif model_type == 'ff3_mom':
        factors = ['ER_mkt', 'SMB', 'HML', 'MOM']

    X = factor_df[factors]
    X = sm.add_constant(X)
    y = factor_df["CMA"]
    model = sm.OLS(y, X).fit()
    alpha = model.params["const"]
    t_stat = model.tvalues["const"]
    p_value = model.pvalues["const"]

    return alpha, t_stat, p_value

def regress_portfolio():

    alphas = []
    alphas_t_stats = []
    betas = []
    betas_t_stats = []
    r_squareds = []
```

```python
    factors = ['ER_mkt', 'SMB', 'HML', 'CMA']

    for col in excess_returns.columns:
        X = factor_df[factors]
        X = sm.add_constant(X)
        y = excess_returns[col]
        model = sm.OLS(y, X).fit()
        alphas.append(model.params["const"])
        alphas_t_stats.append(model.tvalues["const"])
        betas.append(model.params[1:])
        betas_t_stats.append(model.tvalues[1:])
        r_squareds.append(model.rsquared)



    return alphas, alphas_t_stats, betas, betas_t_stats, r_squareds

def fama_macbeth_regression(model_type):
    #TIME SERIES FOR BETAS
    if model_type == 'capm':
        factors = ['ER_mkt']
    elif model_type == 'ff3':
        factors = ['ER_mkt', 'SMB', 'HML']
    elif model_type == 'ff3_mom':
        factors = ['ER_mkt', 'SMB', 'HML', 'MOM']
    elif model_type == 'base':
        factors = ['ER_mkt', 'SMB', 'HML', 'CMA']



    betas = []
    avg_returns = []

    for col in excess_returns.columns:
        X = factor_df[factors]
        X = sm.add_constant(X)
```

```
        y = excess_returns[col]
        model = sm.OLS(y, X).fit()
        betas.append(model.params[1:])
        avg_returns.append(excess_returns[col].mean())

    betas = np.array(betas)
    avg_returns = np.array(avg_returns)

    #CROSS SECTIONAL
    X = sm.add_constant(betas)
    model = sm.OLS(avg_returns, X).fit()

    #SHANKEN ADJUSTMENT
    n = len(excess_returns)
    k = len(factors)
    sigma_f = factor_df[factors].cov()
    sigma_e = np.diag([sm.OLS(excess_returns[col],
sm.add_constant(factor_df[factors])).fit().resid.var()
                    for col in excess_returns.columns])



    lambda_hat = model.params[1:]
    sigma_lambda = model.cov_params()[1:, 1:]
    adj_factor = 1 + lambda_hat @ np.linalg.inv(sigma_f) @ lambda_hat
    sigma_lambda_adj = sigma_lambda * adj_factor

    #T-STATS
    t_stats_adj = lambda_hat / np.sqrt(np.diag(sigma_lambda_adj))

    return {
        'lambda': lambda_hat,
        't_stats': t_stats_adj,
        'r_squared': model.rsquared,
        'model': model
    }
```

```python
def subperiod_analysis():

    mid_date = "1994-01"
    results = {}

    for period in ["Early", "Late"]:
        if period == "Early":
            start, end = START_DATE, mid_date
        else:
            start, end = mid_date, END_DATE


        portfolios_sub = portfolios.loc[start:end]
        factor_df_sub = factor_df.loc[start:end]
        excess_returns_sub = portfolios_sub.subtract(factor_df_sub["RF"], axis=0)


        alphas = []
        alphas_t_stats = []
        betas = []
        betas_t_stats = []
        r_squareds = []

        factors = ['ER_mkt', 'SMB', 'HML', 'CMA']

        for col in excess_returns_sub.columns:
            X = factor_df_sub[factors]
            X = sm.add_constant(X)
            y = excess_returns_sub[col]
            model = sm.OLS(y, X).fit()
            alphas.append(model.params["const"])
            alphas_t_stats.append(model.tvalues["const"])
            betas.append(model.params[1:])
            betas_t_stats.append(model.tvalues[1:])
```

```
            r_squareds.append(model.rsquared)



        results[period] = {
            'alphas': alphas,
            'alphas_t_stats': alphas_t_stats,
            'betas': betas,
            'betas_t_stats': betas_t_stats,
            'r_squareds': r_squareds
        }

    return results

if __name__ == "__main__":
    print_summary_stats()
    alpha, t_stat, p_value = factor_alpha_test('capm')
    print(f"\nCAPM: CMA ~ ER_MKT: alpha = {alpha:.4f}, t-stat = {t_stat:.4f}, p-value = 
{p_value:.4f}")
    alpha, t_stat, p_value = factor_alpha_test('ff3')
    print(f"\nFF3: CMA ~ ER_MKT + SMB + HML: alpha = {alpha:.4f}, t-stat = {t_stat:.4f}, p-
value = {p_value:.4f}")
    alpha, t_stat, p_value = factor_alpha_test('ff3_mom')
    print(f"\nFF3_MOM: CMA ~ ER_MKT + SMB + HML + MOM: alpha = {alpha:.4f}, t-stat = 
{t_stat:.4f}, p-value = {p_value:.4f}")

    alphas, alphas_t_stats, betas, betas_t_stats, r_squareds = regress_portfolio()

    #RESHAPE DATA
    def reshape_to_grid(data, index):
        grid = np.zeros((5, 5))
        for i, idx in enumerate(index):
            me = int(idx.split()[0][2]) - 1  # Extract ME number (1-5)
            inv = int(idx.split()[1][3]) - 1  # Extract INV number (1-5)
            grid[me, inv] = data[i]
        return grid
```

```
#CREATE 5x5 GRIDS
alphas_grid = reshape_to_grid(alphas, excess_returns.columns)
alphas_t_stats_grid = reshape_to_grid(alphas_t_stats, excess_returns.columns)
r_squareds_grid = reshape_to_grid(r_squareds, excess_returns.columns)

#CREATE 5x5x4 GRIDS
betas_grid = np.zeros((5, 5, 4))
betas_t_stats_grid = np.zeros((5, 5, 4))
for i, idx in enumerate(excess_returns.columns):
    me = int(idx.split()[0][2]) - 1
    inv = int(idx.split()[1][3]) - 1
    betas_grid[me, inv] = betas[i]
    betas_t_stats_grid[me, inv] = betas_t_stats[i]


me_labels = ['ME1', 'ME2', 'ME3', 'ME4', 'ME5']
inv_labels = ['INV1', 'INV2', 'INV3', 'INV4', 'INV5']

print("\nAlphas:")
print(pd.DataFrame(alphas_grid, index=me_labels, columns=inv_labels).round(4))

print("\nAlpha t-stats:")
print(pd.DataFrame(alphas_t_stats_grid, index=me_labels, columns=inv_labels).round(4))

print("\nR-squareds:")
print(pd.DataFrame(r_squareds_grid, index=me_labels, columns=inv_labels).round(4))

print("\nBetas:")
for i, factor in enumerate(['ER_mkt', 'SMB', 'HML', 'CMA']):
    print(f"\n{factor} Betas:")
    print(pd.DataFrame(betas_grid[:,:,i], index=me_labels, columns=inv_labels).round(4))
    print(f"\n{factor} Beta t-stats:")
    print(pd.DataFrame(betas_t_stats_grid[:,:,i], index=me_labels,
columns=inv_labels).round(4))
```

```
#FAMA-MACBETH ANALYSIS
print("\nFama-MacBeth Regression Results:")
models = ['capm', 'ff3', 'ff3_mom', 'base']
for model in models:
    results = fama_macbeth_regression(model)
    print(f"\n{model.upper()} Model:")
    print(f"Risk Premiums (Lambda):")
    if model == 'ff3_mom':
        factors = ['ER_mkt', 'SMB', 'HML', 'MOM']
    else:
        factors = ['ER_mkt', 'SMB', 'HML', 'CMA']
    for i, factor in enumerate(factors[:len(results['lambda'])]):
        print(f"{factor}: {results['lambda'][i]:.4f} (t-stat: {results['t_stats'][i]:.4f})")
    print(f"R-squared: {results['r_squared']:.4f}")

# SUB-PERIOD ANALYSIS
print("\n=== Sub-period Analysis ===")
subperiod_results = subperiod_analysis()

for period, results in subperiod_results.items():
    print(f"\n{period} Period Results:")



    alphas_grid = reshape_to_grid(results['alphas'], excess_returns.columns)
    alphas_t_stats_grid = reshape_to_grid(results['alphas_t_stats'],
excess_returns.columns)
    r_squareds_grid = reshape_to_grid(results['r_squareds'], excess_returns.columns)

    betas_grid = np.zeros((5, 5, 4))
    betas_t_stats_grid = np.zeros((5, 5, 4))
    for i, idx in enumerate(excess_returns.columns):
        me = int(idx.split()[0][2]) - 1
        inv = int(idx.split()[1][3]) - 1
        betas_grid[me, inv] = results['betas'][i]
```

```
            betas_t_stats_grid[me, inv] = results['betas_t_stats'][i]

        print("\nAlphas:")
        print(pd.DataFrame(alphas_grid, index=me_labels, columns=inv_labels).round(4))

        print("\nAlpha t-stats:")
        print(pd.DataFrame(alphas_t_stats_grid, index=me_labels, columns=inv_labels).round(4))

        print("\nR-squareds:")
        print(pd.DataFrame(r_squareds_grid, index=me_labels, columns=inv_labels).round(4))

        print("\nCMA Betas:")
        print(pd.DataFrame(betas_grid[:,:,3], index=me_labels, columns=inv_labels).round(4))

        print("\nCMA Beta t-stats:")
        print(pd.DataFrame(betas_t_stats_grid[:,:,3], index=me_labels, columns=inv_labels).round(4))
```