UNIVERSITY OF BIRMINGHAM | BIRMINGHAM BUSINESS SCHOOL

Assessment and Feedback: Student Template

Student ID Number(s): 2244195

Programme: B.Sc. Mathematical Economics and Statistics

Module: Economics Extended Essay

Name of Tutor: William Pouliot

Assignment Title: Evaluation of Performance in the Cross-section of Mutual Fund Returns: Improvement on the Fama French 3-factor CAPM

Date and Time of Submission: 16th May 12pm

Actual Word Count: 4991

Extension: N **Extension Due Date:**

I do not wish my assignment to be considered for including as an exemplar in the **School Bank of Assessed Work. ***
delete as appropriate

The purpose of this template is to ensure you receive targeted feedback that will support your learning. It is a requirement to complete to complete all 3 sections, and to include this completed template as the first page of every assignment that is submitted for marking (your School will advise on exceptions).
Section One: Reflecting on the feedback that I have received on previous assessments, the following issues/topics have been identified as areas for improvement: (add 3 bullet points). *NB – for first year students/PGTs in the first term, this refers to assessments in your previous institution* • logic • language • application of knowledge
Section Two: In this assignment, I have attempted to act on previous feedback in the following ways (3 bullet points) • logic • language

• application of knowledge
Section Three: Feedback on the following aspects of this assignment (i.e. content/style/approach) would be particularly helpful to me: (3 bullet points) • logic • language • application of knowledge

Please ensure that you complete and attach this template to the front of all work that is submitted.

By submitting your work online you are confirming that your work is your own and that you understand and have read the University's rules regarding authorship and plagiarism and the consequences that will arise should you submit work not complying with University's Code of Practice on Academic Integrity.

I confirm that I have / have not used a proof-reader(s) (delete as appropriate). If I have used a proof-reader(s) I confirm that I understand the guidance on use of a proof-reader, as specified in the Code of Practice and School guidance.

Evaluation of Performance in the Cross-section of Mutual Fund Returns: Improvement on the Fama French 3-factor CAPM

An Essay Submitted
for the Degree of

BACHELOR OF SCIENCE
in
University of Birmingham

Student ID: 2244195
Word count: 4991

to
BUSINESS SCHOOL
UNIVERSITY OF BIRMINGHAM
May 2022

ABSTRACT

Fama and French (2010) use an unconditional model to test the performance of actively managed mutual funds. However, there is still part of fund returns that can't be explained by their 3-factor CAPM and the market beta in their model may be time-varying. In this paper, we use the 5-factor unconditional CAPM, 3-factor conditional CAPM and the 5-factor conditional CAPM to find the performance of actively managed mutual funds compared with that found by 3-factor unconditional CAPM. We find that the 5-factor unconditional CAPM makes the overall performance of fund managers worse, but the performance of top managers better. The conditional model improves the performance of almost every fund. The 5-factor conditional CAPM exerts a combined effect of the previous two.

Keywords: actively managed mutual fund; US; Fama French; fund performance; conditional CAPM

Contents

1 Introduction: Studies in actively managed mutual funds and why do we care about them? 1

1.1 Overview of Section 1 1

1.2 How are actively managed mutual funds studies going? 1

1.3 The Fama French model - a double-edged sword 1

1.4 What shall we do in this research? 2

2 Empirical framework: What data and model do we use? 2

2.1 Overview of the Section 2 2

2.2 Data - our building bricks 2

2.2.1 The fund returns 2

2.2.2 The Fama French factors 3

2.2.3 The information variables 4

2.3 Regression framework: what models do we use? 4

2.3.1 Fama French 3-factor CAPM (1993) 4

2.3.2 Fama French 5-factor CAPM (2015) 5

2.3.3 Ferson and Schadt (1996) 5

3 Results: Any difference among different models? 7

3.1 Overview of Section 4 7

3.2 Summary of results: What can we discover? 9

4 Discussion: Why is the result like that and are there any limitations? 11

4.1 Overview of Section 4 11

4.2 Why is the result like that? 12

4.3 Limitations: Our study is not enough. 12

5 Conclusions: What does the study reveal? 13

6 References 13

7 Appendix 1: distributions of t(α) estimates 15

8 Appendix 2: Filtering passively managed funds 17

9 Appendix 3: Matlab code to conduct the study 18

List of Figures

1 Comparison of conditional and unconditional model: 3-factor gross return t-value distribution . . . 15

2 Comparison of conditional and unconditional model: 5-factor gross return t-value distribution . . . 16

3 Comparison of conditional and unconditional model: 3-factor net return t-value distribution . . . 16

4 Comparison of conditional and unconditional model: 5-factor net return t-value distribution . . . 17

List of Tables

1 descriptive statistics of t(α) estimates . . . 7

2 percentiles of t(α) estimates . . . 8

3 Keyword requirements for different sorts of passively managed funds . . . 18

1 Introduction: Studies in actively managed mutual funds and why do we care about them?

1.1 Overview of Section 1

This section provides an overview of the studies in performance of mutual funds and the drawbacks of Fama-French models, as well as what we do in this research.

1.2 How are actively managed mutual funds studies going?

There have been various studies in testing the performance of actively managed mutual funds. Jensen (1969) shows that mutual funds on average cannot outperform buy-and-hold strategies that use stock selection. Henriksson (1984) suggests that mutual fund managers also cannot take advantage of market timing to win market portfolios. Although it is well established that actively managed mutual funds have on average lower returns than passively managed mutual funds (Sharpe, 1991; Fama and French, 2010), this does not stop investors from trying to make large gains from actively managed mutual funds. One reason may be due to the persistence of mutual fund performance both in the short run (Hendricks, Patel, and Zeckhauser, 1993; Brown and Goetzmann, 1995) and in the long run (Grinblatt and Titman, 1992), despite the presence of survivorship bias (Malkiel, 1995). In fact, actively managed mutual funds are becoming increasingly popular in the U.S. economy as investors may be more rational than we assume (Gruber, 2011). Because of this, the assessment of actively managed mutual funds is still valuable today.

1.3 The Fama French model - a double-edged sword

Fama and French use 3-Factor and 4-Factor models to test the performance of actively managed mutual funds and find that few funds have sufficient skill to cover their costs (Fama and French, 2010). The models they use are improved versions of the Capital Asset Pricing Model (CAPM), which was developed by Sharpe (1964), Lintner (1965), and Mossin (1966).The CAPM is the first model to show theoretically that expected returns are systematically affected by market risk, but Fama and French (1992) present evidence that static CAPM cannot explain the cross-section of asset returns. In contrast, the Fama-French 3-factor model can effectively explain the CAPM anomaly (Fama and French, 1993; Fama and French, 1996), making it the most popular model in empirical asset pricing.

However, the Fama-French 3-Factor and 4-Factor models fail to capture the time variation in factor exposure, and their 2010 paper has been criticized for using unconditional versions of CAPM. Jagannathan and Wang (1996) argue that the relative firm risk measured by CAPM beta varies over the business cycle. For example, during a recession, the financial leverage of a troubled firm may rise, leading to an increase in its beta. Brooks et al. (1992) state that the riskiness of a firm is related to the maturity and growth of the firm. Therefore, the beta depends on the available information and changes over time. There is also support for the use of conditional models to make the regression intercept centripetal to zero (Ferson and Harvey, 1991; Ferson and Schadt, 1996), implying that conditional models do a better job than unconditional models in explaining stock returns.

With such ideas and evidence, a number of conditional models were designed. Ferson and Schadt (1996) use information variable to account for the variations in market beta and find that fund managers do better than in traditional unconditional models. Iqbal, Brooks, Galagedera (2010) reveal that the Fama-French model involving "trading volume" as a scaling variable outperforms the original Fama-French model in explaining asset pricing.

1.4 What shall we do in this research?

Our task is to firstly add RMW and CMA, two factors in Fama French 5-factor CAPM (Fama and French, 2015), to the 3-factor unconditional CAPM used in the Fama French 2010 paper to capture a larger part of fund returns. Then we use conditional information to improve the unconditional models, which essentially follows the procedure of Ferson and Schadt (1996) to allow for variations in fund beta. We also combine RMW, CMA and conditional information together to make further improvement. After that we evaluate the t-values of α estimates (we call $t(\alpha)$ estimates for short) of actively managed mutual funds under each model and assess their performance compared to the original 3-factor CAPM.

2 Empirical framework: What data and model do we use?

2.1 Overview of the Section 2

This section describes the data we use, the construction of variables and the models used to evaluate actively managed mutual funds performance.

2.2 Data - our building bricks

2.2.1 The fund returns

Our fund data (including return index, total assets and expense ratios) come from our supervisor, ranging from December 1983 through September 2006 (so that calculated monthly returns start from January 1984). They cover 7,408 live funds and 1,616 dead funds in different share classes, for a total of 9,024.

To get fund returns, we first calculate monthly gross returns of each fund's component share class using the return index. The net return is the gross return minus 1/12 of the fund's expense ratio for the year. If a fund's expense ratio is missing, we assume it to be the average value of the fund's existing expense ratio from January 1984 to September 2006. If a fund's net asset value is missing, we dismiss the fund at month t to avoid imprecise information. We then exclude passively managed funds by keyword filtering, the keywords and detailed process of which can be found in Appendix 2. We then combine the value-weighted returns of the fund's share classes based on total net asset (TNA) to calculate aggregated funds' monthly gross return and net return. Besides, we only incorporate a fund once it reaches 5 million 2006 dollars in TNA, and 3,006 funds reach that. We also set 250 million and 1 billion TNA threshold, and 1,260 and 498 funds reach them respectively. This mitigates the incubation bias as newly born funds are typically low in TNA. After these preprocessing procedures, we obtain the funds' gross and net return data.

Our monthly fund return data covers funds that invested primarily in U.S. common stocks between January 1984 and September 2006, and they involve live and dead funds, so the data are free of survivorship bias.

2.2.2 The Fama French factors

The underlying monthly Fama French factors (market return minus risk-free rate (Mkt - Rf), small minus big (SMB), high minus low (HML), robust minus weak (RMW), conservative minus aggressive (CMA)) and risk-free rate (Rf) data are also from our supervisor and cover the period from January 1984 to September 2006.

Their construction follows Fama and French (1993). NYSE, AMEX and NASDAQ stocks are classified as Small and Big, where Small includes those stocks with market capitalization below the NYSE median and Big includes those stocks with market capitalization above the NYSE median. All stocks are also divided into three groups based on book-to-market (B/M). Those stocks in the bottom 30%, middle 40% and top 30% of NYSE B/M are classified as Growth, Neutral and Value, respectively. These two classifications are then combined to get six value-weighted portfolios (S/V, S/M, S/G, B/V, B/M, B/G). Then SMB_t is defined as the average return of the three small-cap portfolios (S/V, S/M, and S/G) in month t minus the average return of the three large-cap portfolios (B/V, B/M, and B/G). HML_t is the average return of the two value portfolios (S/V and B/V) in month t minus the average return of the two growth portfolios (S/G and B/G).

The construction of RMW and CMA are similar. For RMW, all stocks are grouped into three categories based on operating profitability. Those stocks that are in the lowest 30%, middle 40% and highest 30% of robustness of stock profits are classified in the Weak (W), Neutral (N) and Robust (R) groups. The size and operating profitability classifications are then combined to get six value-weighted portfolios (S/W, S/N, S/R, B/W, B/N, B/R). Then RMW_t is the average return of the two Robust portfolios (S/R and B/R) minus the average return of the two Weak portfolios (S/W and B/W) in month t.

Likewise, for CMA, the stocks are divided into three groups based on their firms' investment. The stocks that are in the lowest 30%, middle 40% and highest 30% of investment are classified in the Conservative (C), Neutral (N), or Aggressive (A) groups. The size and investment classifications are then combined to get six value-weighted portfolios (S/C, S/N, S/A, B/C, B/N, B/A). And CMA_t is the average return of the two Conservative portfolios (S/C and B/C) minus the average return of the two Aggressive portfolios (S/A and B/A) in month t.

2.2.3 The information variables

There are also monthly information variables which we use in this study. The variables are:

1. lagged one-month Treasury bill yield
2. lagged term spread
3. lagged quality spread of corporate bond
4. dummy variable for month of January

Term spread is defined as the difference between the return on the 10-year and the 3-month government bond interest rate. Quality spread of corporate bond is the yield spread between Moody's BAA and AAA corporate bonds. The data are all monthly, starting from December 1983 to August 2006 so that they are conditional variables from January 1984 to September 2006. They are obtained from the Federal Reserve Bank of St. Louis.

2.3 Regression framework: what models do we use?

2.3.1 Fama French 3-factor CAPM (1993)

Fama and French (1993) introduce SMB and HML to the traditional CAPM and form a 3-factor CAPM. Despite the lack of economic explanation for the newly added factors, the model explains much of the variation in asset returns. In this study, we use the 3-factor CAPM as the base model.

3-factor CAPM:

$$R_{it} - R_{ft} = \alpha_i + \beta_{i,1}(R_{Mt} - R_{ft}) + s_i SMB_t + h_i HML_t + e_{it} \tag{1}$$

In this regression, R_{it} is the return of fund i in month t, R_{ft} is the risk-free rate (1-month U.S. Treasury yield) in month t, R_{Mt} is the market return in month t (the return of the value-weighted portfolio of NYSE, Amex, and Nasdaq stocks), and SMB_t and HML_t are the Fama and French (1993) size and value-growth returns.

2.3.2 Fama French 5-factor CAPM (2015)

After the propose of the 3-factor model, some researchers (Novy-Marx, 2013; Titman, Wei and Xie, 2004) point out that the 3-factor model fails to capture the returns related to profitability and investment. Using a dividend discount model, Fama and French (2015) also theoretically show that future profitability and investment determine expected return. Then, to further explain the missed returns, Fama and French (2015) propose a 5-factor CAPM which includes RMW and CMA, the two factors reflecting operation profitability and investment. The resulting model explains up to 94% of the variation in cross-sectional expected returns for typical portfolios in their paper. In our study, we assume that the model more accurately captures fund returns and use it as a comparison to the 3-factor CAPM.

5-factor CAPM:

$$R_{it} - R_{ft} = \alpha_i + \beta_{i,1}(R_{Mt} - R_{ft}) + s_i SMB_t + h_i HML_t + r_i RMW_t + c_i CMA_t + e_{it} \tag{2}$$

The additional factors not shown in the 3-factor CAPM, RMW_t and CMA_t, are the Fama and French (2015) operation profitability and investment factors. Other factors have the same meaning as those in the 3-factor CAPM.

2.3.3 Ferson and Schadt (1996)

The previous two models assume that the market beta is unconditionally stable. However, if market risk exposure for a fund changes over time, the two models may not be reliable and a conditional model with varying market beta will be more appropriate. Ferson (1996) suggests the reasons why models incorporating changing market beta are more attractive. The first is that traditional methods are unable to explain the dynamic behaviors of asset returns. The second reason states that the complex trading strategies adopted by managers lead to more dynamic changes in asset returns. They also present an example of a mutual fund to illustrate that a fund's market beta can change. Suppose expected market return is in positive proportion with market volatility. As the mutual fund wishes to keep its volatility stable over time, it will lower its market beta when market volatility is high and raise it when low. If an average beta is used, its application to average market premium will be higher than the fund's average excess return because the low market return with a high beta cannot compensate for the high market return forgone with a low beta. This will lead to a negative alpha in the unconditional model while the true alpha is zero. This reflects the drawbacks of unconditional models and call for conditional ones.

To capture the confounding changes in fund beta, we need to assume that the available public information is fully reflected in the market price, since the use of public information should not lead to abnormal returns in a semi-strong market efficiency.

Then we construct the conditional 3- and 5-factor conditional CAPM as follows.

$$\begin{aligned} R_{it} - R_{ft} =& \alpha_i + \beta_{i,1}(R_{Mt} - R_{ft}) + s_i SMB_t + h_i HML_t \\ &+ B_{i,2}(Z_{t-1} - \bar{Z})(R_{Mt} - R_{ft}) + e_{it} \end{aligned} \tag{3}$$

$$\begin{aligned} R_{it} - R_{ft} =& \alpha_i + \beta_{i,1}(R_{Mt} - R_{ft}) + s_i SMB_t + h_i HML_t + r_i RMW_t + c_i CMA_t \\ &+ B_{i,2}(Z_{t-1} - \bar{Z})(R_{Mt} - R_{ft}) + e_{it} \end{aligned} \tag{4}$$

where Z_{t-1} is the vector of information variables measured at t-1, $\bar{Z}$ is their mean and $B_{2,i}$ is the vector of coefficients of information variables. Other variables are the same with those in unconditional models.

These two models are based on the previous 3- and 5-factor CAPMs and therefore retain their form. The only difference is the inclusion of an information variable in the market beta to make it reflect the influence of public information. In both models, the dynamic market beta is set to $(\beta_{i,1} + B_{i,2}(Z_{t-1} - \bar{Z}))$. It is a linear function of the information variables and thus makes the models easy to manipulate and interpret.

We use each of these 4 models to regress each fund's return on the explanatory factors. For each fund and each model, we can obtain the t-value of the regression intercept estimate (we denote as t(α) estimate). By pooling the t(α) estimates of all funds, we can obtain the distribution of t(α) estimates for each model. The α estimates measure the return of the actively managed funds in access of that of a passive fund. A positive α estimate indicates a good performance in the long run, while a negative α estimate indicates the opposite. Thus, we can examine the distribution of t(α) estimates and see the overall performance of actively managed mutual funds.

3 Results: Any difference among different models?

3.1 Overview of Section 4

This section provides the distributions of the t(α) estimates and relevant analysis.

Table 1 gives descriptive statistics (mean and standard deviation) of the cross-section of t(α) estimates for each combination of model, gross or net returns and TNA threshold. We have combinations of number of factors and gross or net return so there are four groups (3-Factor Gross Returns, 5-Factor Gross Returns, 3-Factor Net Returns and 5-Factor Net Returns). There are also Unconditional and Conditional columns, representing results of unconditional models and conditional models respectively. There are three sub-columns under each column (5 million, 250 million or 1 billion), and each column represents a TNA threshold. This means that each column only shows the cross-sections of t(α) estimates of funds that reach the TNA threshold (5 million, 250 million or 1 billion).

Table 2 shows the percentiles corresponding to the cross-section of estimated t(α) estimates for each combination of model, gross or net returns and TNA threshold. Columns are the same as Table 1.

Besides, we show graphs of the distributions of t(α) estimates under each combination of model, gross or net returns and TNA threshold in Appendix 1.

Table 1: descriptive statistics of t(α) estimates

	Unconditional			Conditional		
	5 million	250 million	1 billion	5 million	250 million	1 billion
3-Factor Gross Returns						
mean	-0.015	-0.018	0.024	0.163	0.132	0.132
std	0.769	0.689	0.721	0.797	0.731	0.763
5-Factor Gross Returns						
mean	-0.087	-0.095	-0.060	0.098	0.063	0.041
std	0.833	0.746	0.761	0.872	0.815	0.858
3-Factor Net Returns						
mean	-0.278	-0.241	-0.178	-0.098	-0.088	-0.067
std	0.799	0.710	0.742	0.824	0.749	0.781
5-Factor Net Returns						
mean	-0.347	-0.314	-0.257	-0.163	-0.155	-0.154
std	0.862	0.766	0.780	0.897	0.830	0.874

Table 2: percentiles of t(α) estimates

	Unconditional			Conditional		
Pct	5 million	250 million	1 billion	5 million	250 million	1 billion
3-Factor Gross Returns						
1	-1.85	-1.71	-1.55	-1.80	-1.61	-1.55
2	-1.55	-1.45	-1.45	-1.48	-1.42	-1.43
3	-1.43	-1.27	-1.30	-1.37	-1.30	-1.36
4	-1.31	-1.17	-1.22	-1.25	-1.21	-1.25
5	-1.22	-1.11	-1.17	-1.18	-1.12	-1.17
10	-0.95	-0.87	-0.90	-0.85	-0.81	-0.89
20	-0.63	-0.59	-0.57	-0.51	-0.48	-0.50
30	-0.42	-0.38	-0.35	-0.24	-0.21	-0.23
40	-0.25	-0.21	-0.16	-0.01	-0.03	-0.04
50	-0.06	-0.04	0.02	0.18	0.15	0.16
60	0.14	0.14	0.18	0.38	0.33	0.34
70	0.37	0.36	0.37	0.58	0.54	0.51
80	0.63	0.56	0.57	0.83	0.75	0.77
90	0.99	0.85	0.95	1.16	1.02	1.07
95	1.25	1.08	1.31	1.43	1.27	1.37
96	1.33	1.21	1.37	1.48	1.36	1.44
97	1.44	1.30	1.45	1.59	1.50	1.59
98	1.59	1.47	1.59	1.71	1.68	1.68
99	1.80	1.72	1.89	1.98	1.85	1.90
5-Factor Gross Returns						
1	-2.06	-1.91	-1.80	-2.10	-1.93	-2.15
2	-1.75	-1.66	-1.58	-1.78	-1.68	-1.78
3	-1.62	-1.51	-1.41	-1.53	-1.48	-1.67
4	-1.47	-1.37	-1.31	-1.42	-1.39	-1.48
5	-1.40	-1.29	-1.26	-1.34	-1.31	-1.38
10	-1.10	-0.99	-0.99	-1.01	-0.98	-1.08
20	-0.75	-0.70	-0.72	-0.63	-0.63	-0.62
30	-0.53	-0.48	-0.45	-0.33	-0.32	-0.37
40	-0.32	-0.31	-0.26	-0.10	-0.10	-0.16
50	-0.13	-0.11	-0.05	0.13	0.08	0.07
60	0.08	0.07	0.12	0.32	0.28	0.29
70	0.32	0.26	0.31	0.55	0.50	0.50
80	0.60	0.51	0.49	0.82	0.74	0.74
90	0.98	0.84	0.89	1.17	1.04	1.03
95	1.30	1.11	1.19	1.46	1.35	1.39
96	1.39	1.19	1.34	1.54	1.45	1.48
97	1.50	1.30	1.61	1.65	1.54	1.55
98	1.67	1.57	1.71	1.80	1.77	1.81
99	1.84	1.96	1.95	2.01	2.02	2.14

continue

	Unconditional			Conditional		
Pct	5 million	250 million	1 billion	5 million	250 million	1 billion
3-Factor Net Returns						
1	-2.18	-1.88	-1.88	-2.16	-1.87	-1.87
2	-1.91	-1.76	-1.72	-1.86	-1.66	-1.66
3	-1.78	-1.59	-1.52	-1.65	-1.56	-1.59
4	-1.66	-1.46	-1.44	-1.57	-1.48	-1.53
5	-1.57	-1.40	-1.38	-1.47	-1.38	-1.47
10	-1.27	-1.14	-1.12	-1.19	-1.06	-1.07
20	-0.93	-0.82	-0.77	-0.79	-0.71	-0.73
30	-0.68	-0.62	-0.55	-0.51	-0.45	-0.41
40	-0.51	-0.44	-0.36	-0.27	-0.24	-0.27
50	-0.31	-0.26	-0.18	-0.07	-0.08	-0.04
60	-0.10	-0.08	-0.01	0.14	0.13	0.15
70	0.15	0.15	0.15	0.37	0.33	0.33
80	0.41	0.39	0.39	0.61	0.52	0.60
90	0.76	0.64	0.76	0.93	0.84	0.88
95	1.03	0.90	1.19	1.19	1.03	1.20
96	1.09	1.04	1.23	1.27	1.18	1.27
97	1.19	1.16	1.31	1.33	1.30	1.37
98	1.36	1.23	1.39	1.48	1.44	1.55
99	1.54	1.43	1.55	1.73	1.66	1.80
5-Factor Net Returns						
1	-2.37	-2.17	-2.11	-2.35	-2.20	-2.27
2	-2.11	-1.98	-1.86	-2.09	-1.91	-2.11
3	-1.95	-1.77	-1.66	-1.88	-1.76	-1.90
4	-1.84	-1.66	-1.51	-1.76	-1.67	-1.71
5	-1.75	-1.58	-1.48	-1.68	-1.56	-1.61
10	-1.42	-1.22	-1.23	-1.33	-1.18	-1.29
20	-1.02	-0.94	-0.92	-0.89	-0.87	-0.82
30	-0.79	-0.70	-0.66	-0.61	-0.56	-0.56
40	-0.59	-0.53	-0.44	-0.36	-0.31	-0.35
50	-0.39	-0.34	-0.25	-0.13	-0.14	-0.15
60	-0.15	-0.15	-0.05	0.09	0.09	0.11
70	0.09	0.10	0.11	0.32	0.30	0.32
80	0.38	0.32	0.33	0.59	0.54	0.56
90	0.77	0.65	0.70	0.94	0.83	0.86
95	1.05	0.93	1.05	1.23	1.14	1.29
96	1.13	1.02	1.15	1.29	1.24	1.32
97	1.27	1.12	1.40	1.38	1.33	1.40
98	1.41	1.37	1.55	1.53	1.53	1.58
99	1.61	1.72	1.75	1.84	1.78	2.01

3.2 Summary of results: What can we discover?

We shall compare the 3-factor unconditional CAPM with the 5-factor unconditional CAPM, 3-factor conditional CAPM and 5-factor conditional CAPM respectively to see how the distributions of t(α) estimates change depending on the model specification.

We firstly concentrate on the Unconditional column of Table 1 , on a gross return basis. The

5-factor unconditional CAPM has lower means (-0.087 <-0.015, -0.095 <-0.018, -0.060 <0.024) but higher standard deviations (0.833 >0.769, 0.746 >0.689, 0.761 >0.721) than the 3-factor model under any TNA threshold. This suggests that fund managers' performance under the 5-factor CAPM is generally worse than under the 3-factor CAPM, but more diversified. This phenomenon can also be seen on the basis of net returns.

Then we turn to Table 2, concentrating on the Unconditional column on a gross return basis, and compare the percentiles of the 3-factor model with those of the 5-factor model. The 1st percentile of 3-factor model is -1.85, indicating that the fund performing better than precisely 1% of all funds have a t(α) estimate of -1.85, which shows a poor performance. We find that under 5 million and 250 million threshold, t(α) estimates of 5 factor model are smaller than those of the 3 factor model at the 1th to 90th percentile and larger at the 95th to 99th percentile. It suggests that most fund managers perform worse in the 5 factor model. It also provides the additional information that it is the ordinary managers (managers who neither achieve a high t(α) estimate nor a low t(α) estimate) and bottom managers (managers who achieve a low t(α) estimate) that perform worse while the top managers (managers who achieve a high t(α) estimate) perform even better. Similar conclusions can be drawn under 1 billion threshold with fund managers in 5-factor model performing worse at the 1st to 96th percentile and better at the 97th to 99th percentile than in 3-factor model. A similar phenomenon can be seen in the comparison between the two models on a net return basis.

After comparing the 3-factor unconditional CAPM with the 5-factor unconditional CAPM, we make comparisons between the 3-factor unconditional CAPM and the 3-factor conditional CAPM. In Table 1, we first focus on the 3-Factor Gross Returns group and find that Conditional column exhibits higher means under each TNA threshold (0.163 >-0.015, 0.132 >-0.018, 0.132 >0.024) with slightly larger standard deviations, indicating that the conditional CAPM's t(α) estimates have distributions similar to that of the unconditional CAPM after a positive level shift. Thus, it appears that fund managers in general show higher skill under Conditional column. Besides, the 3-Factor Net Returns group shows similar results.

We then look at Table 2 and focus on the 3-Factor Gross Returns group. We find that the t(α) estimates of the conditional model are overall higher than those of the unconditional model except for the 3rd, 4th and 5th percentile under the 250 million threshold and the 3rd and 4th percentile under the 1 billion threshold. This again shows that the distribution of t(α) estimates of the conditional model is like that of the unconditional model after a horizontal shift, indicating that all top, ordinary and bottom fund managers perform better in the conditional model. This is different from the 5-factor unconditional CAPM which reveals worse performance of ordinary and bottom managers and better performance of top managers. 3-Factor Net Returns group also shows similar results with almost all t(α) estimates under Conditional column larger than their counterparts under Unconditional column.

Finally we compare the 3-factor unconditional CAPM with the 5-factor conditional CAPM. In Table 1, we find that the means and standard deviations of 5-Factor Gross Returns group under Conditional column are all larger than those in 3-Factor Gross Returns under Unconditional column under any TNA threshold (0.098 >-0.015, 0.063 >-0.018, 0.041 >0.024; 0.872 >0.769, 0.815 >0.689, 0.858 >0.721). It implies that performances of fund managers are better but more diversified under 5-factor conditional CAPM, which is like the combined effect of 5-factor unconditional CAPM and 3-factor conditional CAPM. Similar findings also hold on a net return basis.

We then turn our attention to Table 2. Under any TNA threshold, the 1st to 10th percentile of 5-Factor Gross Returns group under Conditional column are less than their counterparts in 3-Factor Gross Returns group under Unconditional column, suggesting that bottom managers perform worse under the 5-factor conditional CAPM. For percentiles from 20th to 99th, the t(α) estimates of 5-factor conditional CAPM are generally greater than those of the original 3-factor unconditional CAPM, regardless of TNA threshold. And as the percentile approaches 99th, the difference between the t(α) estimates of the two models becomes more stable. For example, under 5 million TNA threshold, the 20th percentile of the t(α) estimate of the 5-factor conditional CAPM is -0.63, the same as that of the 3 factor unconditional CAPM. However, the difference between the two soon becomes larger as percentile increases, and keeps around 0.2 from the 50th percentile onwards. The phenomenon implies that ordinary managers perform better under the 5-factor conditional CAPM, while the top managers perform even much better. A similar finding can be made on a net return basis.

Overall, the use of unconditional 5-factor CAPM shows that most fund managers (ordinary and bottom) may perform worse than originally found in 3-factor unconditional CAPM and only the top managers who can produce high t(α) estimates seem to perform better. Using the 3-factor conditional CAPM reveals that fund managers, not only the top but also the ordinary and bottom, can perform better than previously shown in the 3-factor unconditional CAPM. This is consistent with what's revealed by Ferson and Schadt (1996). The 5-factor conditional model, however, is a hybrid of the previous two, showing that bottom managers may perform worse. In contrast, the ordinary managers perform better, but not to the extent of top managers.

4 Discussion: Why is the result like that and are there any limitations?

4.1 Overview of Section 4

In this section we discuss the possible reasons of the results we see and the limitations of the research.

4.2 Why is the result like that?

Funds overall perform worse under the 5-factor unconditional model. The change is due to the newly added factors RMW and CMA. This suggests that these two factors can explain positive aspects of fund return.

For the 3-factor conditional CAPM, the improvement in fund performance is probably due to a negative relationship between fund betas and conditional expected market returns (Ferson and Schadt, 1996), indicating that in a world where expected market return is positively correlated with market return variation, funds normally forego a large part of return brought by high expected market return and earn meager excess return when market return is low, which results in lower fund returns than if keeping a constant average beta. Unconditional models fail to account for it, overestimate the fund return explained by market returns, and therefore estimate a lower fund α and conclude with underestimated fund performance.

The effects of RMW, CMA and information variables don't intersect each other and therefore the 5-factor conditional CAPM, which is a combination of 5-factor unconditional CAPM and 3-factor conditional CAPM, exerts a combined effect of the two models.

4.3 Limitations: Our study is not enough.

In this study we construct models in comparison with the 3-factor unconditional CAPM in Fama French (2010). They include 5-factor unconditional CAPM, 3-factor conditional CAPM and 5-factor conditional CAPM. These models take into account more factors affecting α estimates and information variables used in fund managers' trading strategies, in an attempt to better explain mutual fund abnormal returns. However, there are also some limitations in this study.

On one hand, the economic interpretation of size, book-to-market ratio, operation profitability and investment in explaining fund returns is not clear, which is also stated in Fama and French (1993, p. 450).

On the other hand, there are some details that may be considered inappropriate when dealing with mutual fund data. We combined the share classes of a fund by following a fund whose share classes have the same first four codes. This can be inappropriate because we observe that some funds have share classes with different first four digit codes. However, there are not many such fund share classes, so this is not a serious issue. In addition, we exclude index funds and hedge funds by keyword filtering, which may not be accurate because some index funds may not show evidence of their category in their names.

5 Conclusions: What does the study reveal?

To summarize, we find that the inclusion of RMW and CMA in the 5-factor CAPM leads to better performance for top funds and worse performance for other funds, while the information variables added to market beta implies an overall better performance. RMW, CMA and conditional term of market return altogether pose a combined effect, which indicates worse performance for bottom managers, better performance for ordinary managers and much better performance for top managers. This study is a small step in improving the Fama French (2010) 3-factor unconditional CAPM, with new factors to account for fund αs. Future researches are needed to explore more reasonable models to evaluate actively managed mutual funds' performance.

6 References

Brooks, R.D., Faff, R.W. and Lee, J.H.H. (1992) 'The form of time variation of systematic risk: some Australian evidence', *Applied Financial Economics*, 2 (4), pp. 191-198.

Brown, S.J. and Goetzmann, W.N. (1995) 'Performance Persistence', *The Journal of Finance*, 50 (2), pp. 679-698.

Chistopherson, J.A., Ferson, W.E. and Glassman, D.A. (1998) 'Conditioning Manager Alphas on Economic Information: Another Look at the Persistence of Performance', *The Review of Financial Studies*, 11 (1), pp. 111-142.

Cochrane, J.H. (2009) *Asset pricing: Revised edition.* Princeton university press.

Fama, E.F. and French, K.R. (1992) 'The Cross-Section of Expected Stock Returns', *The Journal of Finance*, 47 (2), pp. 427–465.

Fama, E.F. and French, K.R. (1993) 'Common risk factors in the returns on stocks and bonds', *Journal of Financial Economics*, 33 (1), pp. 3–56.

Fama, E.F. and French, K.R. (1996) 'Multifactor Explanations of Asset Pricing Anomalies', *The Journal of Finance*, 51 (1), pp. 55–84.

Fama, E.F. and French, K.R. (2010) 'Luck versus Skill in the Cross-Section of Mutual Fund Returns', *The Journal of Finance*, 65 (5), pp. 1915–1947.

Fama, E.F. and French, K.R. (2015) 'A five-factor asset pricing model', *Journal of Financial Economics*, 116 (1), pp. 1–22.

Ferson, W. and Harvey, C. (1991) 'The variation of economic risk premiums', *Journal of Political Economy*, 99 (2), pp. 527–566.

Ferson, W.E. and Schadt, R.W. (1996) 'Measuring Fund Strategy and Performance in Changing Economic Conditions', *The Journal of Finance*, 51 (2), pp. 425–461.

Grinblatt, M. and Titman, S. (1992) 'The Persistence of Mutual Fund Performance', *The Journal of Finance*, 47 (5), pp. 1977-1984.

Gruber, M.J. (2011) 'Another Puzzle: The Growth in Actively Managed Mutual Funds', *The Journal of Finance*, 51 (3), pp. 783-810.

Hendricks, D., Patel, J. and Zeckhauser, R. (1993) 'Hot Hands in Mutual Funds: Short-Run Persistence of Relative Performance, 1974–1988', *The Journal of Finance*, 48 (1), pp. 93-130.

Huang, R., Asteriou, D. and Pouliot, W. (2020) 'A reappraisal of luck versus skill in the cross-section of mutual fund returns', *Journal of Economic Behavior and Organization*, 176, pp. 166-187.

Henriksson, R.D. (1984) 'Market Timing and Mutual Fund Performance: An Empirical Investigation', *The Journal of Business*, 57 (1), pp. 73–96.

Iqbal, J., Brooks, R. and Galagedera, D.U.A. (2010) 'Testing conditional asset pricing models: An emerging market perspective', *Journal of International Money and Finance*, 29 (5), pp. 897-918.

Jagannathan, R. and Wang, Z. (1996) 'The Conditional CAPM and the Cross-Section of Expected Returns', *The Journal of Finance*, 51 (1), pp. 3–53.

Jensen, M.C. (1968) 'The Performance of Mutual Funds in the Period 1945-1964', *The Journal of Finance*, 23 (2), pp. 389–416.

Lintner, J. (1965) 'Security Prices, Risk, and Maximal Gains From Diversification', *The Journal of Finance*, 20 (4), pp. 587–615.

Malkiel, B.G. (1995) 'Returns from Investing in Equity Mutual Funds 1971 to 1991', *The Journal of Finance*, 50(2), pp. 549–572.

Merton, R.C. (1973) 'An Intertemporal Capital Asset Pricing Model', *Econometrica*, 41 (5), pp. 867–887.

Mossin, J. (1966) 'Equilibrium in a Capital Asset Market', *Econometrica*, 34 (4), pp. 768–783.

Novy-Marx, R. (2013) 'The other side of value: The gross profitability premium', *Journal of Financial Economics*, 108 (1), pp. 1–28.

Sharpe, W.F. (1964) 'Capital Asset Prices: A Theory Of Market Equilibrium Under Conditions Of Risk', *The Journal of Finance*, 19 (3), pp. 425-442.

Sharpe, W.F. (1991) 'The Arithmetic of Active Management', *Financial Analysts Journal*, 47 (1), pp. 7–9.

Titman, S., Wei, K. and Xie, F. (2004) 'Capital Investments and Stock Returns', *Journal of Financial and Quantitative Analysis*, 39 (4), pp. 677–700.

7 Appendix 1: distributions of t(α) estimates

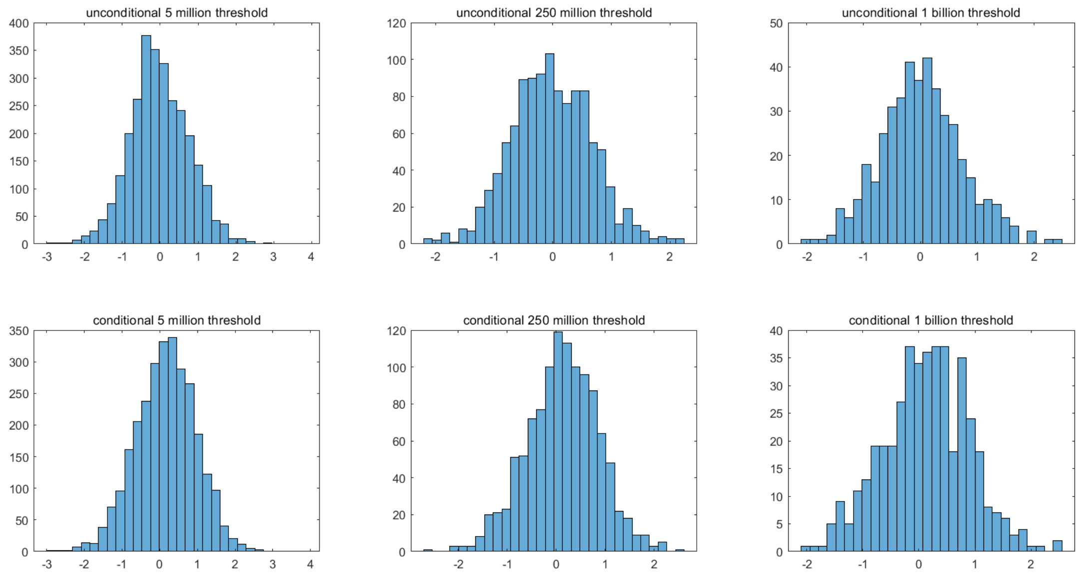

Figure 1: Comparison of conditional and unconditional model: 3-factor gross return t-value distribution

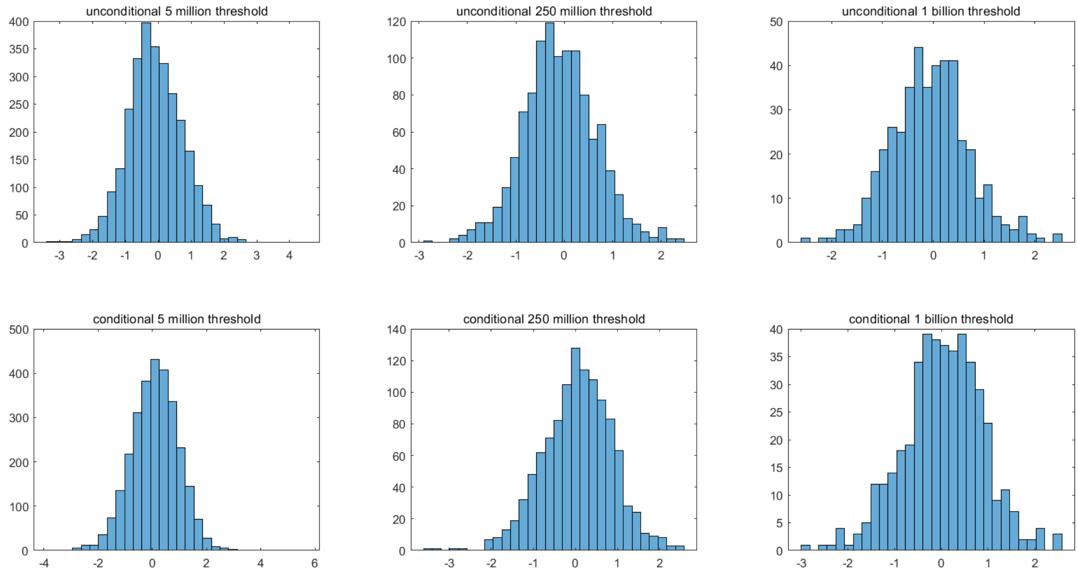

Figure 2: Comparison of conditional and unconditional model: 5-factor gross return t-value distribution

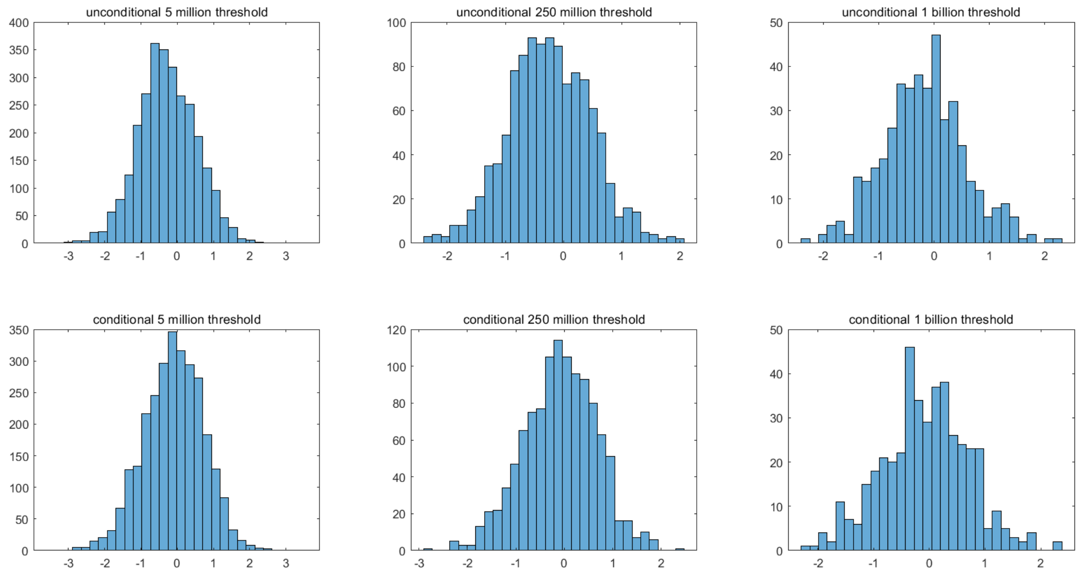

Figure 3: Comparison of conditional and unconditional model: 3-factor net return t-value distribution

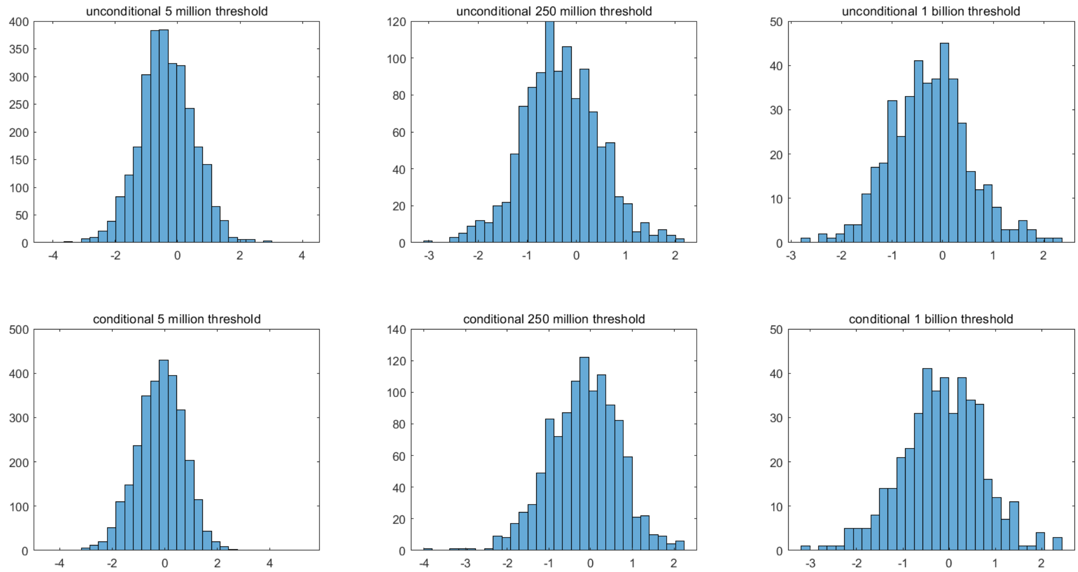

Figure 4: Comparison of conditional and unconditional model: 5-factor net return t-value distribution

8 Appendix 2: Filtering passively managed funds

As we study actively managed mutual funds, we need to exclude passively managed funds and keep only the funds mainly invested in US equities. We use keyword filtering to exclude passively managed funds by checking whether there are certain keywords in the funds' names. We mainly follow the procedures in Huang, Asteriou and Pouliot (2020).

The passively managed funds are classified as Enhanced index fund, Target-date funds and Leveraged index fund, and the requirements to be any of the funds are listed in the right column in Table 3. If a fund doesn't meet any of the requirements, then it won't be classified as any of the categories and is not a passively managed fund, which will be retained for further analysis.

Table 3: Keyword requirements for different sorts of passively managed funds

Category	Keywords requirement in a fund's name
Enhanced index fund	contain any of {"INDEX","INDX","IDX","S&P","SCHWAB 1000", "NASDAQ-100","DOW","JONES","ETF","ISHARE","PROFOUND", "RUSSELL","PROSHARE","POWERSHARE","VIPER","SPIDER", "SPDR","WILSHARE","ETN","EXCHANGE TRADED", "EXCHANGE-TRADED","HEDGE","MANAGE","ENHANCE", "PLUS"}
Target-date fund	contain any of {"1970","1975","1980","1985","1990","1995", "2000","2005","2010","2015","2020","2025","2030","2035","2040", "2045","2050","2055","2060","TARGET","LIFESTYLE"}
leveraged index fund	contain any of {"INVERSE","SHORT","ULTRA","1.25X","1.5X", "2X","2.5X","3X","4X","5X","6X","7X","8X","9X","0X"} and contain none of {"SHORT TERM","SHORT TM", "SHORT BOND","SHORT BND","SHORT BD","LG SHORT", "LONG/SHORT","LONG-SHORT","LONG SHORT"}

9 Appendix 3: Matlab code to conduct the study

```
%% get and combine live and dead RI, and get gross return

% get US live fund data
clear ID Ind opts2
opts2 = spreadsheetImportOptions('Numvariables',465);
opts2.Sheet = 'total return index';
opts2.VariableTypes(1:2) = {'string'};
opts2.VariableTypes(3:end) = {'double'};
opts2.DataRange = 'A2:QW7409';
USliveRI = readtable('USLiveEquityFunds.xlsx',opts2);
USliveRI(:,3:13) = [];
USliveRI(:,277:end) = [];

% get US dead fund data
clear ID Ind opts2
opts2 = spreadsheetImportOptions('Numvariables',627);
opts2.Sheet = 'total return index';
```

```
opts2.VariableTypes(1:2) = {'string'};

opts2.VariableTypes(3:end) = {'double'};

opts2.DataRange = 'A2:XC1617';

USDeadRI = readtable('USDeadEquityFunds.xlsx', opts2);

USDeadRI(:,3:175) = [];

USDeadRI(:,277:end) = [];

nRowCol = size(USDeadRI);

for ind = 1:nRowCol(1)

    md = mode(USDeadRI{ind,3:end},'all');

    a = USDeadRI{ind,3:end} == md;

    a = logical([0 0 a]);

    USDeadRI{ind,a} = NaN;

end

% change variable name to combine

allVars = 1:width(USDeadRI);

newNames = append("Reading",string(allVars));

USliveRI = renamevars(USliveRI,allVars,newNames);

USDeadRI = renamevars(USDeadRI,allVars,newNames);

% combine live and dead funds and calculate gross return

USRI = [USliveRI; USDeadRI];

USRIName = USRI(:,1); % get fund name

USId = USRI(:,2); % get fund id

USRI(:,1) = [];

nRowCol = size(USRI);

USRetGrs = NaN(nRowCol(1),(nRowCol(2)-1));

for ind = 1:nRowCol(1)
```

```
    row = table2array(USRI(ind,2:end));
    notnan = ~isnan(row);
    noNaN = row(notnan);
    grt = NaN(1,length(noNaN));
    for in = 2:length(noNaN)
        grt(in) = (noNaN(in)/noNaN(in-1))-1;
    end
    row(notnan) = grt;
    zero = row == 0;
    row(zero) = NaN;
    USRetGrs(ind,2:end) = row(2:end);
end

%% get and combine live and dead TER

% get US live fund data
clear ID Ind opts2
opts2 = spreadsheetImportOptions('Numvariables',465);
opts2.Sheet = 'total expense ratio';
opts2.VariableTypes(1:2) = {'string'};
opts2.VariableTypes(3:end) = {'double'};
opts2.DataRange = 'A2:QW7409';
USliveTER = readtable('USLiveEquityFunds.xlsx',opts2);
USliveTER(:,2:13) = [];
USliveTER(:,276:end) = [];

% get US dead fund data
clear ID Ind opts2
```

```
opts2 = spreadsheetImportOptions('Numvariables',627);

opts2.Sheet = 'total expense ratio';

opts2.VariableTypes(1:2) = {'string'};

opts2.VariableTypes(3:end) = {'double'};

opts2.DataRange = 'A2:XC1617';

USDeadTER = readtable('USDeadEquityFunds.xlsx', opts2);

USDeadTER(:,2:175) = [];

USDeadTER(:,276:end) = [];

nRowCol = size(USDeadTER);

for ind = 1:nRowCol(1)

    md = mode(USDeadTER{ind,2:end},'all');

    a = USDeadTER{ind,2:end} == md;

    a = logical([0 a]);

    USDeadTER{ind,a} = NaN;

end

allVars = 1:width(USDeadTER);

newNames = append("Reading",string(allVars));

USliveTER = renamevars(USliveTER,allVars,newNames);

USDeadTER = renamevars(USDeadTER,allVars,newNames);


% combine the two dataset

USTER = [USliveTER; USDeadTER];

USTERTrim = USTER;

USTERTrim(:,2) = []; % USTERTrim is USTER without data for Dec 1983


%% get and combine live and dead TNA


% get US live fund data
```

```
clear ID Ind opts2

opts2 = spreadsheetImportOptions('Numvariables',465);

opts2.Sheet = 'total net asset';

opts2.VariableTypes(1:2) = {'string'};

opts2.VariableTypes(3:end) = {'double'};

opts2.DataRange = 'A2:QW7409';

USliveTNA = readtable('USLiveEquityFunds.xlsx',opts2);

USliveTNA(:,2:13) = [];

USliveTNA(:,276:end) = [];

% get US dead fund data

clear ID Ind opts2

opts2 = spreadsheetImportOptions('Numvariables',627);

opts2.Sheet = 'total net asset';

opts2.VariableTypes(1:2) = {'string'};

opts2.VariableTypes(3:end) = {'double'};

opts2.DataRange = 'A2:XC1617';

USDeadTNA = readtable( 'USDeadEquityFunds.xlsx', opts2);

USDeadTNA(:,2:175) = [];

USDeadTNA(:,276:end) = [];

nRowCol = size(USDeadTNA);

for ind = 1:nRowCol(1)

    md = mode(USDeadTNA{ind,2:end},'all');

    a = USDeadTNA{ind,2:end} == md;

    a = logical([0 a]);

    USDeadTNA{ind,a} = NaN;

end
```

```
allVars = 1:width(USDeadTNA);

newNames = append("Reading",string(allVars));

USliveTNA = renamevars(USliveTNA,allVars,newNames);

USDeadTNA = renamevars(USDeadTNA,allVars,newNames);


% combine the two dataset

USTNA = [USliveTNA; USDeadTNA];

USTNATrim = USTNA;

USTNATrim(:,2) = []; % USTNATrim is USTNA without data for Dec 1983


%% read factor data


opts = spreadsheetImportOptions('Numvariables',7);

opts.Sheet = '5factor';

opts.VariableTypes = {'string' 'double' 'double' 'double' 'double' 'double' 'double'};

opts.DataRange = 'A251:G523';

US5Factor = readtable('USFactors.xlsx', opts);

US3Factor = US5Factor;

US3Factor(:,5:6) = [];


%% read factor data with information variable


% conditional model based on 3 factor CAPM

clear ID Ind opts

opts = spreadsheetImportOptions('Numvariables',9);

opts.Sheet = '3fcon';

opts.VariableTypes(1) = {'string'};

opts.VariableTypes(2:end) = {'double'};

opts.DataRange = 'A1:I273';
```

```
US3Factor_con = readtable('conFactors.xlsx', opts);

% conditional model based on 5 factor CAPM
clear ID Ind opts
opts = spreadsheetImportOptions('Numvariables',11);
opts.Sheet = '5fcon';
opts.VariableTypes(1) = {'string'};
opts.VariableTypes(2:end) = {'double'};
opts.DataRange = 'A1:K273';
US5Factor_con = readtable('conFactors.xlsx', opts);

%% get net return

USTER_temp = USTER(:,2:end);
USRI_temp = USRI(:,2:end);
nRowCol = size(USRI);
USRetNet = NaN(nRowCol(1),nRowCol(2)-1);
row1 = NaN(1,(nRowCol(2)-1));
row2 = NaN(1,(nRowCol(2)-1));
row4 = NaN(1,(nRowCol(2)-1));
j=1;
for ind = 1:nRowCol(1)
    notnan = ~isnan(table2array(USRI(ind,2:end)));
    row1(notnan) = table2array(USRI_temp(ind,notnan));
    rows = row1(notnan);
    row2(notnan) = [NaN rows(2:end)];
    row4(notnan) = [NaN rows(1:(end-1))];
    ret(j,notnan) = row2(notnan)./row4(notnan)-1;
```

```
        row3 = table2array(USTER_temp(ind,notnan));

        row3nonan =  isnan(row3);

        mrow3 = mean(row3(row3nonan),'all');

        for ind2 = 1:length(row3)

            if isnan(row3(ind2))

                row3(ind2) = mrow3;

            end

        end

        USRetNet(j,notnan) = ret(j,notnan) - row3./1200;

        j = j+1 ;

    end

    %% remove funds with no data and exclude passively managed funds

    nColRowUSRI = size(USRI);

    array_USRetGrs = USRetGrs;

    USRetGrs = array2table(USRetGrs);

    array_USRetNet = USRetNet;

    USRetNet = array2table(USRetNet);

    array_USTNATrim = table2array(USTNATrim(:,2:end));

    array_USTNA = table2array(USTNA(:,2:end));

    array_USRIName = table2array(USRIName);

    % funds with any of these keywords are regarded as enhanced index fund or target-date fund

    Enhanced_index_fund = ["INDEX", "INDX", "IDX", "SP", "SCHWAB 1000", "NASDAQ-100", "DOW",...
"JONES", "ETF", "ISHARE", "PROFOUND", "RUSSELL", "PROSHARE", "POWERSHARE", "VIPER",...
"SPIDER", "SPDR", "WILSHARE", "ETN", "EXCHANGE TRADED", "EXCHANGE-TRADED", "HEDGE",...
"MANAGE", "ENHANCE", "PLUS", "1970", "1975", "1980", "1985", "1990", "1995", "2000", "2005", "2010",...
"2015", "2020", "2025", "2030", "2035", "2040", "2045", "2050", "2055", "2060", "TARGET", "LIFESTYLE"];
```

% funds with any of these keywords in leveraged_index_fund_in and none of the keywords in lever-%aged_index_fund_ex are regarded as leveraged index fund

leveraged_index_fund_in = ["INVERSE", "SHORT", "ULTRA", "1.25X", "1.5X", "2X", "2.5X", "3X",... "4X", "5X", "6X", "7X", "8X", "9X", "0X"];

leveraged_index_fund_ex = ["SHORT TERM", "SHORT TM", "SHORT BOND", "SHORT BND",... "SHORT BD", "LG SHORT", "LONG/SHORT", "LONG-SHORT", "LONG SHORT"];

% conduct fund filter to filter funds with no data and passively managed funds

array_USRetGrs_filter = array_USRetGrs(sum(array_USRetNet >-100,2) >0 & sum(array_USTNA... >0,2) >0 & ~contains(array_USRIName, Enhanced_index_fund) & ~(contains (array_USRIName, ... leveraged_index_fund_in) & ~contains(array_USRIName, leveraged_index_fund_ex)),:);

array_USRetNet_filter = array_USRetNet(sum(array_USRetNet >-100,2) >0 & sum(array_USTNA >0,2)... >0 & ~contains(array_USRIName, Enhanced_index_fund) & ~(contains(array_USRIName, leveraged_index_fund_in)... & ~contains(array_USRIName, leveraged_index_fund_ex)),:);

array_USTNATrim_filter = array_USTNATrim(sum(array_USRetNet >-100,2) >0 & sum(array_USTNA... >0 , 2) >0 & ~contains (array_USRIName, Enhanced_index_fund) & ~(contains (array_USRIName,... leveraged_index_fund_in) & ~contains(array_USRIName, leveraged_index_fund_ex)),:);

USRIName_filter = USRIName(sum(array_USRetNet >-100,2) >0 & sum(array_USTNA >0,2) >0 & ... ~contains(array_USRIName, Enhanced_index_fund) & ~(contains (array_USRIName, leveraged_index_fund_in...) & ~contains(array_USRIName, leveraged_index_fund_ex)),:);

USId_filter = USId(sum(array_USRetNet >-100,2) >0 & sum (array_USTNA >0, 2) >0 & ~contains (... array_USRIName, Enhanced_index_fund) & ~(contains (array_USRIName, leveraged_index_fund_in) ... & ~contains(array_USRIName, leveraged_index_fund_ex)),:);

%% combine the same fund in different classes (using funds' ids)

% get the position of funds which change classes

nColRowarray_USRetNet_filter = size(array_USRetNet_filter);

diff_sig = ones(1,nColRowarray_USRetNet_filter(1)+1);

for i=2:nColRowarray_USRetNet_filter(1)

char_seq_1 = convertStringsToChars(string(USId_filter.Reading2(i,1)));

char_seq_2 = convertStringsToChars(string(USId_filter.Reading2(i-1,1)));

if char_seq_1(1:4) == char_seq_2(1:4)

diff_sig(1,i) = 0;

```
    end
end

% get AGTNA
diff_index = find(diff_sig);
USAGTNA = array_USTNATrim_filter;
USAGTNA = array2table(USAGTNA);
diff_index_length = length(diff_index);
for i=1:diff_index_length-1
    temp = array2table(sum(array_USTNATrim_filter(diff_index( i):diff_index(i+1)-1,:),1));
    for j=diff_index(i):diff_index(i+1)-1
        USAGTNA(j,:) = temp;
    end
end

% get combined fund return
weight = array_USTNATrim_filter./table2array(USAGTNA);
temp_RetGrs = weight.*array_USRetGrs_filter(:,2:end);
temp_RetNet = weight.*array_USRetNet_filter(:,2:end);
RetGrs = USRetGrs(1:diff_index_length-1,:);
RetNet = USRetNet(1:diff_index_length-1,:);
for i=1:diff_index_length-1
    temp = array2table(sum(temp_RetGrs(diff_index(i):diff_index(i+1)-1,:),1));
    RetGrs(i,2:end) = temp;
end
for i=1:diff_index_length-1
    temp = array2table(sum(temp_RetNet(diff_index(i):diff_index(i+1)-1,:),1));
    RetNet(i,2:end) = temp;
```

```
end

% get unique AGTNA

AGTNA_unique = USAGTNA(1:diff_index_length-1,:);

for i=1:diff_index_length-1

    AGTNA_unique(i,:) = USAGTNA(diff_index(i),:);

end

%% set TNA threshold get get three groups of funds (5 million, 250 million and 1 billion)

array_AGTNA_unique = double(table2array(AGTNA_unique));

nRowCol = size(RetGrs);

temp = array_AGTNA_unique >= 5;

temp2 = double(temp);

temp = zeros(nRowCol(1),nRowCol(2));

temp(:,2:end) = temp2;

temp(temp==0)=NaN;

temp = fillmissing(temp,'previous',2);

RetGrs_above5 = table2array(RetGrs).*temp;

RetGrs_above5 = RetGrs_above5(array_AGTNA_unique(:,end) >= 5,:);

RetGrs_above5 = array2table(RetGrs_above5);

RetNet_above5 = table2array(RetNet).*temp;

RetNet_above5 = RetNet_above5(array_AGTNA_unique(:,end) >= 5,:);

RetNet_above5 = array2table(RetNet_above5);

temp = array_AGTNA_unique >= 250;

temp2 = double(temp);

temp = zeros(nRowCol(1),nRowCol(2));

temp(:,2:end) = temp2;
```

```
temp(temp==0)=NaN;

temp = fillmissing(temp,'previous',2);

RetGrs_above250 = table2array(RetGrs).*temp;

RetGrs_above250 = RetGrs_above250(array_AGTNA_unique(:,end) >= 250,:);

RetGrs_above250 = array2table(RetGrs_above250);

RetNet_above250 = table2array(RetNet).*temp;

RetNet_above250 = RetNet_above250(array_AGTNA_unique(:,end) >= 250,:);

RetNet_above250 = array2table(RetNet_above250);

temp = array_AGTNA_unique >= 1000;

temp2 = double(temp);

temp = zeros(nRowCol(1),nRowCol(2));

temp(:,2:end) = temp2;

temp(temp==0)=NaN;

temp = fillmissing(temp,'previous',2);

RetGrs_above1000 = table2array(RetGrs).*temp;

RetGrs_above1000 = RetGrs_above1000(array_AGTNA_unique(:,end) >= 1000,:);

RetGrs_above1000 = array2table(RetGrs_above1000);

RetNet_above1000 = table2array(RetNet).*temp;

RetNet_above1000 = RetNet_above1000(array_AGTNA_unique(:,end) >= 1000,:);

RetNet_above1000 = array2table(RetNet_above1000);


%% get percentiles of t(α)s unconditional


% get each fund's t-value under each of the 12 sets (gross or net; TNA threshold and 3 factor CAPM or 5
% factor CAPM

nRowCol = size(RetGrs_above5);

Tstat_gross_above5_3factor = calculatealpha12Aug2021(RetGrs_above5,US3Factor);

Tstat_gross_above250_3factor = calculatealpha12Aug2021(RetGrs_above250,US3Factor);
```

```
Tstat_gross_above1000_3factor = calculatealpha12Aug2021(RetGrs_above1000,US3Factor);

Tstat_gross_above5_5factor = calculatealpha12Aug2021(RetGrs_above5,US5Factor);

Tstat_gross_above250_5factor = calculatealpha12Aug2021(RetGrs_above250,US5Factor);

Tstat_gross_above1000_5factor = calculatealpha12Aug2021(RetGrs_above1000,US5Factor);

Tstat_net_above5_3factor = calculatealpha12Aug2021(RetNet_above5,US3Factor);

Tstat_net_above250_3factor = calculatealpha12Aug2021(RetNet_above250,US3Factor);

Tstat_net_above1000_3factor = calculatealpha12Aug2021(RetNet_above1000,US3Factor);

Tstat_net_above5_5factor = calculatealpha12Aug2021(RetNet_above5,US5Factor);

Tstat_net_above250_5factor = calculatealpha12Aug2021(RetNet_above250,US5Factor);

Tstat_net_above1000_5factor = calculatealpha12Aug2021(RetNet_above1000,US5Factor);

% put the t(alpha)s into a single table

t_table_5factor = zeros(nRowCol(1),12)*nan;

t_table_5factor(1:length(Tstat_gross_above5_3factor),1) = Tstat_gross_above5_3factor;

t_table_5factor(1:length(Tstat_gross_above250_3factor),2) = Tstat_gross_above250_3factor;

t_table_5factor(1:length(Tstat_gross_above1000_3factor),3) = Tstat_gross_above1000_3factor;

t_table_5factor(1:length(Tstat_gross_above5_5factor),4) = Tstat_gross_above5_5factor;

t_table_5factor(1:length(Tstat_gross_above250_5factor),5) = Tstat_gross_above250_5factor;

t_table_5factor(1:length(Tstat_gross_above1000_5factor),6) = Tstat_gross_above1000_5factor;

t_table_5factor(1:length(Tstat_net_above5_3factor),7) = Tstat_net_above5_3factor;

t_table_5factor(1:length(Tstat_net_above250_3factor),8) = Tstat_net_above250_3factor;

t_table_5factor(1:length(Tstat_net_above1000_3factor),9) = Tstat_net_above1000_3factor;

t_table_5factor(1:length(Tstat_net_above5_5factor),10) = Tstat_net_above5_5factor;

t_table_5factor(1:length(Tstat_net_above250_5factor),11) = Tstat_net_above250_5factor;

t_table_5factor(1:length(Tstat_net_above1000_5factor),12) = Tstat_net_above1000_5factor;

% get t-value percentile table - t_table_percentile_5factor

percentile = [1,2,3,4,5,10,20,30,40,50,60,70,80,90,95,96,97,98,99];
```

```
for i=1:19

    t_table_percentile_5factor(i,:) = prctile(t_table_5factor,percentile(i));

end


%% get percentiles of t(alpha)s conditional


% get each fund's t-value under each of the 12 sets (gross or net; TNA threshold and 3 factor CAPM or 5
% factor CAPM)

Tstat_gross_above5_3factor_con = calculatealpha12Aug2021(RetGrs_above5,US3Factor_con);

Tstat_gross_above250_3factor_con = calculatealpha12Aug2021(RetGrs_above250,US3Factor_con);

Tstat_gross_above1000_3factor_con = calculatealpha12Aug2021(RetGrs_above1000,US3Factor_con);

Tstat_gross_above5_5factor_con = calculatealpha12Aug2021(RetGrs_above5, US5Factor_con);

Tstat_gross_above250_5factor_con = calculatealpha12Aug2021(RetGrs_above250, US5Factor_con);

Tstat_gross_above1000_5factor_con = calculatealpha12Aug2021(RetGrs_above1000,US5Factor_con);

Tstat_net_above5_3factor_con = calculatealpha12Aug2021(RetNet_above5,US3Factor_con);

Tstat_net_above250_3factor_con = calculatealpha12Aug2021(RetNet_above250,US3Factor_con);

Tstat_net_above1000_3factor_con = calculatealpha12Aug2021(RetNet_above1000,US3Factor_con);

Tstat_net_above5_5factor_con = calculatealpha12Aug2021(RetNet_above5,US5Factor_con);

Tstat_net_above250_5factor_con = calculatealpha12Aug2021(RetNet_above250,US5Factor_con);

Tstat_net_above1000_5factor_con = calculatealpha12Aug2021(RetNet_above1000,US5Factor_con);


% put the t(α)s into a single table

t_table_con = zeros(nRowCol(1),12)*nan;

t_table_con(1:length(Tstat_gross_above5_3factor_con),1) = Tstat_gross_above5_3factor_con;

t_table_con(1:length(Tstat_gross_above250_3factor_con),2) = Tstat_gross_above250_3factor_con;

t_table_con(1:length(Tstat_gross_above1000_3factor_con),3) = Tstat_gross_above1000_3factor_con;

t_table_con(1:length(Tstat_gross_above5_5factor_con),4) = Tstat_gross_above5_5factor_con;

t_table_con(1:length(Tstat_gross_above250_5factor_con),5) = Tstat_gross_above250_5factor_con;

t_table_con(1:length(Tstat_gross_above1000_5factor_con),6) = Tstat_gross_above1000_5factor_con;
```

```
t_table_con(1:length(Tstat_net_above5_3factor_con),7) = Tstat_net_above5_3factor_con;
t_table_con(1:length(Tstat_net_above250_3factor_con),8) = Tstat_net_above250_3factor_con;
t_table_con(1:length(Tstat_net_above1000_3factor_con),9) = Tstat_net_above1000_3factor_con;
t_table_con(1:length(Tstat_net_above5_5factor_con),10) = Tstat_net_above5_5factor_con;
t_table_con(1:length(Tstat_net_above250_5factor_con),11) = Tstat_net_above250_5factor_con;
t_table_con(1:length(Tstat_net_above1000_5factor_con),12) = Tstat_net_above1000_5factor_con;

% get t-value percentile table - t_table_percentile_con
percentile = [1,2,3,4,5,10,20,30,40,50,60,70,80,90,95,96,97,98,99];
for i=1:19
    t_table_percentile_con(i,:) = prctile(t_table_con,percentile(i));
end

%% get mean and standard deviation of t-values (unconditional and conditional respectively)

mean_and_std_table(1,1:12) = mean(t_table_5factor,'omitnan');
mean_and_std_table(2,1:12) = std(t_table_5factor,'omitnan');
mean_and_std_table(3,1:12) = mean(t_table_con,'omitnan');
mean_and_std_table(4,1:12) = std(t_table_con,'omitnan');

%% plot t(alpha) distributions

% gross_3factor
subplot(2,3,1)
histogram(t_table_5factor(:,1),30)
title('unconditional 5 million threshold')
subplot(2,3,2)
histogram(t_table_5factor(:,2),30)
title('unconditional 250 million threshold')
```

```
subplot(2,3,3)

histogram(t_table_5factor(:,3),30)

title('unconditional 1 billion threshold')

subplot(2,3,4)

histogram(t_table_con(:,1),30)

title('conditional 5 million threshold')

subplot(2,3,5)

histogram(t_table_con(:,2),30)

title('conditional 250 million threshold')

subplot(2,3,6)

histogram(t_table_con(:,3),30)

title('conditional 1 billion threshold')

% gross_5factor

subplot(2,3,1)

histogram(t_table_5factor(:,4),30)

title('unconditional 5 million threshold')

subplot(2,3,2)

histogram(t_table_5factor(:,5),30)

title('unconditional 250 million threshold')

subplot(2,3,3)

histogram(t_table_5factor(:,6),30)

title('unconditional 1 billion threshold')

subplot(2,3,4)

histogram(t_table_con(:,4),30)

title('conditional 5 million threshold')

subplot(2,3,5)

histogram(t_table_con(:,5),30)
```

```
title('conditional 250 million threshold')
subplot(2,3,6)
histogram(t_table_con(:,6),30)
title('conditional 1 billion threshold')
% net_3factor
subplot(2,3,1)
histogram(t_table_5factor(:,7),30)
title('unconditional 5 million threshold')
subplot(2,3,2)
histogram(t_table_5factor(:,8),30)
title('unconditional 250 million threshold')
subplot(2,3,3)
histogram(t_table_5factor(:,9),30)
title('unconditional 1 billion threshold')
subplot(2,3,4)
histogram(t_table_con(:,7),30)
title('conditional 5 million threshold')
subplot(2,3,5)
histogram(t_table_con(:,8),30)
title('conditional 250 million threshold')
subplot(2,3,6)
histogram(t_table_con(:,9),30)
title('conditional 1 billion threshold')
% net_5factor
subplot(2,3,1)
histogram(t_table_5factor(:,10),30)
title('unconditional 5 million threshold')
```

```
subplot(2,3,2)

histogram(t_table_5factor(:,11),30)

title('unconditional 250 million threshold')

subplot(2,3,3)

histogram(t_table_5factor(:,12),30)

title('unconditional 1 billion threshold')

subplot(2,3,4)

histogram(t_table_con(:,10),30)

title('conditional 5 million threshold')

subplot(2,3,5)

histogram(t_table_con(:,11),30)

title('conditional 250 million threshold')

subplot(2,3,6)

histogram(t_table_con(:,12),30)

title('conditional 1 billion threshold')
```