Claremont McKenna College

BETTING AGAINST ALL BETAS: DO FAMA-FRENCH-CARHART FACTORS SHARE THE BETA ANOMALY?

Submitted to

Professor Benjamin Gillen

By

Kevin Jiang

For

Senior Thesis in Economics and Data Science

Fall 2024

December 3, 2024

Acknowledgments

First and foremost, I would like to thank Professor Benjamin Gillen for his continued support throughout my time at Claremont McKenna College. He has provided invaluable insights into the world of quantitative finance through my summer research experience and this thesis, a world I hope to continue to be involved in throughout my career. I am also incredibly grateful for my friends, family, and other professors who all have contributed to my success and development each in their own way.

Abstract

Frazzini and Pedersen's (2013) Betting Against Beta is an investment strategy that exploits a well-documented anomaly in the Capital Asset Pricing Model. This anomaly is called the beta anomaly, which states that the model overestimates the risk-adjusted returns of high-beta assets and underestimates the risk-adjusted returns of low-beta assets. According to the beta anomaly, betas and alphas should be negatively correlated. Frazzini and Pedersen prove that an investor is able to generate positive abnormal returns by holding a long position in low-beta assets and a short position in high-beta assets. They also show that their Betting Against Beta factor delivers statistically significant returns when adjusted for multiple factors. My research seeks to test the significance of the beta anomaly in the market factor when adjusted for factors used in the Fama-French 5-Factor and Carhart 4-Factor models that control for the known size, value, momentum, profitability, and investment factors. I find that the beta anomaly loses significance in the market factor when adjusted for the Fama-French 5-Factor model, proving that Frazzini and Pedersen's results are not robust. I also test for a factor beta anomaly across the size, value, momentum, profitability, and investment factors, as well as the significance of style beta anomalies when adjusted for additional factors. I find that the HML, RMW, and CMA factors show statistically strong patterns that can be made into effective investment strategies.

Table of Contents

1. Introduction

In the vast world of modern investing, all investors seek alpha. Capturing positive alpha represents gaining excess returns above the market or a benchmark. There are many ways in which investors attempt to capture alpha, and one of these ways is factor investing. Factor investing has been a prevalent investing strategy dating back to the creation of the Capital Asset Pricing Model (CAPM) in the early 1960s. This model uses the overall market performance, which is called the market factor (Mkt-RF), to estimate the returns of assets. The model is widely used across many areas of finance. After the CAPM model was developed, additional factors were created to better explain asset returns. Some popular factors include the size (SMB), value (HML), momentum (MOM), profitability (RMW), and investment (CMA) factors. These factors are used in prevalent factor models, which include the Fama-French 3-Factor model (1993) that uses Mkt-RF, SMB, and HML, the Carhart 4-Factor model (1997) that adds MOM to the Fama-French 3-Factor model, and the Fama-French 5-Factor model that adds RMW and CMA to their 3-Factor model. In addition to estimating the returns of an asset, these models are also used to show if the risk-adjusted average returns are statistically significant and how returns are correlated with each of the factors in the model. However, following the creation of the CAPM model, researchers found that it has a pricing anomaly. They found that the model overestimates the risk-adjusted returns of high-beta assets and underestimates the risk-adjusted returns of low-beta assets, a phenomenon commonly referred to as the beta anomaly. This anomaly can be characterized by a negative correlation between betas and alphas. More recently, Frazzini and Pedersen (2013) of hedge fund AQR popularized a strategy to exploit this anomaly. They called this strategy “Betting Against Beta”, in which they invested in low-beta assets and shorted high-beta assets. They found that the beta anomaly is significant when adjusted for the market factor, as well as in combination with other factors. At the time that their paper was written, the Fama-French 5-Factor model was not yet developed, so they could only test the significance of their returns using limited factors. After adjusting for those factors, they found that their Betting Against Beta (BAB) portfolio delivered statistically significant positive alpha. Since Frazzini and Pedersen's paper, many other authors have created and researched similar strategies to exploit the beta anomaly.

However, the beta anomaly has only been researched for the market factor, as it is the only factor used in the CAPM model. In my thesis, I test for the existence of the beta anomaly across Mkt-RF, SMB, HML, MOM, RMW, and CMA by grouping the returns of stocks by their factor betas, then adjusting those returns on a CAPM, 3-Factor, 4-Factor, and 5-Factor level to evaluate the significance of their alphas. One focus of my analysis is whether the beta anomaly is significant for the Mkt-RF factor after adjusting for the more recently developed Fama-French 5-Factor model. The other focus is to test whether the beta anomaly persists for the remaining 5 factors from my selection, along with their significance at different levels. Proving a strong persistence of the beta anomaly in other factors indicates possible effective investing strategies that deliver significant alpha for investors. I find that although the beta anomaly is significant for the Mkt-RF factor when adjusted for the CAPM model, it loses significance when adjusting for Fama-French 5-Factor model. Furthermore, upon investigating the beta anomaly across the other popular factors, I find 3 new significant patterns. First, the HML factor shows a strong positive correlation of 0.90 between its betas and CAPM-adjusted alphas. This is the opposite of the beta anomaly, so any investment strategy using this pattern would invest in high HML beta assets and short low HML beta assets. Using a simple strategy that shorts the lowest beta decile and invests in the highest beta decile, this portfolio results in a CAPM alpha with a t-stat of 2.57. Second, the RMW factor displays the beta anomaly pattern at the 5-Factor level. This pattern has a correlation coefficient of -0.91, and a betting against beta portfolio with a t-stat of 2.46 for its alpha. Lastly, the CMA factor also displays the beta anomaly pattern at the 5-Factor level. This pattern has a lower correlation coefficient of -0.73 but has a significant t-stat of 2.89 for its alpha. I also find intriguing results for the other factors, but these 3 have the most potential to be made into an effective investment strategy.

The remainder of my thesis is as follows. Section 2 discusses the relevant literature surrounding Frazzini and Pedersen's Betting Against Beta strategy and the factor models it builds upon. Section 3 describes my data and how I format it for my analysis. Section 4 explains the methodology for my study, which is broken down into 5 parts. Section 5 displays and outlines my key results. Section 6 discusses the significance of my results in the context of my goal to create an effective investment strategy that captures alpha. Section 7 summarizes my thesis. Following my conclusion, I include my references and an appendix including all less significant results and the code used for my analysis.

2. Literature Review

My research expands on Frazzini and Pedersen's (2013) Betting Against Beta strategy. Their strategy exploits an anomaly in the Capital Asset Pricing Model, which has been empirically proven to overestimate the risk-adjusted returns of high-beta assets and underestimate the risk-adjusted returns of low-beta assets. This anomaly is widely referred to as the beta anomaly. The paper became widely popular because of its robust returns, and generated discourse among the quantitative finance community. Outside of Frazzini and Pedersen, other authors have provided empirical support for the strategy, all attributing different causes for its efficacy. However, some authors argue that the strategy is flawed and overestimates its returns. None of the existing research investigates whether the anomaly persists in the SMB, HML, MOM, RMW, and CMA factors, which I investigate.

Research surrounding the beta anomaly can be dated back to Black, et.al. (1962), who found that the security market line was too flat relative to CAPM. This meant that excess market returns (Mkt-RF) gave an inaccurate estimate of asset returns, with high-beta stocks delivering negative abnormal returns and low-beta stocks delivering positive abnormal returns. In their paper, Frazzini and Pedersen provide further evidence to support this and develop a strategy to exploit it, by constructing a BAB (Betting Against Beta) factor that takes a long position in low-beta assets and a short position in high-beta assets. This factor is synthesized to be market neutral, with the long positions leveraged to a beta of 1 and the short positions de-leveraged to a beta of 1. Using the CRSP tape, Xpressfeed Global database, Barclays Captial Bond.Hub database, and AQR Capital Management's internal pricing data, they find that this factor realizes "statistically significant risk-adjusted returns, accounting for its realized exposure to market, value, size, momentum, and liquidity factors… and realizes a significant positive return in each of the four 20-year subperiods between 1926 and 2012". The U.S. BAB factor realizes a Sharpe ratio of 0.78 between 1926 and 2012, which is about twice the value effect and 40% higher than that of momentum over the same period. Although they prove that the BAB factor is statistically significant when adjusted for the market, size, value, momentum, and liquidity factors, I find that it loses significance when profitability and investment factors replace momentum and liquidity. That is, while it is significant when using a 5-Factor model with Carhart's (1994) Momentum and Pastor and Stambaugh's (2003) liquidity factors, it is not significant when using the Fama-French 5-Factor model. This shows that their BAB factor is not as robust as they suggest and that there needs to be further adjustments to account for other factors in their strategy.

Other sources add further support for Frazzini and Pedersen's findings, such as Berrada et.al. (2014) creating a similar strategy that bets against beta which displays significant performance across equity country indices, equity sectors, and individual stocks. They also use the CRSP database with a sample period of 07/1925 to 12/2011 and re-estimate their betas monthly using a 60-month rolling window. After constructing a long-short portfolio that is alpha optimal, they find that its alphas range from 0.5% to 2.5% per year across each 10-year sub-sample period. Furthermore, Frazzini and Pedersen, alongside Asness (2013), published a corresponding paper providing further proof that their BAB strategy works independently of industry bets. They proved that their BAB strategy does not rely on heavily weighting certain well-performing industries, as it still generated positive returns when picking stocks within individual industries. In fact, this industry-neutral BAB strategy performed better than the regular BAB strategy, delivering positive returns in each of the 49 U.S. industries and in 61 of 70 global industries.

Frazzini and Pedersen attribute the anomaly to the funding constraints of investors, claiming that investors who cannot take on leverage overweight risky securities in order to generate the highest expected excess return power unit of risk. During this process, investors will overweight high-beta assets, driving the prices of those assets up and leading to negative BAB returns. However, funding constraints will eventually loosen, compressing betas toward one and making the BAB factor realize positive returns. Using the TED spread as a measure of funding conditions, their model shows that a higher TED spread predicts low contemporaneous BAB returns, but higher expected returns. This proves their theory, as a higher TED spread is associated with tighter funding constraints. They also find empirical evidence that more constrained investors hold riskier assets, while less constrained investors use leverage to buy low-beta stocks.

In a similar line of reasoning, Hedegaard (2018) also uses investor constraints to explain BAB returns. He cites market performance as a cause for investor constraints, claiming that when past markets have been performing well, there is an outward shift in the investors' demand function. This causes constrained investors to overweight high-beta assets, leading to higher expected BAB returns. To support this, he first regresses BAB returns on market returns to determine if there is a statistically significant relationship between the two. He finds that when regressing BAB on contemporaneous market returns, there is a daily beta of -0.3 with a t-statistic of -40. On the flip side, it also predicts future returns of BAB, which has a significantly positive t-statistic over 2 for up to 10 months. This means that as the market performs well, the current BAB factor generates negative returns, but its expected return increases. He also supports this by sorting BAB returns based on 1-month, 6-month, and 12-month lagged market returns, and in all these cases, low past returns correspond with low BAB returns, while high past returns correspond with high BAB returns. This behavior is prevalent in both U.S. and International equity markets.

Another explanation for the anomaly comes from Berrada et.al.(2014), who claim that the diversity of financial markets guarantees that the betas of stock eventually converge. Similar to how market capitalization can never be concentrated in a single stock and large companies eventually grow at a lower rate, they claim that this mechanism applies to stock betas as well. By creating a model of the financial market with implied diversity, they claim to derive the source of beta arbitrage.

Bali et.al. (2017) take a different approach, claiming that the investors' demand for "lottery-like" stocks causes high-beta assets to be overweight. They state that investors have a disproportionately higher demand for stocks with "high probabilities of large short-term up moves in the stock price", as they can make a larger profit in a shorter period of time. These stocks that are highly sensitive to the market are high-beta stocks, by definition. They prove this by defining a term called MAX, which is the average of the five highest daily returns of the given stock in the given month. Constraining MAX to be neutral, the abnormal returns of a BAB portfolio are no longer significant. Using Fama and MacBeth regressions “indicate a significantly positive relation between beta and stock returns when MAX is included in the regression specification”.

Other authors point to the CAPM model itself as the cause of the anomaly, such as Buchner and Wagner (2015). They suggest that the CAPM model is flawed because it does not take non-linearities in stock returns into account. The Black-Scholes-Merton model of corporate debt and equity valuation states that the equity of a levered firm is equivalent to the call option written on the underlying value of the firm’s asset. Option returns are highly skewed and non-linear related to the company’s stock return, so the model correctly prices the equity of levered firms. By using this model to derive expressions for the pricing error made by CAPM, they find that the pricing error is negative and becomes economically large as the firm's leverage increases. This explains why there is a negative correlation between equity betas and their returns, which becomes significant as betas move further away from 1. Similarly, Ehsani and Linnainmaa (2021) create a correction factor, which explains all distortions in the CAPM security market line. This correction factor has a high explanatory power to explain returns when used with models with more factors. For example, when used with the Fama-French 5-Factor model, the correction factor “increases the model’s out of sample ratio by 40%, increases its power to explain the cross section of returns, and lowers its errors in pricing 208 anomalies”.

However, Novy-Marx and Velikov (2021) refute the BAB strategy, claiming that Frazzini and Pedersen exaggerate the returns of their strategy by not accounting for transaction costs, as well as using unconventional calculations to derive their BAB portfolio and stock betas. They argue that the BAB strategy implemented by Frazzini and Pedersen holds large positions in micro- and nano-cap stocks, which are illiquid and have high transaction costs. They find that “accounting for these transaction costs reduces BAB profitability by almost 60%”. Furthermore, they point out that Frazzini and Pedersen’s BAB portfolio construction procedure and beta estimation calculations are not standard. However, they find that these practices do not significantly alter the results. My analysis follows a more traditional beta calculation and portfolio construction procedure.

All existing research revolves around the low-beta anomaly and Frazzini and Pedersen’s Betting Against Beta strategy. These articles discuss the shortcomings of the CAPM model and only focus on the primary market beta. However, there are other relevant factors that also explain the returns of an asset. I seek to investigate whether the beta anomaly persists in these other factors. My analysis does not provide an explanation for the anomaly, only to discover whether it exists. A strong persistence of the anomaly indicates the potential for a new effective investing strategy.

3. Data

In order to discover whether the beta anomaly persists in the other factor betas, I gathered 2 pieces of data: stock returns and factor returns. Following Frazzini and Pedersen, Berrada et.al, Hedegaard, and Bali et.al., I get my stock return data from CRSP and my factor return data from Kenneth R. French’s Data Library.

The stock return data comes from the CRSP (Center for Research in Security Prices) database at WRDS (Wharton Research Data Services). The CRSP database contains the data of all securities traded on the NYSE (New York Stock Exchange), American Stock Exchange (AMEX), and Nasdaq Stock Market (Nasdaq). Within CRSP, I use the Monthly Stock file, which has monthly data for each stock being queried. I choose a monthly frequency because it will provide ample information without being too large of a dataset. The CRSP Monthly Stock file has data ranging from 12/1925 to 12/2023, but I use data ranging from 01/1973 to 12/2023 for more accurate returns and to ensure that there is corresponding factor data to run my regressions on. All CRSP datasets include a company identifier (PERMNO) and date (date) columns, and I choose to add price data (PRC), returns (RET), and number of shares outstanding (SHROUT) for my analysis. This query results in a csv file containing 4,361,137 entries, each with the corresponding 5 columns. In addition to these columns, I add a market capitalization column to make my analysis easier. This is done by multiplying the Price or Bid/Ask Average (prc) with the Number of Shares Outstanding (shrout) for all stocks.

Table 3.1: First 13 Rows of CRSP Data After Adding Market Capitalization

PERMNO	Date	PRC	RET	SHROUT	MktCap
10000	12/31/1985				
10000	1/31/1986	-4.375	C	3680	-16100
10000	2/28/1986	-3.25	-0.25714	3680	-11960
10000	3/31/1986	-4.4375	0.365385	3680	-16330
10000	4/30/1986	-4	-0.09859	3793	-15172
10000	5/31/1986	3.10938	-0.22266	3793	-11793.9
10000	6/30/1986	3.09375	-0.00503	3793	-11734.6
10000	7/31/1986	2.84375	-0.08081	3793	-10786.3
10000	8/31/1986	1.09375	-0.61539	3793	-4148.59
10000	9/30/1986	1.03125	-0.05714	3793	-3911.53
10000	10/31/1986	0.78125	-0.24242	3843	-3002.34
10000	11/30/1986	0.82813	0.06	3843	-3182.5
10000	12/31/1986	0.51563	-0.37736	3843	-1981.57

To make this dataset more manageable for my regressions, I pivot the dataset so that the columns are all PERMNOs in the dataset, the rows are the date range from 01/1973 to 12/2023, and the values are the monthly returns for each given PERMNO and date. This way I can access the returns by their date and PERMNO, which will help with regressions. Many of the entries are nan since the dataset contains all PERMNOs that exist during this window, and many stocks do not have returns for every month throughout this entire period.

Table 3.2: Sample Returns Data After Pivoting

Date	10006	10014
1/31/1973	-6.04%	-3.57%
2/28/1973	-7.60%	-18.52%
3/31/1973	11.04%	4.55%
4/30/1973	-0.28%	-17.39%
5/31/1973	-1.16%	-10.53%
6/30/1973	3.41%	5.88%
7/31/1973	5.77%	0.00%
8/31/1973	-9.40%	-11.11%
9/30/1973	8.72%	0.00%
10/31/1973	0.53%	12.50%
11/30/1973	-14.68%	-16.67%
12/31/1973	46.52%	0.00%

The factor return data comes from Kenneth R. French's Data Library, which provides information on 3-Factor returns, the Momentum factor returns, and 5-Factor returns. This website is updated regularly, and I use the monthly version of the Fama/French 3 Factors, Momentum Factor (Mom), and Fama/French 5 Factors (2x3) datasets. The Fama/French 3 Factors dataset ranges from 07/1925 to 08/2024, the Momentum Factor (MOM) dataset ranges from 01/1927 to 08/2024, and the Fama/French 5 Factors (2x3) dataset ranges from 07/1963 to 08/2024. For my analysis, I merge these datasets by date to get a resulting dataset with 6 columns: Mkt-RF, SMB, HML, MOM, RMW, CMA, and RF. Lastly, I limit the data to match the date range from my CRSP dataset, which is from 01/1973 to 12/2023.

Table 3.3: First 13 Rows of Factor Data After Merging

Date	Mkt-RF	SMB	HML	MOM	RMW	CMA	RF
1/31/1973	-3.29%	-3.49%	2.68%	3.73%	0.42%	0.90%	0.44%
2/28/1973	-4.85%	-3.87%	1.60%	2.16%	-0.26%	0.02%	0.41%
3/31/1973	-1.30%	-2.82%	2.62%	3.59%	-1.07%	0.62%	0.46%
4/30/1973	-5.68%	-3.85%	5.41%	6.36%	-1.58%	2.60%	0.52%
5/31/1973	-2.94%	-6.30%	0.41%	7.14%	1.95%	-1.57%	0.51%
6/30/1973	-1.57%	-2.86%	1.20%	4.30%	-0.21%	0.11%	0.51%
7/31/1973	5.05%	7.97%	-5.31%	-11.57%	-0.05%	-3.28%	0.64%
8/31/1973	-3.82%	-2.13%	1.24%	3.46%	-1.31%	1.30%	0.70%
9/30/1973	4.75%	3.04%	2.01%	-7.00%	-2.33%	1.77%	0.68%
10/31/1973	-0.83%	-0.45%	1.94%	6.87%	-1.90%	2.71%	0.65%
11/30/1973	-12.75%	-7.67%	3.87%	8.66%	-2.63%	1.73%	0.56%
12/31/1973	0.61%	-5.35%	3.85%	10.38%	-2.78%	2.48%	0.64%

I also create an excess returns data frame equal to the stock returns adjusted for the risk-free rate. I do this by subtracting the monthly RF rate from the factor returns data for each entry of the returns data frame.

To clarify, the Fama/French 3 Factors dataset contains information about 3 factors: Mkt-RF, SMB, and HML. The Mkt-RF returns are calculated by subtracting the risk-free rate from the return of the market. The risk-free rate is defined by the one-month Treasury bill rate, taken from Ibbotson Associates. The market return is defined by the value-weight return of all CRSP firms incorporated in the US and listed on the NYSE, AMEX, or NASDAQ that have a CRSP share code of 10 or 11 at the beginning of month t and good shares and price data at the beginning of t.

The SMB (Small Minus Big) factor is the average return of 3 small stock portfolios minus the average return of 3 large stock portfolios, where the size is measured by a firm's Market Equity. The 3 portfolios in each size category are separated into high value, neutral value, and low value (growth), where value is measured by a firm's Book Equity to Market Equity ratio. Thus, the formula for calculating SMB is

$$SMB = \frac{1}{3}\ (Small\ Value + Small\ Neutral + Small\ Growth)$$

$$-\ \frac{1}{3}\ (Big\ Value + Big\ Neutral + Big\ Growth)$$

The HML (High Minus Low) factor is the average return of 2 high-value portfolios minus the average returns of 2 low-value (growth) portfolios, where value is once again determined by the firm's Book Equity to Market Equity ratio. The portfolios in each value category are separated into big and small, where size is measured by the firm's Market Equity. Thus, the formula for calculating HML is

$$HML = \frac{1}{2}\ (Small\ Value + Big\ Value)$$

$$-\ \frac{1}{2}\ (Small\ Growth + Big\ Growth)$$

Carhart's 4-Factor model uses the same factors as the Fama-French 3-Factor model, as well as a Monthly Momentum Factor to explain an asset's returns. The MOM (Monthly Momentum Factor), is the average return of 2 high past-return portfolios minus the average return of 2 low past-return portfolios, where past returns are measured by the prior 12-month returns of an asset. The past returns are sorted into 3 portfolios – low, neutral, and high – but the neutral portfolio is not used for the calculation. Thus, the high past-return portfolio consists of the top third of asset returns, and the low past-return portfolio consists of the bottom third of asset returns. The 2 portfolios in each past return category are separated into big and small, where size is measured by the firm's Market Equity. Thus, the formula for calculating MOM is:

$$MOM= \frac{1}{2}\ (Small\ High + Big\ High)$$

$$-\ \frac{1}{2}\ (Small\ Low + Big\ Low)$$

Fama/French's 5-Factor model takes their 3-Factor model, slightly modifies the definition of SMB, and adds two extra factors: RMW and CMA. For my regressions, I choose to only use the SMB calculated from their 3-Factor model instead of the new SMB from the 5-Factor model.

The RMW (Robust Minus Weak) factor is the average return of 2 robust operating profitability portfolios minus the average return of 2 weak operating profitability portfolios, where operating profitability is defined by the firm's annual revenue minus its cost of goods sold, selling, general, and administrative expenses, and interest expense. The 2 portfolios in each operating profitability category are separated into big and small, where size is measured by the firm's Market Equity. Thus, the formula for calculating RMW is:

RMW= ½ (Small Robust + Big Robust)

- ½ (Small Weak+ Big Weak)

The CMA (Conservative Minus Aggressive) factor is the average return of 2 conservative investment portfolios minus the average of 2 aggressive investment portfolios, where investments are measured by a firm's growth of total assets over the course of the past year. The 2 portfolios in each investment category are separated into big and small, where size is measured by the firm's Market Equity. Thus, the formula for calculating CMA is:

CMA= ½ (Small Conservative + Big Conservative)

- ½ (Small Aggressive+ Big Aggressive)

The Fama-French 3-Factor model uses Mkt-RF, SMB, and HML according to the following formula:

$$r^* = \alpha + \beta_{MKT-RF}(Mkt - RF) + \beta_{SMB}(SMB) + \beta_{HML}(HML) + \varepsilon$$

Carhart's (1997) 4-Factor model adds MOM to the Fama-French 3-Factor model according to the following formula:

$$r^* = \alpha + \beta_{MKT-RF}(Mkt - RF) + \beta_{SMB}(SMB) + \beta_{HML}(HML) + \beta_{MOM}(MOM) + \varepsilon$$

Lastly, Fama-French's 5-Factor model adds RMW and CMA to their 3-Factor model according to the following formula:

$$r^* = \alpha + \beta_{MKT-RF}(Mkt - RF) + \beta_{SMB}(SMB) + \beta_{HML}(HML) + \beta_{RMW}(RMW) + \beta_{CMA}(CMA) + \varepsilon$$

With this data, I run regressions to analyze how each of these factors explains stock returns in a process that I outline in the next section.

4. Methodology

According to the beta anomaly, high-beta stocks are associated with low alpha, and low-beta stocks are associated with high alpha. In my implementation, I create a self-financing, beta-neutral portfolio that has a long position in the bottom decile of betas and a short position in the top decile of betas. In theory, this portfolio should be factor-neutral and have a positive alpha. To achieve this with code, I break down the process into 5 parts. Choose a factor to bet against and calculate the monthly beta deciles and average excess returns of those deciles (4.1). Regress each decile's return against the Mkt-RF factor, 3-Factor, 4-Factor, and 5-Factor returns to get estimated betas for each factor (4.2). Calculate and visualize the Mkt-RF, 3-Factor, 4-Factor, and 5-Factor-neutral returns equal to alphas for each decile (4.3). Create a zero-cost self-financing factor-neutral betting against beta portfolio (4.4). Evaluate performance metrics for each factor-neutral decile portfolio and the betting against beta portfolio (4.5). Each of these 5 steps will be repeated for the 6 factors I am analyzing.

4.1 Calculating Monthly Beta Deciles and Risk-Adjusted Returns

To start, I define my regression window as 5 years, with the start being the first date in the dataset (01/31/1973), and the end being 5 years after (12/31/1977). This window will be shifted monthly in each iteration of my code. In the first iteration, I define a new variable to track when the end of a year is reached. This is used to calculate the top 500 stocks by market capitalization of each year, which will be recalculated at the end of every year. I do this to limit my analysis to highly liquid stocks that are easily tradable, in response to Novy-Marx and Velikov's (2021) criticisms of Frazzini and Pedersen's (2013) strategy. Once the top 500 stocks are selected, I use an OLS model to regress the 5-year excess return of each stock against my selected factor with the following formula to calculate the beta and alpha for that stock. In this process, I make sure that the stock has at least 50% of its return data from the regression period and the most recent year's worth of data filled before running the regression. Each regression has the following formula.

$$R_s^* = \alpha_s + \beta_s * R_{factor} + \varepsilon$$

for s in top 500 stocks

R_s^* is the stock's excess return on the risk-free rate, α_s is the intercept representing the value of R_s^* when R_{factor} is 0, β_s is the coefficient representing the sensitivity of R_s^* to R_{factor}, R_{factor} is the factor returns, and ε is the error term from the regression. Unlike Frazzini and Pedersen's beta estimation, my approach is more standard and uses a simple regression. Theoretically, the alphas should be negatively correlated with betas. My goal is to construct a factor-neutral portfolio that captures the discrepancies in alphas between the high-and low-beta stocks. After the regression is run for each stock, I store the stock's beta and next month's excess return in an array. The next month's excess return simulates the 1-month excess return of holding that stock after running a 5-year regression. After all regressions are run for that month, I use the stored beta information to calculate the thresholds for beta deciles. I then iterate through the array and index the returns to their corresponding decile bucket using the calculated thresholds. At the end of this process, the next month's excess returns for all stocks are sorted into 10 beta deciles. I then average the returns for each of the deciles, which represents the equal-weight average return for the stocks in each beta decile, and store the 1x10 array as a row into a larger array. This process is repeated monthly until I reach the end of my data, which signals a completed data frame for that factor and results in a 552x10 data frame. Because these are excess returns adjusted for the risk-free rate, each decile's portfolio becomes self-financing.

Table 4.1: First 13 Rows of Sample Decile Excess Returns

	1	2	3	4	5	6	7	8	9	10
1/31/1978	-6.20%	-4.05%	-6.98%	-6.09%	-5.22%	-6.72%	-6.76%	-7.41%	-6.13%	-6.60%
2/28/1978	-2.63%	-1.19%	-1.90%	-1.53%	-1.54%	-1.52%	-1.95%	-2.61%	-2.07%	-2.63%
3/31/1978	3.46%	4.31%	3.76%	1.96%	3.78%	3.54%	2.65%	4.90%	3.66%	4.43%
4/30/1978	6.29%	1.86%	3.92%	5.22%	6.22%	9.02%	9.60%	9.67%	10.68%	10.47%
5/31/1978	1.36%	2.25%	1.30%	1.50%	3.27%	1.56%	1.20%	1.44%	4.36%	4.89%
6/30/1978	-1.34%	-1.44%	-0.93%	-2.08%	-1.36%	-2.23%	-1.19%	-1.94%	-0.27%	-1.61%
7/31/1978	4.14%	3.30%	2.60%	5.08%	5.35%	6.79%	4.85%	5.87%	8.09%	8.01%
8/31/1978	2.32%	1.89%	2.94%	1.48%	5.23%	3.31%	3.40%	5.15%	1.89%	1.80%
9/30/1978	0.96%	-0.30%	-0.80%	0.08%	-0.56%	-2.14%	-1.25%	-1.87%	-1.98%	-3.11%
10/31/1978	-11.31%	-9.36%	-9.67%	-10.44%	-10.01%	-11.94%	-13.23%	-12.02%	-13.38%	-11.60%
11/30/1978	0.68%	1.81%	3.27%	2.44%	1.08%	2.41%	0.74%	2.41%	3.94%	3.05%
12/31/1978	-0.71%	0.48%	-1.10%	0.48%	-1.39%	0.58%	-0.59%	1.76%	0.95%	1.31%
1/31/1979	6.66%	4.41%	6.35%	4.96%	6.34%	5.87%	4.75%	4.49%	2.22%	1.85%

4.2 Regressing Decile Returns on Factor Returns

Using the average returns by beta decile data frame, I regress each decile column returns on the Mkt-RF factor, 3-Factor, 4-Factor, and 5-Factor returns to calculate the estimated beta for each factor in the decile. For each regression, I regress the entire decile returns data (01/31/1983 – 12/31/2023) on the selected factors. I choose the Mkt-RF factor for the 1-Factor regression because it is used in the CAPM model and holds the most explanatory power on decile returns. As I add more factors, the explanatory power of each model should increase, making the returns neutralized to more factors. The regression formulas are as follows.

Mkt-RF:

$$R^*_{decile} = \alpha_{decile} + \beta_{Mkt-RF,decile} * R_{Mkt-RF} + \varepsilon$$

3-Factor:

$$R^*_{decile} = \alpha_{decile} + \beta_{Mkt-RF,decile} * R_{Mkt-RF} + \beta_{SMB,decile} * R_{SMB} + \beta_{HML,decile} * R_{HML} + \varepsilon$$

4-Factor:

$$R^*_{decile} = \alpha_{decile} + \beta_{Mkt-RF,decile} * R_{Mkt-RF} + \beta_{SMB,decile} * R_{SMB} + \beta_{HML,decile} * R_{HML} + \beta_{MOM,decile} * R_{MOM} + \varepsilon$$

5-Factor:

$$R^*_{decile} = \alpha_{decile} + \beta_{Mkt-RF,decile} * R_{Mkt-RF} + \beta_{SMB,decile} * R_{SMB} + \beta_{HML,decile} * R_{HML} + \beta_{RMW,decile} * R_{RMW} + \beta_{CMA,decile} * R_{CMA} + \varepsilon$$

In each regression, R^*_{decile} is the estimated excess return for that decile, α_{decile} is the intercept representing the value of R^*_{decile} when R_{factor} is 0, $\beta_{factor,decile}$ is the coefficient representing the sensitivity of R^*_{decile} to R_{factor}, R_{factor} is the factor returns, and ε is the error term from the regression. For each regression, I also record the corresponding R-squared values to evaluate the power that the factors have in explaining the decile's excess returns.

4.3 Calculating and Visualizing Factor-Neutral Decile Returns

As I run each regression from 4.2, I use the estimated betas to calculate the factor-neutral returns for each decile. These factor-neutral returns are equal to the alphas of each decile's returns and represent the estimated returns when all factors are 0. The factor-neutral returns are calculated with the following formulas.

CAPM:

$$R_{decile}^{FN} = R_{decile}^{*} - \beta_{Mkt-RF,decile} * R_{Mkt-RF}$$

3-Factor:

$$R_{decile}^{FN} = R_{decile}^{*} - \beta_{Mkt-RF,decile} * R_{Mkt-RF} - \beta_{SMB,decile} * R_{SMB} - \beta_{HML,decile} * R_{HML}$$

4-Factor:

$$R_{decile}^{FN} = R_{decile}^{*} - \beta_{Mkt-RF,decile} * R_{Mkt-RF} - \beta_{SMB,decile} * R_{SMB} - \beta_{HML,decile} * R_{HML}$$

$$\beta_{MOM,decile} * R_{MOM}$$

5-Factor:

$$R_{decile}^{FN} = R_{decile}^{*} - \beta_{Mkt-RF,decile} * R_{Mkt-RF} - \beta_{SMB,decile} * R_{SMB} - \beta_{HML,decile} * R_{HML} -$$

$$\beta_{RMW,decile} * R_{RMW} - \beta_{CMA,decile} * R_{CMA}$$

R_{decile}^{FN} is the calculated factor-neutral excess returns for that decile, and it is equal to the alphas from the regressions in 4.2. $R_{decile}^{*}$is the excess returns by decile calculated from 4.1. $\beta_{factor,decile}$ is the estimated factor betas from the previous regression. Because the decile factor-neutral returns capture excess returns, each factor-neutral decile portfolio is self-financing as well. By capturing the alphas for each decile, the factor-neutral returns are insensitive to changes in each factor. Furthermore, as the returns are adjusted for more factors, they become more robust. Lastly, after taking averages for each decile's factor-neutral returns, I can easily visualize the relationship between alphas and betas.

4.4 Creating a Self-Financing Factor-Neutral Betting Against Beta Portfolio

Using the decile factor-neutral returns, I can construct a betting against beta portfolio that is self-financing and factor-neutral. The beta anomaly suggests that the bottom decile has higher alphas than the top decile, which means that the calculated factor-neutral returns for the top decile should be higher than the bottom decile. There are many ways to construct a betting against beta portfolio, but I choose a simple approach that takes a long position in the factor-neutral bottom beta decile portfolio and an equal short position in the factor-neutral top beta decile portfolio. This is far simpler than Frazzini and Pedersen's Betting Against Beta portfolio, which uses betas as weights to include all assets in the universe. To calculate my portfolio, I subtract the factor-neutral returns of the top beta decile portfolio from the factor-neutral returns of the bottom beta decile portfolio. The formula for this is

$$R_{BAB}^{FN} = R_{10}^{FN} - R_{1}^{FN}$$

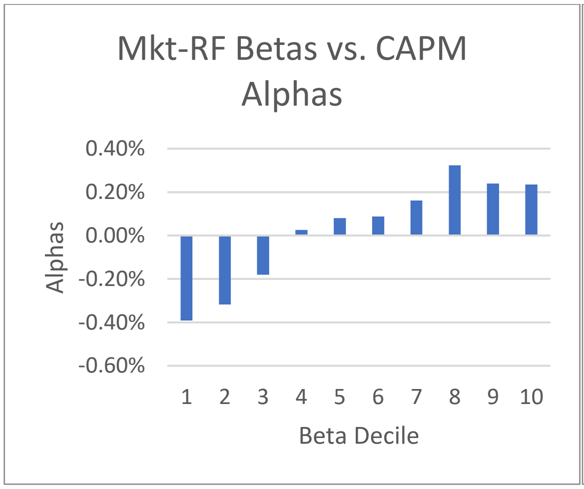
R_{BAB}^{FN} is the returns of the BAB, R_{10}^{FN} is the factor-neural returns of the bottom beta decile, and R_{1}^{FN} is the factor-neural returns of the top beta decile.

4.5 Evaluating Performance

With all the returns computed, I analyze their performance using certain performance metrics. The first part of my analysis involves plotting betas and alphas from 4.3 in a bar plot. According to the beta anomaly, alphas should decrease as betas increase. The second part of my analysis is calculating performance metrics for each of the factor-neutral beta decile portfolios. I choose to calculate the annual arithmetic mean, t-stat of the alpha, annual geometric mean, annual volatility, and Sharpe Ratios for each decile's returns. I also calculate the same performance metrics for my betting against beta portfolio returns. Lastly, to evaluate the persistence of the beta anomaly pattern, I calculate the coefficient of correlation between the beta decile indices and their corresponding alphas.

5. Results

Figures 5.1 – 5.6 show the visual representation of the alphas for each beta decile. Each pair of figures represents betting against a certain factor during beta decile construction, and the figures within the set represent CAPM and 5-Factor alphas. The 3-Factor and 4-Factor alphas are added in the appendix. In my figures, beta decile 1 corresponds to the top decile of betas, and beta decile 10 corresponds to the bottom decile. Visually, a beta anomaly pattern corresponds to alphas increasing as betas decrease. The clearer this pattern is, the stronger the beta anomaly, and the better a betting against beta strategy would work.



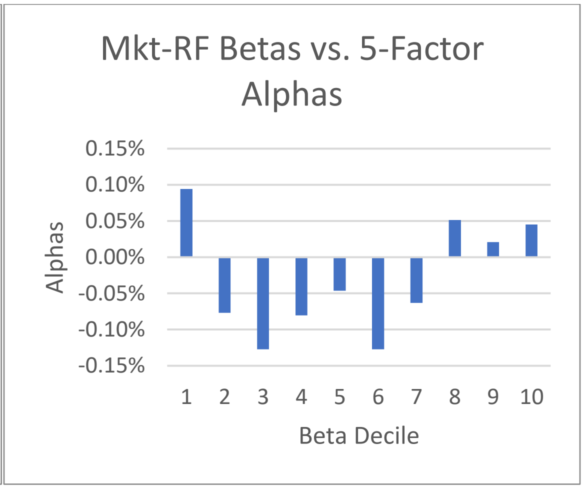


Figure 5.1a and 5.1b: Mkt-RF Beta Deciles vs. CAPM and 5-Factor Alphas

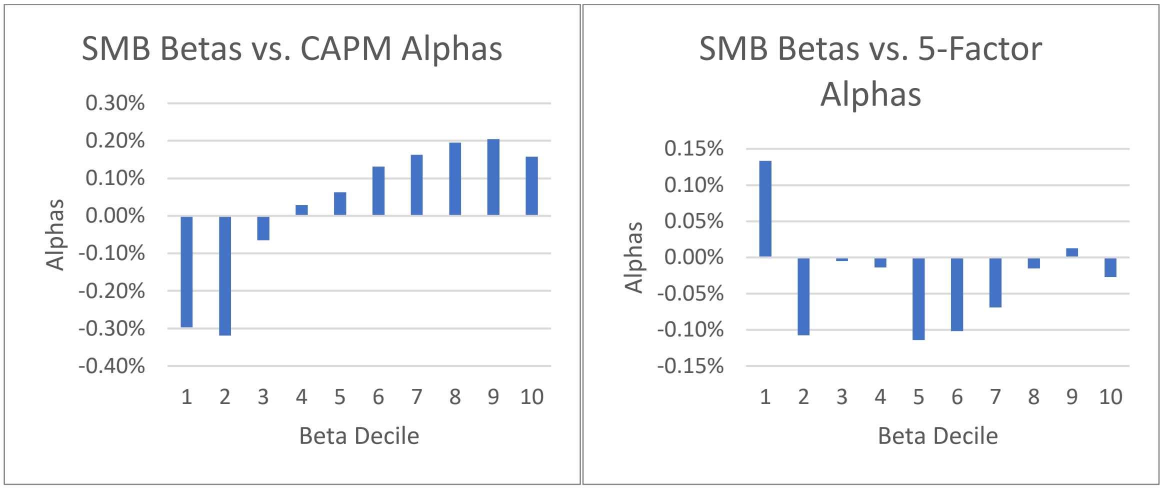


Figure 5.2a and 5.2b: SMB Beta Deciles vs. CAPM and 5-Factor Alphas

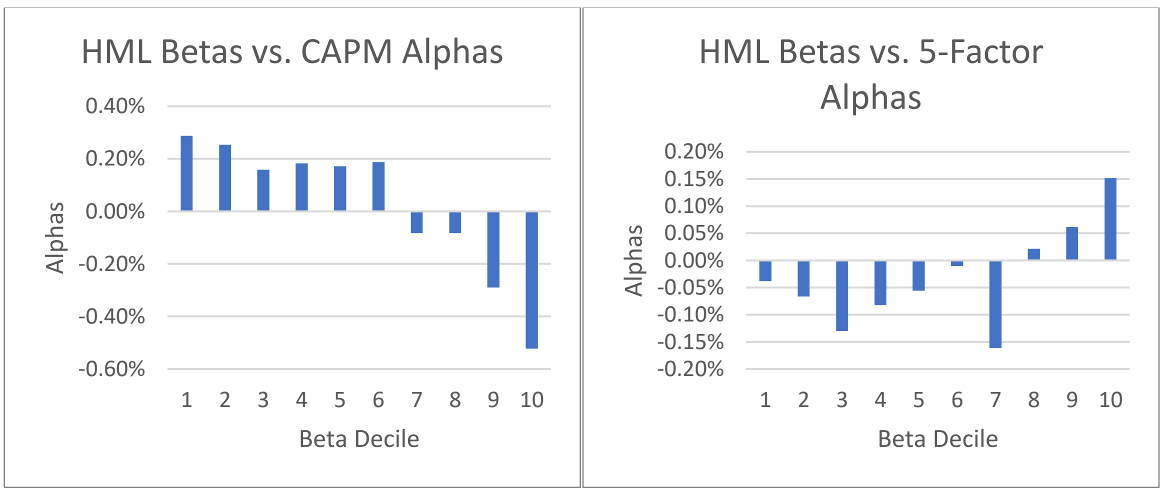


Figure 5.3a and 5.3b: HML Beta Deciles vs. CAPM and 5-Factor Alphas

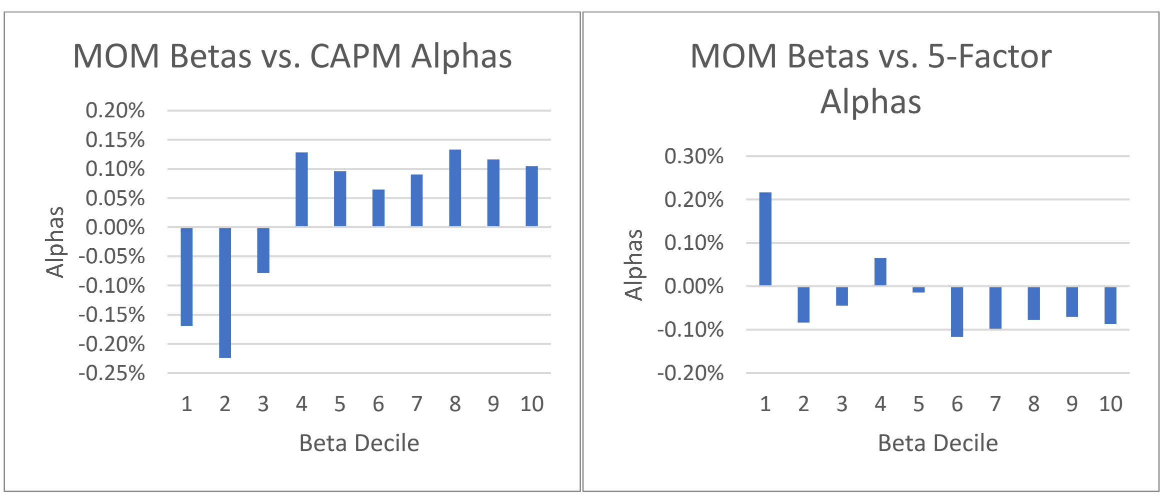


Figure 5.4a and 5.4b: MOM Beta Deciles vs. CAPM and 5-Factor Alphas

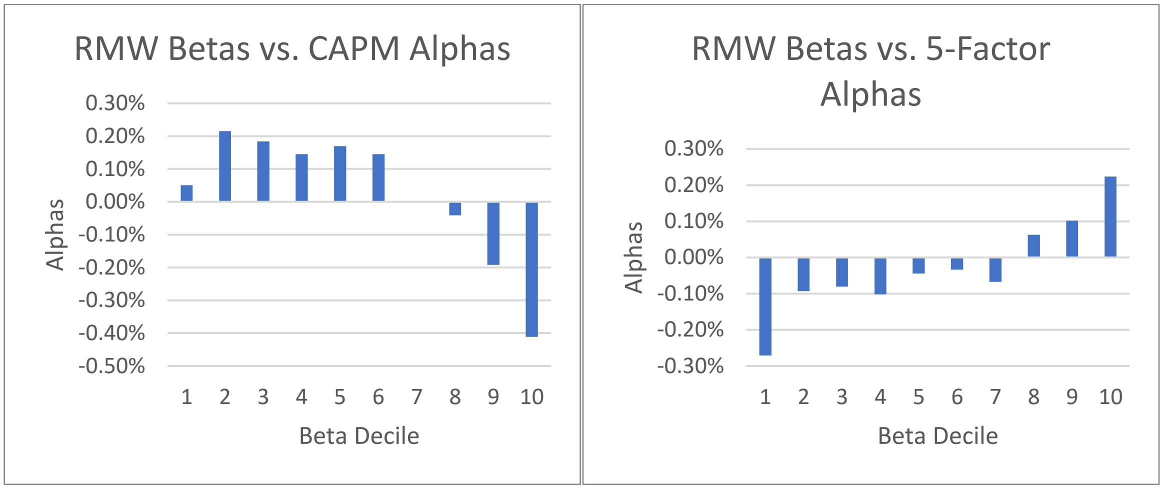


Figure 5.5a and 5.5b: RMW Beta Deciles vs. CAPM and 5-Factor Alphas

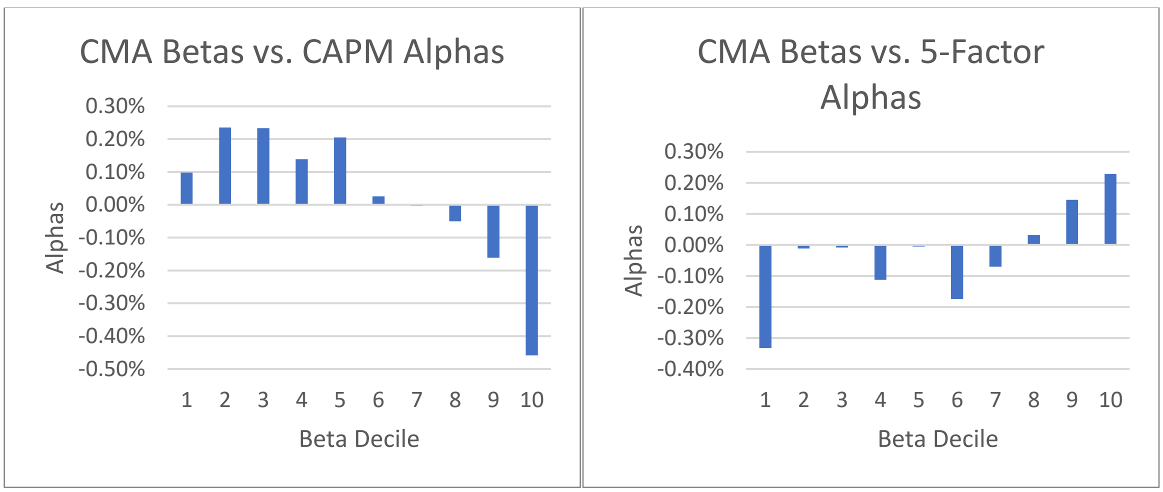


Figure 5.6a and 5.6b: CMA Beta Deciles vs. CAPM and 5-Factor Alphas

Tables 5.7 – 5.11 show the performance metrics of each factor-neutral beta decile returns, as well as the explanatory power of the factors used to neutralize the returns. Each table shows the annual arithmetic mean, t-stat of the alpha, annual geometric mean, annual volatility, Sharpe Ratio, and R-squared for the 10 beta decile portfolios. The alpha for each decile is equal to the annual arithmetic mean. According to the beta anomaly, as betas decrease, alphas should increase. This pattern should appear in the arithmetic mean, geometric mean, and Sharpe Ratios. Furthermore, as the decile returns are adjusted for more factors, the returns are better explained by those factors. This is reflected in the R-squared values, which increase as more factors are added. The R-squared represents the proportion that the factors explain the variance in returns. Once again beta decile 1 corresponds to the highest beta decile, and each decile after that corresponds to lower beta deciles.

Table 5.7a: Mkt-RF Beta Decile CAPM Statistics

Beta Decile	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Arith Mean	-4.69%	-3.82%	-2.16%	0.30%	0.97%	1.06%	1.93%	3.88%	2.87%	2.81%
t-stat	-2.18	-2.95	-2.20	0.34	1.21	1.19	2.16	3.96	2.76	2.11
Geo Mean	-5.58%	-4.12%	-2.36%	0.12%	0.82%	0.88%	1.76%	3.72%	2.66%	2.43%
Volatility	14.62%	8.77%	6.65%	5.99%	5.41%	6.04%	6.08%	6.65%	7.05%	9.05%
Sharpe Ratio	-0.32	-0.44	-0.32	0.05	0.18	0.18	0.32	0.58	0.41	0.31
R-Squared	0.76	0.86	0.90	0.89	0.90	0.86	0.84	0.77	0.66	0.37

Table 5.7b: Mkt-RF Beta Decile 5-Factor-Neutral Statistics

Beta Decile	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Arith Mean	1.13%	-0.92%	-1.53%	-0.97%	-0.56%	-1.53%	-0.76%	0.62%	0.25%	0.54%
t-stat	0.62	-0.78	-1.61	-1.19	-0.76	-2.05	-1.00	0.78	0.29	0.45
Geo Mean	0.38%	-1.23%	-1.72%	-1.11%	-0.68%	-1.65%	-0.89%	0.47%	0.08%	0.21%
Volatility	12.44%	7.98%	6.41%	5.51%	4.96%	5.07%	5.12%	5.40%	5.87%	8.15%
Sharpe Ratio	0.09	-0.12	-0.24	-0.18	-0.11	-0.30	-0.15	0.11	0.04	0.07
R-Squared	0.83	0.88	0.90	0.91	0.91	0.90	0.88	0.85	0.77	0.49

Table 5.8a: SMB Beta Decile CAPM Statistics

Beta Decile	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Arith Mean	-3.56%	-3.83%	-0.78%	0.35%	0.76%	1.58%	1.95%	2.34%	2.45%	1.90%
t-stat	-1.72	-2.92	-0.88	0.42	0.85	1.71	2.12	2.53	2.30	1.42
Geo Mean	-4.42%	-4.14%	-0.95%	0.19%	0.58%	1.39%	1.77%	2.17%	2.21%	1.50%
Volatility	14.04%	8.89%	5.97%	5.72%	6.04%	6.27%	6.25%	6.27%	7.24%	9.07%
Sharpe Ratio	-0.25	-0.43	-0.13	0.06	0.13	0.25	0.31	0.37	0.34	0.21
R-Squared	0.75	0.85	0.90	0.90	0.87	0.85	0.83	0.80	0.69	0.55

Table 5.8b: SMB Beta Decile 5-Factor-Neutral Statistics

Beta Decile	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Arith Mean	1.60%	-1.29%	-0.06%	-0.17%	-1.37%	-1.22%	-0.83%	-0.18%	0.15%	-0.33%
t-stat	0.96	-1.11	-0.07	-0.21	-1.73	-1.61	-1.12	-0.24	0.18	-0.29
Geo Mean	0.98%	-1.59%	-0.21%	-0.32%	-1.50%	-1.34%	-0.95%	-0.31%	-0.02%	-0.60%
Volatility	11.31%	7.91%	5.60%	5.45%	5.36%	5.13%	5.02%	5.08%	5.91%	7.51%
Sharpe Ratio	0.14	-0.16	-0.01	-0.03	-0.26	-0.24	-0.16	-0.04	0.03	-0.04
R-Squared	0.84	0.88	0.91	0.91	0.90	0.90	0.89	0.87	0.80	0.69

Table 5.9a: HML Beta Decile CAPM Statistics

Beta Decile	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Arith Mean	3.45%	3.04%	1.90%	2.19%	2.06%	2.25%	-1.00%	-1.00%	-3.48%	-6.27%
t-stat	1.68	2.21	1.67	2.05	2.21	2.58	-1.17	-1.03	-2.35	-2.71
Geo Mean	2.52%	2.64%	1.61%	1.95%	1.87%	2.10%	-1.16%	-1.21%	-3.90%	-7.22%
Volatility	13.90%	9.32%	7.74%	7.24%	6.31%	5.93%	5.77%	6.56%	10.05%	15.68%
Sharpe Ratio	0.25	0.33	0.25	0.30	0.33	0.38	-0.17	-0.15	-0.35	-0.40
R-Squared	0.53	0.68	0.78	0.79	0.83	0.85	0.88	0.87	0.77	0.67

Table 5.9b: HML Beta Decile 5-Factor-Neutral Statistics

Beta Decile	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Arith Mean	-0.46%	-0.80%	-1.56%	-0.98%	-0.67%	-0.13%	-1.94%	0.26%	0.74%	1.82%
t-stat	-0.32	-0.83	-1.91	-1.13	-0.88	-0.17	-2.35	0.28	0.61	1.10
Geo Mean	-0.93%	-1.01%	-1.70%	-1.15%	-0.80%	-0.26%	-2.07%	0.07%	0.40%	1.21%
Volatility	9.81%	6.58%	5.55%	5.90%	5.12%	5.21%	5.59%	6.22%	8.28%	11.26%
Sharpe Ratio	-0.05	-0.12	-0.28	-0.17	-0.13	-0.02	-0.35	0.04	0.09	0.16
R-Squared	0.77	0.84	0.89	0.86	0.89	0.89	0.88	0.88	0.85	0.83

Table 5.10a: MOM Beta Decile CAPM Statistics

Beta Decile	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Arith Mean	-2.03%	-2.69%	-0.94%	1.54%	1.15%	0.78%	1.09%	1.60%	1.40%	1.26%
t-stat	-1.10	-2.22	-0.98	2.08	1.49	0.85	0.99	1.38	1.06	0.66
Geo Mean	-2.79%	-2.99%	-1.14%	1.42%	1.02%	0.59%	0.81%	1.30%	1.00%	0.42%
Volatility	12.58%	8.22%	6.51%	5.00%	5.26%	6.18%	7.45%	7.84%	8.95%	13.01%
Sharpe Ratio	-0.16	-0.33	-0.14	0.31	0.22	0.13	0.15	0.20	0.16	0.10
R-Squared	0.67	0.79	0.84	0.90	0.89	0.85	0.80	0.80	0.76	0.67

Table 5.10b: MOM Beta Decile 5-Factor-Neutral Statistics

Beta Decile	1	2	3	4	5	6	7	8	9	10
Arith Mean	2.60%	-1.00%	-0.53%	0.78%	-0.17%	-1.40%	-1.17%	-0.93%	-0.84%	-1.04%
t-stat	1.67	-0.90	-0.57	1.08	-0.23	-1.78	-1.25	-0.98	-0.81	-0.67
Geo Mean	2.07%	-1.27%	-0.73%	0.66%	-0.29%	-1.53%	-1.37%	-1.14%	-1.08%	-1.58%
Volatility	10.54%	7.48%	6.35%	4.92%	4.98%	5.33%	6.36%	6.47%	7.06%	10.58%
Sharpe Ratio	0.25	-0.13	-0.08	0.16	-0.03	-0.26	-0.18	-0.14	-0.12	-0.10
R-Squared	0.77	0.82	0.84	0.90	0.90	0.89	0.85	0.86	0.85	0.78

Table 5.11a: RMW Beta Decile CAPM Statistics

Beta Decile	1	2	3	4	5	6	7	8	9	10
Arith Mean	0.61%	2.58%	2.21%	1.75%	2.03%	1.74%	0.00%	-0.49%	-2.30%	-4.94%
t-stat	0.42	2.25	2.18	1.89	2.22	1.91	0.00	-0.49	-1.66	-2.15
Geo Mean	0.12%	2.31%	1.99%	1.56%	1.86%	1.56%	-0.16%	-0.72%	-2.71%	-5.95%
Volatility	9.90%	7.78%	6.89%	6.25%	6.22%	6.18%	5.63%	6.79%	9.42%	15.58%
Sharpe Ratio	0.06	0.33	0.32	0.28	0.33	0.28	0.00	-0.07	-0.24	-0.32
R-Squared	0.65	0.74	0.80	0.83	0.84	0.86	0.89	0.87	0.80	0.66

Table 5.11b: RMW Beta Decile 5-Factor-Neutral Statistics

Beta Decile	1	2	3	4	5	6	7	8	9	10
Arith Mean	-3.25%	-1.12%	-0.97%	-1.22%	-0.54%	-0.41%	-0.81%	0.75%	1.22%	2.68%
t-stat	-2.58	-1.20	-1.16	-1.67	-0.69	-0.54	-1.11	0.82	1.05	1.56
Geo Mean	-3.55%	-1.31%	-1.13%	-1.34%	-0.67%	-0.54%	-0.92%	0.56%	0.91%	2.04%
Volatility	8.53%	6.34%	5.68%	4.96%	5.24%	5.19%	4.94%	6.18%	7.88%	11.62%
Sharpe Ratio	-0.38	-0.18	-0.17	-0.25	-0.10	-0.08	-0.16	0.12	0.15	0.23
R-Squared	0.74	0.82	0.86	0.90	0.89	0.90	0.92	0.89	0.86	0.81

Table 5.12a: CMA Beta Decile CAPM Statistics

Beta Decile	1	2	3	4	5	6	7	8	9	10
Arith Mean	1.17%	2.82%	2.80%	1.67%	2.47%	0.31%	-0.04%	-0.61%	-1.94%	-5.51%
t-stat	0.72	2.48	2.68	1.61	2.73	0.35	-0.05	-0.64	-1.38	-2.40
Geo Mean	0.55%	2.55%	2.58%	1.43%	2.30%	0.13%	-0.19%	-0.81%	-2.36%	-6.48%
Volatility	11.11%	7.73%	7.08%	7.03%	6.14%	6.08%	5.62%	6.44%	9.53%	15.53%
Sharpe Ratio	0.11	0.37	0.40	0.24	0.40	0.05	-0.01	-0.09	-0.20	-0.35
R-Squared	0.57	0.70	0.76	0.79	0.84	0.86	0.90	0.88	0.82	0.70

Table 5.12b: CMA Beta Decile 5-Factor-Neutral Statistics

Beta Decile	1	2	3	4	5	6	7	8	9	10
Arith Mean	-3.99%	-0.13%	-0.10%	-1.35%	-0.06%	-2.09%	-0.84%	0.39%	1.75%	2.75%
t-stat	-3.20	-0.15	-0.13	-1.67	-0.08	-2.75	-1.05	0.42	1.39	1.61
Geo Mean	-4.27%	-0.32%	-0.26%	-1.50%	-0.19%	-2.20%	-0.99%	0.19%	1.40%	2.11%
Volatility	8.45%	6.10%	5.53%	5.50%	5.10%	5.15%	5.45%	6.31%	8.52%	11.58%
Sharpe Ratio	-0.47	-0.02	-0.02	-0.25	-0.01	-0.41	-0.15	0.06	0.20	0.24
R-Squared	0.75	0.81	0.85	0.87	0.89	0.90	0.90	0.89	0.85	0.83

Tables 5.13 – 5.14 show the performance of a betting against beta portfolio for each factor. The betting against beta portfolio returns follows a simple formula of the lowest decile returns minus the highest decile returns. This represents a long position in the low decile returns and an equal short position in the high decile returns. I report the annual arithmetic mean, t-stat, annual geometric mean, annual volatility, Sharpe Ratio, and the correlation coefficient from each table in 5.7-5.11. The closer the correlation coefficient is to -1, the stronger the beta anomaly pattern is. Table 5.13 shows the betting against beta portfolio adjusted for the CAPM model, and Table 5.14 shows the betting against beta portfolio adjusted for the Fama-French 5-Factor model.

Table 5.13: CAPM BAB Performance

Factor	Mkt-RF	SMB	HML	MOM	RMW	CMA
Arithmetic Mean	7.50%	5.45%	-9.72%	3.29%	-5.55%	-6.68%
t-stat	2.53	1.85	-2.57	1.04	-1.68	-1.94
Geometric Mean	5.52%	3.39%	-12.26%	0.99%	-7.74%	-8.96%
Volatility	20.14%	19.97%	25.67%	21.51%	22.41%	23.37%
Sharpe Ratio	0.37	0.27	-0.38	0.15	-0.25	-0.29
Correlation Coefficient	-0.94	-0.90	0.90	-0.80	0.79	0.84

Table 5.14: 5-Factor-Neutral BAB Performance

Factor	Mkt-RF	SMB	HML	MOM	RMW	CMA
Arithmetic Mean	-0.59%	-1.93%	2.28%	-3.64%	5.93%	6.74%
t-stat	-0.24	-0.86	0.94	-1.50	2.46	2.89
Geometric Mean	-2.04%	-3.05%	0.94%	-4.89%	4.71%	5.66%
Volatility	16.86%	15.15%	16.37%	16.46%	16.32%	15.83%
Sharpe Ratio	-0.03	-0.13	0.14	-0.22	0.36	0.43
Correlation Coefficient	-0.26	0.20	-0.63	0.62	-0.91	-0.73

6. Results Discussion

The goal of my research is to test the beta anomaly in the Mkt-RF factor, as well as whether it persists in the SMB, HML, MOM, RMW, and CMA factors. As mentioned before the beta anomaly pattern corresponds to a negative correlation between betas and alphas. The stronger this correlation is and the larger the difference is between the highest and lowest deciles, the better a betting against beta strategy will work. When evaluating the anomaly for each factor, I use a correlation coefficient to determine the strength of the pattern, and the t-stat of the betting against portfolio (BAB) alpha to determine the difference between the highest and lowest deciles. In my discussion, I will refer to factor-neutral average returns as alphas, since they are mathematically equivalent. I will begin my discussion by analyzing patterns across all factors, and then discuss the pattern within each factor.

6.1 Patterns Across All Factors

There are a number of interesting patterns that are consistent across all factors. First, the relationships between betas and alphas significantly change from the CAPM alphas to the 5-Factor alphas. Across all 6 factors, the relationship either loses most of its significance or reverses. This can be seen visually in each pair of figures in 5.1 – 5.7. Quantitatively, the correlation coefficient of each factor from Table 5.13 is reduced or changes sign in Table 5.14. In theory, the relationships should stay similar regardless of the additional factors, especially since all factors outside of Mkt-RF add marginal explanatory power. The second interesting pattern is that for all factors other than Mkt-RF, the volatility for beta deciles towards the tails is higher than for beta deciles in the middle. This results in the betting against beta portfolios becoming highly volatile, as they are the combination of 2 high-volatility portfolios. Lastly, for all factors, the R-squared is higher for beta deciles in the middle than for beta deciles towards the tails. This means that the returns of extreme beta deciles are less predictable by the factors.

6.2 Betting Against Mkt-RF

Current literature involving betting against beta only involves the Mkt-RF factor. The consensus beta anomaly describes the negative correlation between Mkt-RF betas and alphas. I find that this pattern persists for CAPM alphas, but not the 5-Factor alphas. This can be seen visually in Figures 5.1a and 5.1b, where Figure 5.1a shows CAPM alphas increasing as betas decrease, while Figure 5.1b shows 5-Factor alphas largely uncorrelated to betas. Quantifying this relationship, Table 5.13 shows that the correlation coefficient between CAPM alphas and Mkt-RF betas is -0.94, which is the strongest measured across all factors. In Table 5.14, the correlation coefficient for 5-Factor alphas diminishes to -0.26, far lower than the CAPM alphas. This pattern is reflected across tables 5.7a and 5.7b, where table 5.7a shows a consistent increase in arithmetic means, geometric means, and Sharpe Ratios as betas decrease. On the other hand, table 5.7b shows arithmetic means, geometric means, and Sharpe Ratios do not have a structured change as betas decrease. The beta anomaly persisting for Mkt-RF CAPM alphas is in line with the existing research, supporting Frazzini and Pedersen’s (2013) findings. However, the beta anomaly losing most of its significance at a 5-Factor level shows that their results are not robust. In Frazzini and Pedersen's (2013) Table III, it is shown the alphas for beta decile returns begin losing significance as more factors are added. Although the table does not include the Fama-French 5-Factor alphas, this decreasing pattern implies that new outside factors further diminish significance, which is proven by my findings. My result is consistent with Ehsani and Linnainmaa (2021), who wrote their paper after the Fama-French 5-Factor model was published, allowing them to use more factors in their analysis. They also show that the BAB factor loses significance as more factors are added. They claim that when four or more factors are used, their correction factor is more effective at estimating returns than the BAB factor. This shows that there must be additional calculations to be considered for the Mkt-RF BAB strategy to deliver 5-Factor alphas.

6.3 Betting Against SMB

Analyzing the relationship between SMB betas and alphas begins my new research into this topic. I find that similar to Mkt-RF, the beta anomaly pattern for SMB betas is present for CAPM alphas and loses significance for 5-Factor alphas. This visually appears in figures 5.2a and 5.2b, where CAPM alphas generally increase, and 5-Factors alphas have a very weak pattern as betas decrease. This is supported by the correlation coefficients in Tables 5.13 and 5.14. In Table 5.13, the correlation coefficient for CAPM alphas and SMB betas is -0.90, while in Table 5.14, the 5-Factor correlation coefficient is 0.20. This pattern is especially clear for the geometric means across the CAPM beta deciles in Table 5.8a, where the means are strictly increasing from beta deciles 1-9. Table 5.8b shows the geometric means of the 5-Factor-neutral returns have no pattern across beta deciles.

6.4 Betting Against HML

The HML betas and alphas show a new pattern. For CAPM alphas, there is a strong positive correlation between betas and alphas. However, this pattern changes to a moderate negative correlation for 5-Factor alphas. This can be visually noted in figures 5.3a and 5.3b, where CAPM alphas decrease and 5-Factor alphas moderately increase as betas decrease. Table 5.13 shows that CAPM alphas and HML betas have a correlation coefficient of 0.90, which is relatively strong. Furthermore, the t-stat of its BAB alpha being -2.57 means that the difference between its highest and lowest decile is large. This could lead to a successful strategy that has short positions in low HML beta assets and long positions in high HML beta assets. Table 5.14 shows that the 5-Factor alphas and HML betas have a less significant correlation coefficient of -0.63.

6.5 Betting Against MOM

The MOM betas have a positive correlation with CAPM alphas, but a negative correlation with 5-Factor alphas. This is visually displayed in figures 5.4a and 5.4b, where CAPM alphas increase and 5-Factor alphas weakly decrease as betas decrease. Table 5.13 shows that the CAPM alphas and MOM beta deciles indices have a correlation coefficient of -0.80. It is important to note that although this coefficient is relatively high, the difference between the top and bottom deciles is small. This is evidenced by the MOM 5-Factor BAB alpha having a t-stat of 1.04, which is not largely significant. On the other hand, the 5-Factor alphas and MOM betas have a positive correlation with a correlation coefficient of 0.62, as shown in Table 5.14.

6.6 Betting Against RMW

The RMW and CMA factors were developed more recently than the other factors, and therefore have been studied less. In my analysis, I have found that these 2 factors follow similar patterns as well. Similar to the MOM factor, the RMW betas show a negative correlation for CAPM alphas. However, the relationship reverses for the 5-Factor alphas, showing a very strong negative correlation. This is visually represented in Figures 5.5a and 5.5b, where the two figures have opposite patterns. Table 5.13 shows a positive correlation between CAPM alphas and RMW betas, with a correlation coefficient of 0.79. More importantly, there is a strong negative correlation between 5-Factor alphas and RMW betas, with a correlation coefficient of -0.91. This is the second most significant correlation coefficient, just behind the CAPM alphas for Mkt-RF betas, which is the source of the consensus beta anomaly. However, the magnitude of the top and bottom decile difference is slightly smaller, with the BAB alpha having a t-stat of 2.46, as compared to 2.53 for the Mkt-RF CAPM alphas. Nonetheless, this is still a significant finding and proves a clear beta anomaly pattern at the 5-Factor level.

6.7 Betting against CMA

As mentioned before, CMA and RMW have similar patterns. The CMA betas have a positive correlation with CAPM alphas, but similar to RMW, the pattern flips and leads to a negative correlation with 5-Factor alphas. This is visually represented in Figures 5.6a and 5.6b, where the 2 figures have opposite patterns, although each pattern is not as strong as RMW's. Quantitatively, the correlation coefficient for CAPM alphas and CMA betas is 0.84, while the correlation coefficient for 5-Factor alphas and CMA betas is -0.73. Although the 5-Factor alphas do not have the strongest correlation coefficient, the magnitude of the difference between the top and bottom deciles is the largest. This results in the CMA 5-Factor BAB portfolio having an alpha with a t-stat of 2.89, the highest out of all BAB portfolios. This is a promising result and may lead to more successful BAB strategies in the future.

7. Conclusion

Betting Against Beta is a strategy introduced by Frazzini and Pedersen (2013) that exploits a mispricing anomaly in the CAPM model. This anomaly, commonly referred to as the beta anomaly, states that the CAPM model has been found to overestimate the alphas of high-beta assets and underestimate the alphas of low-beta assets. This means that as betas decrease, alphas should increase. Furthermore, Frazzini and Pedersen find that this pattern is significant when adjusting for certain factors. The beta anomaly is well documented by many authors and has been used to support many betting against beta strategies. There are many proposed explanations for the beta anomaly, but most point to two general reasons. One follows that certain investor behaviors lead high-beta assets to be overweight, which drives their price up but future returns down. The same pressure occurs for low beta assets, leading to higher than expected future returns. The other explanation suggests that there is an inherent mistake within the CAPM model and that there are certain calculations that account for the pricing errors. Regardless of the reasons behind the anomaly, no authors have researched the persistence of the anomaly across other popular factors, namely the factors used in Fama-French 3-Factor and 5-Factor models, and Carhart's 4-Factor model. In my thesis, I test the significance of the beta anomaly in the Mkt-RF factor, as well as whether the anomaly persists in the SMB, HML, MOM, RMW, and CMA factors.

Using stock return data from CRSP and factor return data from Kenneth R. French's Data Library from 01/1973-12/2023, I calculate beta decile excess returns for each factor, regress those returns on different combinations of factors to get their CAPM, 3-Factor, 4-Factor, and 5-Factor alphas, derive factor-neutral returns equal to those alphas, and create a betting against beta portfolio that shorts the highest beta decile and goes long in the lowest beta decile. I find that the beta anomaly is statistically significant for the Mkt-RF factor at only the CAPM level, but not the 5-Factor level. This proves that Frazzini and Pedersen's findings are not robust, and there needs to be further considerations to make their BAB factor-neutral to more factors. However, this is supported by Ehsani and Linnainmaa (2021), who show that the Mkt-RF BAB factor loses significance when adjusted for 4 or more factors. This means that a CAPM investor can bet against the Mkt-RF factor to generate positive alpha, but a 5-Factor investor cannot. I also find that the HML factor shows a strong positive correlation between CAPM alphas and betas, which is the reverse of the beta anomaly. This shows up in both the correlation coefficient and t-stat of its BAB alpha. This behavior can be made into a "Betting with Beta" strategy that is long in high-HML-beta assets and short in low-HML-beta assets. The only factors to display a beta anomaly pattern at the 5-Factor level are the RMW and CMA factors. Interestingly, these factors do not display the pattern at the CAPM level, instead showing the opposite pattern. At the 5-Factor level, the alphas have a strong negative correlation with RMW betas, but the difference between the top and bottom deciles is not extremely large. For the CMA factor at the 5-Factor level, the correlation is not as strong, but the difference between the top and bottom deciles is larger. Both factors show promise to be used in a 5-Factor Betting against Beta strategy.

In the future, I hope to expand and refine my research to provide stronger evidence for my results. One avenue of potential research would be to test the returns of my selected factor betas under different periods and market conditions. This approach would be similar to Hedegaard (2018), where I would test if certain market conditions affect the returns of different BAB portfolios. Furthermore, it would be interesting to research the anomaly in other popular factors, such as Pastor and Stambaugh's liquidity factor. I would also like to test different portfolio weightings for each decile portfolio. Currently, the monthly returns for each decile portfolio are an average of all assets within the portfolio, making them equal in weight. There are certain volatility weighting techniques that would reduce the volatility of each portfolio, thus leading to higher Sharpe Ratios. Lastly, I believe there are better methods to construct the BAB portfolio similar to the one used in Frazzini and Pedersen, in which each asset weight is determined by the measured beta. Currently, the BAB portfolio only takes a long position in the lowest decile and a short position in the highest decile. This ignores the returns of all other deciles. I believe a beta-weighted portfolio will lead to better results as it does not depend on the highest and lowest decile returns. Accomplishing these goals is beyond the scope of my senior thesis, and I hope that either myself or another interested individual can extend my research.

References

Asness, C. S., Frazzini, A., & Pedersen, L. H. (2013). Low-Risk Investing without industry bets. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.2259244

Bali, T. G., Brown, S. J., Murray, S., & Tang, Y. (2017). A Lottery-Demand-Based Explanation of the Beta Anomaly. The Journal of Financial and Quantitative Analysis, 52(6), 2369–2397. https://www.jstor.org/stable/26590484

Berrada, T., Messikh, R. J., Oderda, G., & Pictet, O. V. (2014). Beta-Arbitrage strategies: When do they work, and why? SSRN Electronic Journal. https://doi.org/10.2139/ssrn.1976285

Buchner, A., & Wagner, N. (2015). The betting against beta anomaly: Fact or fiction? Finance Research Letters, 16, 283–289. https://doi.org/10.1016/j.frl.2015.12.010

Carhart, M. M. (1997). On Persistence in Mutual Fund Performance. The Journal of Finance, 52(1), 57–82. https://doi.org/10.1111/j.1540-6261.1997.tb03808.x

Ehsani, S., & Linnainmaa, J. T. (2021b). The invisible portfolio. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3855066

Fama, E. F., & French, K. R. (2014). A five-Factor asset pricing model. Journal of Financial Economics, 116(1), 1–22. https://doi.org/10.1016/j.jfineco.2014.10.010

Hedegaard, E. (2018). Time-Varying Leverage Demand and Predictability of Betting-Against-Beta. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3194626

Fama, E. F., & French, K. R. (1993). Common Risk Factors in the Returns on Stocks and Bonds. Journal of Financial Economics, 33(1), 3–56.

Frazzini, A., & Pedersen, L. H. (2013). Betting against beta. Journal of Financial Economics, 111(1), 1–25. https://doi.org/10.1016/j.jfineco.2013.10.005

Novy-Marx, R., & Velikov, M. (2021). Betting against betting against beta. Journal of Financial

Economics, 143(1), 80–106. https://doi.org/10.1016/j.jfineco.2021.05.023

Appendix

Additional 3-Factor and 4-Factor Figures and Tables:

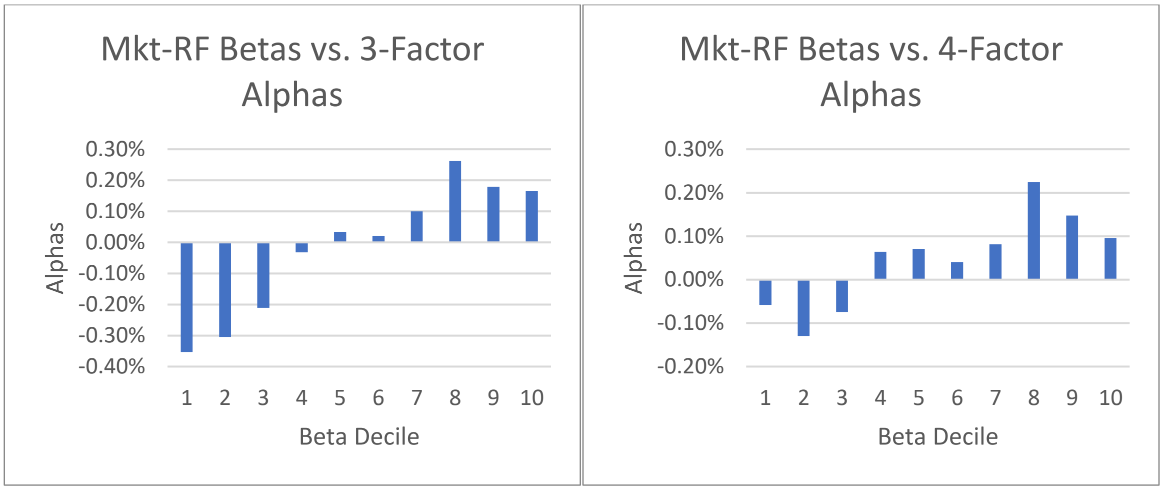


Figure A1a and A1b: Mkt-RF Beta Deciles vs. 3-Factor and 4-Factor-Neutral Returns

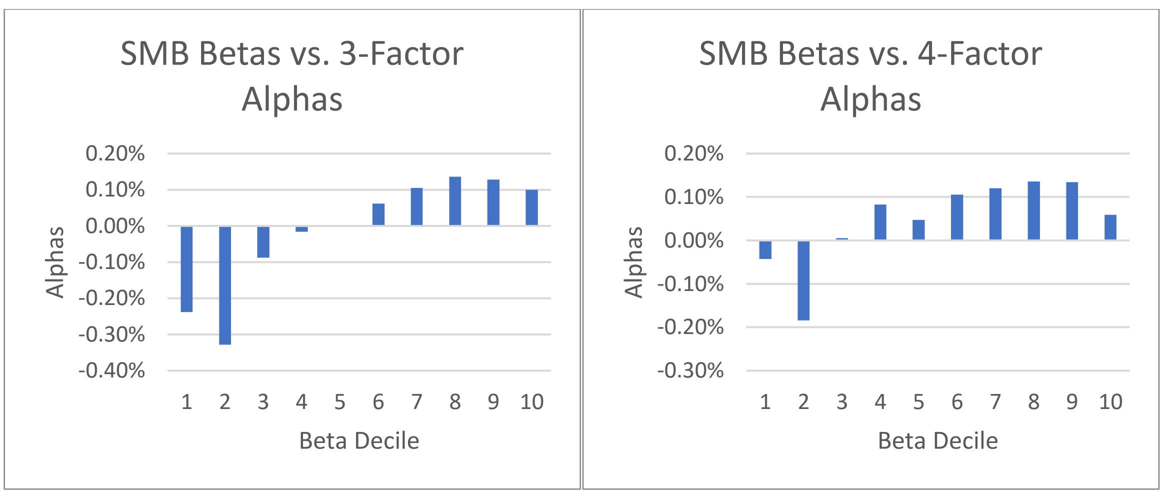


Figure A2a and A2b: SMB Beta Deciles vs. 3-Factor and 4-Factor Alphas

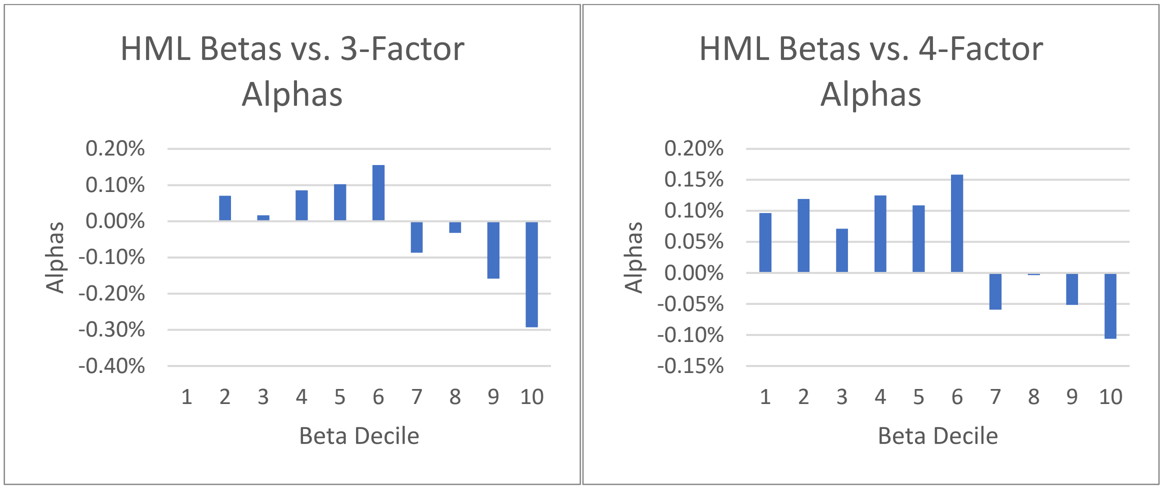


Figure A3a and A3b: HML Beta Deciles vs. 3-Factor and 4-Factor Alphas

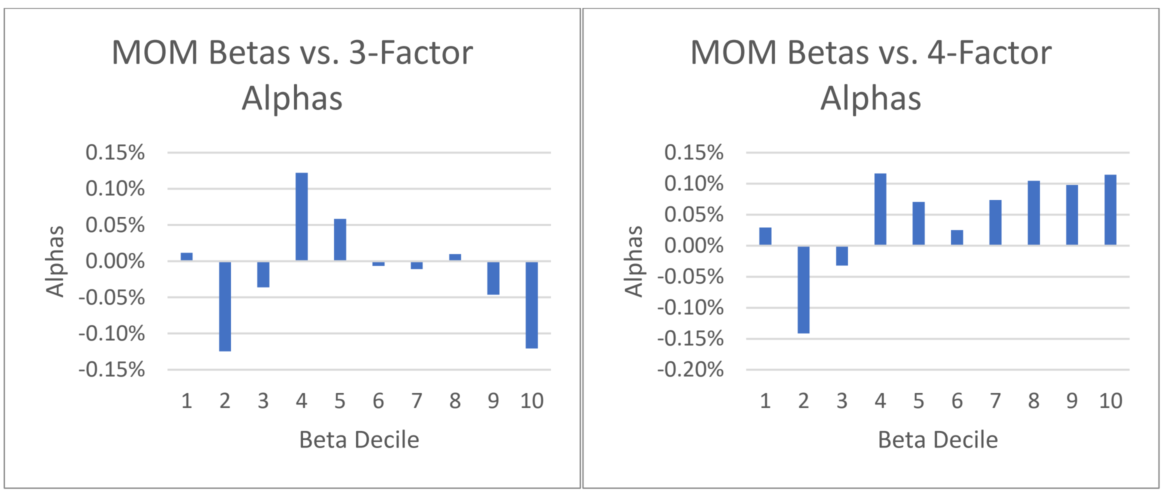


Figure A4a and A4b: MOM Beta Deciles vs. 3-Factor and 4-Factor Alphas

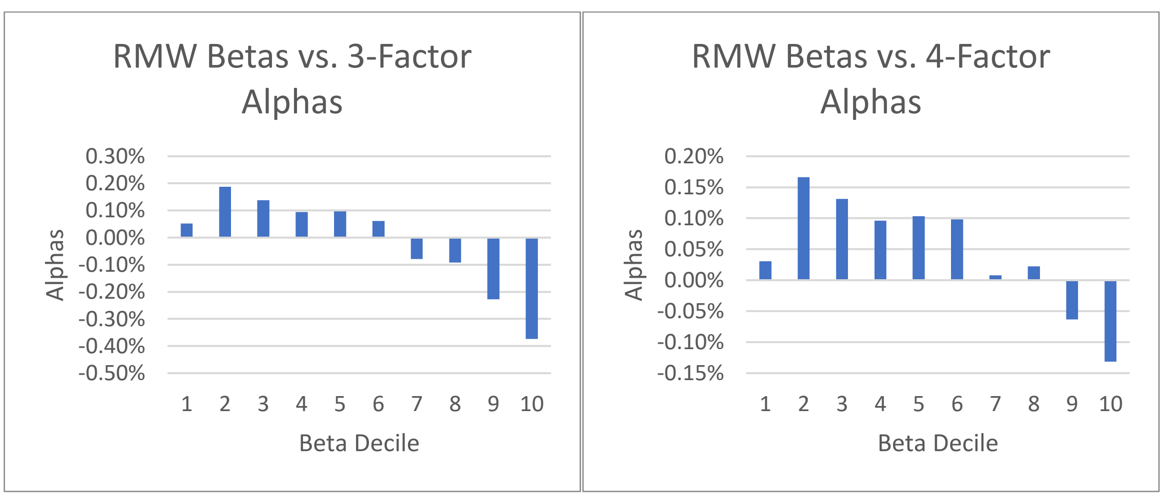
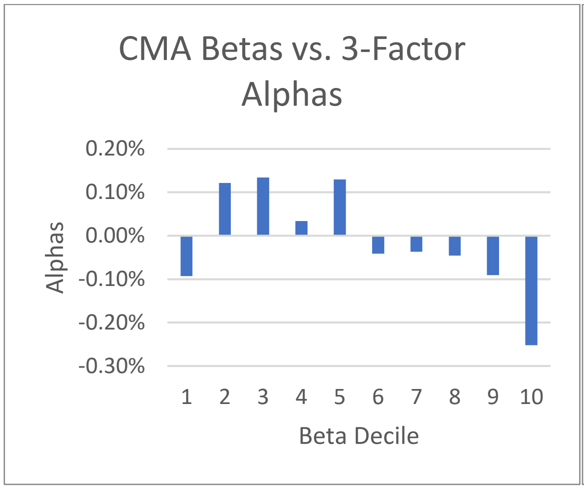


Figure A5a and A5b: RMW Beta Deciles vs. 3-Factor and 4-Factor Alphas



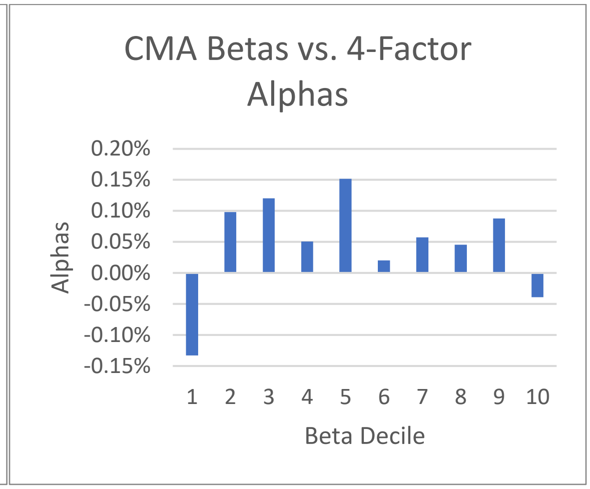


Figure A6a and A6b: CMA Beta Deciles vs. 3-Factor and 4-Factor Alphas

Table A1: Mkt-RF Beta Decile 3-Factor-Neutral Statistics

Beta Decile	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Arith Mean	-4.23%	-3.65%	-2.53%	-0.39%	0.39%	0.25%	1.19%	3.15%	2.15%	1.98%
t-stat	-2.09	-2.92	-2.64	-0.47	0.52	0.30	1.46	3.56	2.35	1.61
Geo Mean	-5.03%	-3.94%	-2.71%	-0.55%	0.26%	0.10%	1.05%	3.01%	1.97%	1.64%
Volatility	13.75%	8.48%	6.50%	5.62%	5.07%	5.48%	5.54%	6.00%	6.20%	8.30%
Sharpe Ratio	-0.31	-0.43	-0.39	-0.07	0.08	0.04	0.22	0.52	0.35	0.24
R-Squared	0.79	0.87	0.90	0.91	0.91	0.88	0.86	0.81	0.74	0.47

Table A2: Mkt-RF Beta Decile 4-Factor-Neutral Statistics

Beta Decile	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Arith Mean	-0.69%	-1.56%	-0.90%	0.77%	0.85%	0.48%	0.97%	2.69%	1.77%	1.14%
t-stat	-0.37	-1.36	-1.02	0.98	1.15	0.59	1.19	3.06	1.94	0.94
Geo Mean	-1.45%	-1.84%	-1.07%	0.63%	0.73%	0.33%	0.82%	2.54%	1.59%	0.81%
Volatility	12.54%	7.79%	5.95%	5.31%	5.02%	5.46%	5.52%	5.96%	6.17%	8.20%
Sharpe Ratio	-0.06	-0.20	-0.15	0.14	0.17	0.09	0.18	0.45	0.29	0.14
R-Squared	0.82	0.89	0.92	0.92	0.91	0.88	0.86	0.81	0.74	0.48

Table A3: SMB Beta Decile 3-Factor-Neutral Statistics

Beta Decile	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Arith Mean	-2.86%	-3.93%	-1.06%	-0.19%	-0.03%	0.74%	1.26%	1.64%	1.54%	1.19%
t-stat	-1.56	-3.18	-1.26	-0.23	-0.03	0.90	1.55	2.05	1.72	1.05
Geo Mean	-3.56%	-4.20%	-1.21%	-0.34%	-0.18%	0.59%	1.11%	1.50%	1.37%	0.90%
Volatility	12.43%	8.40%	5.70%	5.52%	5.59%	5.60%	5.50%	5.42%	6.08%	7.69%
Sharpe Ratio	-0.23	-0.47	-0.19	-0.03	0.00	0.13	0.23	0.30	0.25	0.16
R-Squared	0.80	0.86	0.91	0.90	0.89	0.88	0.87	0.85	0.78	0.67

Table A4: SMB Beta Decile 4-Factor-Neutral Statistics

Beta Decile	1	2	3	4	5	6	7	8	9	10
Arith Mean	-0.52%	-2.21%	0.06%	0.99%	0.57%	1.26%	1.44%	1.63%	1.61%	0.70%
t-stat	-0.30	-1.89	0.07	1.30	0.70	1.55	1.78	2.04	1.79	0.62
Geo Mean	-1.20%	-2.50%	-0.09%	0.86%	0.42%	1.11%	1.30%	1.50%	1.43%	0.41%
Volatility	11.86%	7.94%	5.42%	5.19%	5.51%	5.54%	5.49%	5.42%	6.08%	7.65%
Sharpe Ratio	-0.04	-0.28	0.01	0.19	0.10	0.23	0.26	0.30	0.26	0.09
R-Squared	0.82	0.88	0.92	0.92	0.89	0.88	0.87	0.85	0.78	0.68

Table A5: HML Beta Decile 3-Factor-neutral Statistics

Beta Decile	1	2	3	4	5	6	7	8	9	10
Arith Mean	-0.01%	0.85%	0.20%	1.03%	1.22%	1.86%	-1.04%	-0.38%	-1.90%	-3.52%
t-stat	-0.01	0.84	0.23	1.10	1.50	2.25	-1.25	-0.41	-1.46	-1.86
Geo Mean	-0.49%	0.62%	0.02%	0.83%	1.08%	1.72%	-1.20%	-0.58%	-2.26%	-4.24%
Volatility	9.87%	6.85%	5.92%	6.32%	5.53%	5.60%	5.67%	6.32%	8.85%	12.80%
Sharpe Ratio	0.00	0.12	0.03	0.16	0.22	0.33	-0.18	-0.06	-0.22	-0.27
R-Squared	0.76	0.83	0.87	0.84	0.87	0.87	0.88	0.88	0.83	0.78

Table A6: HML Beta Decile 4-Factor-Neutral Statistics

Beta Decile	1	2	3	4	5	6	7	8	9	10
Arith Mean	1.16%	1.43%	0.85%	1.50%	1.31%	1.90%	-0.71%	-0.04%	-0.62%	-1.28%
t-stat	0.81	1.43	0.99	1.62	1.61	2.30	-0.86	-0.05	-0.49	-0.70
Geo Mean	0.70%	1.21%	0.69%	1.31%	1.16%	1.76%	-0.87%	-0.24%	-0.98%	-2.00%
Volatility	9.69%	6.78%	5.82%	6.28%	5.53%	5.60%	5.64%	6.30%	8.61%	12.29%
Sharpe Ratio	0.12	0.21	0.15	0.24	0.24	0.34	-0.13	-0.01	-0.07	-0.10
R-Squared	0.77	0.83	0.87	0.84	0.87	0.87	0.88	0.88	0.83	0.80

Table A7: MOM Beta Decile 3-Factor-Neutral Statistics

Beta Decile	1	2	3	4	5	6	7	8	9	10
Arith Mean	0.14%	-1.49%	-0.43%	1.47%	0.70%	-0.08%	-0.13%	0.12%	-0.56%	-1.45%
t-stat	0.09	-1.35	-0.46	2.00	0.94	-0.10	-0.14	0.12	-0.54	-0.93
Geo Mean	-0.45%	-1.76%	-0.63%	1.35%	0.58%	-0.23%	-0.34%	-0.10%	-0.80%	-1.98%
Volatility	10.90%	7.50%	6.35%	4.97%	5.07%	5.52%	6.46%	6.56%	7.08%	10.61%
Sharpe Ratio	0.01	-0.20	-0.07	0.29	0.14	-0.01	-0.02	0.02	-0.08	-0.14
R-Squared	0.75	0.82	0.84	0.90	0.90	0.88	0.85	0.86	0.85	0.78

Table A8: MOM Beta Decile 4-Factor-Neutral Statistics

Beta Decile	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Arith Mean	0.35%	-1.70%	-0.38%	1.40%	0.85%	0.30%	0.88%	1.26%	1.17%	1.37%
t-stat	0.22	-1.54	-0.41	1.90	1.13	0.37	0.96	1.35	1.22	0.97
Geo Mean	-0.24%	-1.96%	-0.58%	1.28%	0.72%	0.15%	0.69%	1.06%	0.97%	0.92%
Volatility	10.89%	7.50%	6.35%	4.97%	5.07%	5.49%	6.26%	6.31%	6.52%	9.61%
Sharpe Ratio	0.03	-0.23	-0.06	0.28	0.17	0.05	0.14	0.20	0.18	0.14
R-Squared	0.75	0.82	0.84	0.90	0.90	0.88	0.86	0.87	0.88	0.82

Table A9: RMW Beta Decile 3-Factor-Neutral Statistics

Beta Decile	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Arith Mean	0.63%	2.25%	1.65%	1.13%	1.16%	0.74%	-0.94%	-1.10%	-2.73%	-4.48%
t-stat	0.44	2.07	1.75	1.36	1.41	0.93	-1.29	-1.14	-2.03	-2.09
Geo Mean	0.16%	2.00%	1.46%	0.98%	1.01%	0.60%	-1.06%	-1.31%	-3.10%	-5.38%
Volatility	9.65%	7.34%	6.40%	5.65%	5.58%	5.38%	4.95%	6.55%	9.13%	14.54%
Sharpe Ratio	0.06	0.31	0.26	0.20	0.21	0.14	-0.19	-0.17	-0.30	-0.31
R-Squared	0.67	0.76	0.82	0.86	0.87	0.89	0.92	0.88	0.81	0.71

Table A10: RMW Beta Decile 4-Factor-Neutral Statistics

Beta Decile	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Arith Mean	0.37%	2.00%	1.57%	1.15%	1.24%	1.18%	0.10%	0.27%	-0.76%	-1.58%
t-stat	0.26	1.85	1.67	1.38	1.51	1.50	0.14	0.29	-0.60	-0.78
Geo Mean	-0.10%	1.74%	1.38%	1.00%	1.09%	1.04%	-0.01%	0.08%	-1.12%	-2.48%
Volatility	9.64%	7.33%	6.39%	5.65%	5.58%	5.33%	4.66%	6.18%	8.57%	13.79%
Sharpe Ratio	0.04	0.27	0.25	0.20	0.22	0.22	0.02	0.04	-0.09	-0.11
R-Squared	0.67	0.77	0.82	0.86	0.87	0.90	0.92	0.89	0.83	0.74

Table A11: CMA Beta Decile 3-Factor-Neutral Statistics

Beta Decile	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Arith Mean	-1.12%	1.45%	1.61%	0.40%	1.55%	-0.49%	-0.44%	-0.55%	-1.09%	-3.03%
t-stat	-0.84	1.56	1.88	0.47	1.96	-0.60	-0.55	-0.58	-0.82	-1.57
Geo Mean	-1.51%	1.26%	1.45%	0.23%	1.42%	-0.64%	-0.59%	-0.75%	-1.49%	-3.81%
Volatility	8.97%	6.33%	5.81%	5.83%	5.38%	5.54%	5.47%	6.38%	9.05%	13.12%
Sharpe Ratio	-0.12	0.23	0.28	0.07	0.29	-0.09	-0.08	-0.09	-0.12	-0.23
R-Squared	0.72	0.80	0.84	0.85	0.88	0.89	0.90	0.89	0.83	0.78

Table A12: CMA Beta Decile 4-Factor-Neutral Statistics

Beta Decile	1	2	3	4	5	6	7	8	9	10
Arith Mean	-1.60%	1.17%	1.44%	0.61%	1.82%	0.24%	0.69%	0.55%	1.05%	-0.47%
t-stat	-1.21	1.26	1.68	0.71	2.31	0.30	0.91	0.60	0.85	-0.25
Geo Mean	-1.98%	0.98%	1.28%	0.44%	1.69%	0.10%	0.56%	0.36%	0.70%	-1.22%
Volatility	8.93%	6.31%	5.81%	5.82%	5.36%	5.41%	5.17%	6.14%	8.39%	12.47%
Sharpe Ratio	-0.18	0.19	0.25	0.10	0.34	0.04	0.13	0.09	0.13	-0.04
R-Squared	0.72	0.80	0.84	0.85	0.88	0.89	0.91	0.90	0.86	0.81

Table A13: 3-Factor-Neutral BAB Performance

Factor	Mkt-RF	SMB	HML	MOM	RMW	CMA
Arithmetic Mean	6.21%	4.05%	-3.50%	-1.59%	-5.10%	-1.91%
t-stat	2.30	1.67	-1.34	-0.65	-1.65	-0.71
Geometric Mean	4.57%	2.70%	-4.97%	-2.92%	-7.08%	-3.48%
Volatility	18.31%	16.46%	17.78%	16.61%	21.02%	18.20%
Sharpe Ratio	0.34	0.25	-0.20	-0.10	-0.24	-0.11
Correlation Coefficient	-0.94	-0.89	0.67	0.22	0.87	0.63

Table A14: 4-Factor-Neutral BAB Performance

Factor	Mkt-RF	SMB	HML	MOM	RMW	CMA
Arithmetic Mean	1.83%	1.22%	-2.44%	1.02%	-1.94%	1.13%
t-stat	0.73	0.52	-0.93	0.43	-0.65	0.44
Geometric Mean	0.37%	-0.06%	-3.94%	-0.27%	-3.95%	-0.38%
Volatility	16.93%	15.83%	17.70%	16.08%	20.41%	17.55%
Sharpe Ratio	0.11	0.08	-0.14	0.06	-0.10	0.06
Correlation Coefficient	-0.83	-0.73	0.77	-0.67	0.77	-0.06

Code (Can also be found on GitHub at https://github.com/kjiang25/Betting-Against-All-Betas)

```
# Import Packages
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import statsmodels.api as sm
from math import prod

# Read stock data
data = pd.read_csv("Full dataset.csv")
data['date'] = pd.to_datetime(data['date'])
data.rename(columns={'PERMNO': 'Permno'}, inplace=True)
```

```
data['date'] = data['date'] + pd.offsets.MonthEnd(0)
data['MktCap'] = data['PRC'] * data['SHROUT']

# Check for duplicates and remove if found
duplicates = data.groupby(['date', 'Permno']).size().reset_index(name='count')
duplicates = duplicates[duplicates['count'] > 1]

if not duplicates.empty:
  print("Found duplicate entries:")
  print(duplicates)

  returns = data.groupby(['date', 'Permno'])['RET'].mean().reset_index()

# Pivot stock data to returns dataframe
returns = data.pivot(index='date', columns='Permno', values='RET')
returns = returns.applymap(lambda x: pd.to_numeric(x, errors='coerce'))

# Display returns.head
returns.head()

# Read three factor file
three_factors = pd.read_csv("Fama French 3 Factors.csv")
three_factors.columns = three_factors.iloc[2]
three_factors = three_factors.iloc[3:1181]
three_factors.rename(columns={three_factors.columns[0]: 'Date'}, inplace = True)
three_factors['Date'] = pd.to_datetime(three_factors['Date'], format = '%Y%m') + pd.offsets.MonthEnd(0)

# Read five factor file
five_factors = pd.read_csv("Fama French 5 Factors.csv")
five_factors.columns = five_factors.iloc[2]
five_factors = five_factors.iloc[3:737]
five_factors.rename(columns={five_factors.columns[0]: 'Date'}, inplace = True)
five_factors['Date'] = pd.to_datetime(five_factors['Date'], format = '%Y%m') + pd.offsets.MonthEnd(0)

# Read momentum factor file
Mom_factor = pd.read_csv("Fama French Momentum Factor.csv")
Mom_factor = Mom_factor.iloc[:1172]
Mom_factor['Date'] = pd.to_datetime(Mom_factor['Date'], format = '%Y%m') + pd.offsets.MonthEnd(0)

# Merge factors into one dataframe
factors = pd.merge(three_factors[['Date', 'Mkt-RF', 'SMB', 'HML']], Mom_factor[['Date', 'MOM']], on='Date')
factors = pd.merge(factors, five_factors[['Date', 'RMW', 'CMA', 'RF']], on='Date')
factors = factors.loc[(factors['Date'] >= pd.to_datetime('1973-01')) & (factors['Date'] <= pd.to_datetime('2024-01'))]
factors.set_index('Date', inplace = True)
factors = factors.astype(float)
factors /= 100
```

```python
# Display factors.head
factors.head()
# Create copy for excess returns
excess_returns = returns.copy()

# Subtract RF from all columns
excess_returns = excess_returns.sub(factors['RF'], axis=0)

# Fill na
excess_returns[returns.isna()] = returns[returns.isna()]

# Display excess_returns.head
excess_returns.head()


# Function to calculate deciles thresholds from 0 to 90
def calculate_deciles(data):
  deciles = []
  for i in range(0, 91, 10):
    decile = np.percentile(data, i)
    deciles.append(decile)
  return deciles

# Function to average the innermost values of nested lists, used to calculate average returns of group of stocks
def average_innermost_lists(nested_lists):

    if not isinstance(nested_lists, list):
        return nested_lists

    # Check if we're at the innermost list (contains numbers)
    if nested_lists and not isinstance(nested_lists[0], list):
        return np.mean(nested_lists)

    # Recursively process nested lists
    return [average_innermost_lists(sublist) for sublist in nested_lists]


# Initialize date window for regressions
window_start = pd.to_datetime('1973-01-31')
window_end = window_start + pd.DateOffset(years=5) - pd.offsets.MonthEnd(1)
final_date = pd.to_datetime(returns.index[-1])

date_index = pd.date_range(start=window_end + pd.offsets.MonthEnd(1), end=final_date, freq='M')

# Initialize lists to store decile returns
```

```
decile_returns = []

# Select factor to analyze
factor = 'Mkt-RF'

# Run regressions with monthly frequency
while window_end < final_date:

  # Get top 500 assets at the end of each calendar year
  next_year = window_end + pd.DateOffset(years=1)
  df_date = data[data['date'] == window_end]
  top_assets = df_date.sort_values('MktCap', ascending = False).head(500)
  permnos = top_assets['Permno'].tolist()

  # Monthly loop
  while window_end < next_year and window_end < final_date:

    # Initialize lists to store decile buckets and beta+returns for month
    decile_returns_row = [[] for i in range(10)]
    betas_returns = []

    # Iterate through assets
    for permno in permnos:

      # Filter data by regression window
      filtered_returns = excess_returns.loc[window_start:window_end]
      filtered_factors = factors.loc[window_start:window_end]

      # Run regression
      if filtered_returns[permno].notnull().sum() / len(filtered_returns )> 0.5 and filtered_returns[permno][-
12:].notnull().all(): # at least 50% of returns, most recent year filled
        regression = sm.OLS(filtered_returns[permno], sm.add_constant(filtered_factors[factor]),missing='drop').fit() #
Calculate speficied factor beta
        beta = regression.params[factor] # Get beta
        next_month_excess_return = excess_returns.loc[window_end + pd.offsets.MonthEnd(1), permno] # Get next
month excess return of that asset

        # Store beta and next month excess return pair into list if next month return is not NaN
        if not np.isnan(next_month_excess_return):
          br_entry = [beta, next_month_excess_return]
          betas_returns.append(br_entry)

    # Calculate beta decile thresholds
    betas = [item[0] for item in betas_returns]
    deciles = calculate_deciles(betas)
```

```
  # Iterate through beta+returns pairs and append next month returns to corresponding decile bucket
  for i in range(len(betas_returns)):
    for d in range(9, -1, -1): # Iterate backwards from high beta decile to low beta decile
      if betas_returns[i][0] >= deciles[d]: # If beta is greater than or equal to decile threshold
        decile_returns_row[9-d].append(betas_returns[i][1]) # Append next month return to corresponding decile
bucket, where decile 0 is the highest beta decile and decile 9 is the lowest beta decile
        break

  decile_returns_row = average_innermost_lists(decile_returns_row) # Calculate average excess next month returns
for each decile
  decile_returns.append(decile_returns_row) # Append decile returns to list

  # Increment dates by one month
  window_start += pd.offsets.MonthEnd(1)
  window_end += pd.offsets.MonthEnd(1)


decile_returnsdf = pd.DataFrame(decile_returns, columns = [i for i in range(1, 11)], index = date_index) # Convert
nested decile returns list to dataframe
decile_returnsdf.head()


# Initialize regression window
regression_start = pd.to_datetime('1978-01-31')
regression_end = pd.to_datetime(decile_returnsdf.index[-1])

# Initialize empty factor neutral decile returns dataframe and lists to store data
factor_neutral_decile_returnsdf = pd.DataFrame(columns = [i for i in range(1, 11)], index = decile_returnsdf.index)
alphas = [[] for i in range(10)]
factor_betas = [[] for i in range(10)]
r_squared = [[] for i in range(10)]

# Iterate through each decile's returns
for i in range(10):

 # Initialize lists to store betas
 factor_beta_entries = []

 # Filter factor data
 filtered_factors = factors.loc[regression_start:regression_end, ['Mkt-RF']] # CAPM specification
 # filtered_factors = factors.loc[regression_start:regression_end, ['Mkt-RF', 'SMB', 'HML']] # 3-Factor specification
 # filtered_factors = factors.loc[regression_start:regression_end, ['Mkt-RF', 'SMB', 'HML', 'MOM']] # 4-Factor
specification
 # filtered_factors = factors.loc[regression_start:regression_end, ['Mkt-RF', 'SMB', 'HML', 'RMW', 'CMA']] # 5-Factor
specification

 regression = sm.OLS(decile_returnsdf[i+1], sm.add_constant(filtered_factors), missing='drop').fit() # Run regression
```

```
  mkt_beta = regression.params['Mkt-RF'] # Assign beta value
  # smb_beta = regression.params['SMB'] # Add if necessary
  # hml_beta = regression.params['HML']
  # mom_beta = regression.params['MOM']
  # rmw_beta = regression.params['RMW']
  # cma_beta = regression.params['CMA']
  r_squared[i] = regression.rsquared # Store R-squared value
  alphas[i] = regression.params['const'] # Store alpha value

  # Append factor betas to list
  factor_beta_entries.append(mkt_beta)
  # factor_beta_entries.append(smb_beta) # Add if necessary
  # factor_beta_entries.append(hml_beta)
  # factor_beta_entries.append(mom_beta)
  # factor_beta_entries.append(rmw_beta)
  # factor_beta_entries.append(cma_beta)

  # Calculate factor neutral decile returns equal to alpha of regression
  factor_neutral_decile_returnsdf[i+1] = decile_returnsdf[i+1] - mkt_beta * filtered_factors['Mkt-RF'] # CAPM
specification
  # factor_neutral_decile_returnsdf[i+1] = decile_returnsdf[i+1] - mkt_beta * filtered_factors['Mkt-RF'] - smb_beta *
filtered_factors['SMB'] - hml_beta * filtered_factors['HML'] # 3-Factor specification
  # factor_neutral_decile_returnsdf[i+1] = decile_returnsdf[i+1] - mkt_beta * filtered_factors['Mkt-RF'] - smb_beta *
filtered_factors['SMB'] - hml_beta * filtered_factors['HML'] - mom_beta * filtered_factors['MOM'] # 4-Factor
specification
  # factor_neutral_decile_returnsdf[i+1] = decile_returnsdf[i+1] - mkt_beta * filtered_factors['Mkt-RF'] - smb_beta *
filtered_factors['SMB'] - hml_beta * filtered_factors['HML'] - rmw_beta * filtered_factors['RMW'] - cma_beta *
filtered_factors['CMA'] # 5-Factor specification

  # Append factor betas to list
  factor_betas[i] = factor_beta_entries

# Convert r_squared list to dataframe
r_squareddf = pd.DataFrame(r_squared).T
r_squareddf.columns = [i for i in range(1, 11)]

# Convert alphas list to dataframe
alphasdf = pd.DataFrame(alphas).T
alphasdf.columns = [i for i in range(1, 11)]

# Convert factor_betas list to dataframe
factor_betasdf = pd.DataFrame(factor_betas).T
factor_betasdf.columns = [i for i in range(1, 11)]

factor_betasdf # Display factor betas, each row corresponds to a factor
```

```
r_squareddf # Display R-squared values

# Calculate average factor neutral returns for each decile
returns_avg = pd.DataFrame([factor_neutral_decile_returnsdf.mean()])
returns_avg

alphasdf # Display alphas, equal to average factor neutral returns

plt.bar(alphasdf.columns, alphasdf.iloc[0]) # Plot alphas on beta deciles
plt.xlabel('Decile')
plt.ylabel('Alphas')


# Initialize list to store all performance metrics
performance_metrics = []

# Iterate through each decile and calculate performance metrics
for i in range(10):

  # Initialize list to store performance metrics for each decile
  entry = []
  monthly_arithmetic_mean = factor_neutral_decile_returnsdf[i+1].mean() # Calculate monthly arithmetic mean
  annual_arithmetic_mean = monthly_arithmetic_mean * 12 # Calculate annual arithmetic mean
  monthly_volatility = factor_neutral_decile_returnsdf[i+1].std() # Calculate monthly volatility
  annual_volatility = monthly_volatility * np.sqrt(12) # Calculate annual volatility
  monthly_geometric_mean = (1 + factor_neutral_decile_returnsdf[i+1]).prod() **
(1/len(factor_neutral_decile_returnsdf)) - 1 # Calculate monthly geometric mean
  annual_geometric_mean = (1 + monthly_geometric_mean) ** 12 - 1 # Calculate annual geometric mean
  sharpe_ratio = annual_arithmetic_mean / annual_volatility # Calculate Sharpe ratio
  t_stat = (np.sqrt(len(factor_neutral_decile_returnsdf)) * sharpe_ratio) / np.sqrt(12) # Calculate t-statistic

  # Append performance metrics to list
  entry.append(annual_arithmetic_mean)
  entry.append(t_stat)
  entry.append(annual_geometric_mean)
  entry.append(annual_volatility)
  entry.append(sharpe_ratio)
  performance_metrics.append(entry)

index = [i for i in range(1,11)] # Assign index values for deciles
columns = ['Arithmetic Mean', 't-stat', 'Geometric Mean', 'Volatility', 'Sharpe Ratio'] # Assign column names for
performance metrics
performance_metricsdf = pd.DataFrame(performance_metrics, columns = columns, index = index).T # Convert
performance metrics list to dataframe
performance_metricsdf
```

```
# Create betting against beta portfolio that goes long the lowest beta decile and short the highest beta decile
betting_against_betadf = pd.DataFrame(factor_neutral_decile_returnsdf[10] - factor_neutral_decile_returnsdf[1],
columns=['BAB Return'])

# Calculate performance metrics for betting against beta portfolio
monthly_arithmetic_mean = betting_against_betadf['BAB Return'].mean() # Calculate monthly arithmetic mean
annual_arithmetic_mean = monthly_arithmetic_mean * 12 # Calculate annual arithmetic mean
monthly_volatility = betting_against_betadf['BAB Return'].std() # Calculate monthly volatility
annual_volatility = monthly_volatility * np.sqrt(12) # Calculate annual volatility
monthly_geometric_mean = (1 + betting_against_betadf['BAB Return']).prod() **
(1/len(factor_neutral_decile_returnsdf)) - 1 # Calculate monthly geometric mean
annual_geometric_mean = (1 + monthly_geometric_mean) ** 12 - 1 # Calculate annual geometric mean
sharpe_ratio = annual_arithmetic_mean / annual_volatility # Calculate Sharpe ratio
t_stat = (np.sqrt(len(factor_neutral_decile_returnsdf)) * sharpe_ratio) / np.sqrt(12) # Calculate t-statistic
print('Arithmetic Mean: ', annual_arithmetic_mean)
print('t-stat: ', t_stat)
print('Geometric Mean: ', annual_geometric_mean)
print('Volatility: ', annual_volatility)
print('Sharpe Ratio: ', sharpe_ratio)
```