Model Implied Asset Pricing Beliefs: An Empirical Analysis for S&P 500® Constituents from 2010–2020

By Christian Satzky, FRM
Email: c.satzky@gmail.com

Last update: 13 May 2021

Contents

Abstract

Common investment beliefs result from academic research and conventional wisdom. In this study, I am deriving explicit and implicit investment hypotheses from the Capital Asset Pricing Model (Sharpe, 1964), the Fama and French 5 Factor Model (Fama and French, 1993), and common price multiples. I empirically investigate the credibility of these investment beliefs using data from S&P 500® constituents from January 2010–December 2020. The significance of each investment hypothesis is assessed by two methods of statistical evaluation. The goal of this research is to provide insights to asset managers in support of well-informed financial decision-making.

Keywords: *Asset Pricing*, *Factor Investing*, *Empirical Analysis of Financial Markets*.

1 Introduction

The rational investor seeks to optimize an investment's risk/return profile. An investment's risk is commonly modeled using estimates of volatility. Determining an asset's expected risk is straightforward, as volatility exhibits significant dependencies across time. Determining an asset's expected return is more challenging, as asset return time-series generally lack exploitable time-dependencies. Statistically, these findings are consistent with the asset's price process following a geometric Brownian motion (GBM) with time-dependent standard deviation.

However, academic research in asset pricing suggests that return time-series are impacted by exogenous variables, such as the overall stock market's return. Consequently, a company's expected return is dependent on the company's exposure to such *risk factor*. This section provides a brief overview of determining a stock's expected return based on academic research.

1.1 The Capital Asset Pricing Model (CAPM)[1]

The most elementary pricing model is the Capital Asset Pricing Model (CAPM, e.g. Sharpe, 1964). It considers the relationship between a company's share price return and the overall market's return. Under the CAPM, the expected return $\mathbb{E}[r_i]$ for company i is given by the market risk premium and the company's exposure to it. The market risk premium is defined as the difference between the market's return and the risk-free rate. The exposure of a company's return to the market risk premium is measured by Beta, β_i. More formally:

$$\mathbb{E}[r_i] = \mathbb{E}[r_f] + \beta_i \underbrace{\mathbb{E}[r_m - r_f]}_{\substack{\text{Expected Market} \\ \text{Risk Premium}}} \tag{1}$$

Where:

r_i	Return of company i
r_f	Return of the risk-free asset
r_m	Return of the stock market
β_i	Beta of company i

Equation (1) shows: When the market risk premium is expected to be *positive*, the greater the Beta of a company, the greater the company's expected return. Likewise, if the expected market risk premium is *negative*, a company with a high value of Beta is expected to exceed the losses of a company with a low value of Beta.

[1] In Appendix A I present a method on how to efficiently compute daily estimates of Beta in large datasets.

Beta can be interpreted as "leverage" of the market risk premium. If $\mathbb{E}[r_m - r_f] > 0$, then:

$\beta < 1$	$\beta = 1$	$\beta > 1$
The expected stock return is lower than the market's return	The expected stock return is equal to the market's return	The expected stock return is greater than the market's return

Thus, when the market risk premium is assumed to be positive, it is sensible to look for companies with a high value of Beta. Ordering companies by *descending* values of Beta and forming portfolios out of the top and bottom-ranking companies exploits this idea.[2]

The CAPM considers a company's exposure to the overall market's return. In the next section, fundamental characteristics of a company are included as well.

1.2 The Fama and French 5 Factor Model

Eugene Fama and Kenneth French (1993) extend the CAPM by adding four factors associated with a company's fundamental characteristics. Under this research, a 'factor', similarly to the market risk premium in the CAPM, is a return time-series. The general idea is to construct a 'factor' as follows:

1. Order many companies within the company universe by a specific characteristic;
2. Form a "long portfolio" from top-ranking companies and a "short portfolio" from bottom-ranking companies;
3. Compute the return spread between the long and the short portfolio.

For example, if the relevant firm characteristic is Value, all companies in the company universe are ordered by *descending* values of the Book-to-Market ratio. From this list of companies, the long portfolio includes 30% of the top-ranking companies. The short portfolio includes 30% of bottom-ranking companies. In this example, the return spread of the long and the short portfolio should reasonably approximate the High Minus Low (HML) or the Value-factor within the Fama and French framework.[3]

[2] This idea is investigated in the Beta Portfolios-section

[3] The exact computation of the Fama and French factors differs. Further information is provided in appendix B: Fama and French Factors Computation

Under the Fama and French 5 factor model, a company's expected excess return is given by:

$$\begin{aligned}\mathbb{E}[r_i - r_f] = \alpha &+ \beta_i^{Market}\,\mathbb{E}[r_m - r_f] \\ &+ \beta_i^{Size}\,\mathbb{E}[SMB] \\ &+ \beta_i^{Value}\,\mathbb{E}[HML] \\ &+ \beta_i^{Profitability}\,\mathbb{E}[RMW] \\ &+ \beta_i^{Investment}\,\mathbb{E}[CMA]\end{aligned} \tag{2}$$

Where:

r_i	Return of company i
r_f	Return of the risk-free asset
r_m	Return of the stock market
β_i^f	Company i's exposure to factor f

The firm characteristics to rank companies by to obtain the SMB, HML, RMW, and CMA factors are given in table 1.

Table 1: Fama and French Factors

Order Companies by	Calculation*†	Order	To Compute Factor‡
Size	Market Cap_t	Ascending	Small Minus Big (SMB)
Value	$\frac{\text{Book Equity}_q}{\text{Market Cap}_t}$	Descending	High Minus Low (HML)
Profitability	$\frac{\sum_{j=0}^{3}\text{EBT}_{q-j}}{\text{Book Equity}_q}$	Descending	Robust Minus Weak (RMW)
Investment	$\frac{\text{Assets}_q}{\text{Assets}_{q-4}} - 1$	Ascending	Conservative Minus Aggresssive (CMA)

* Approximate formulas

† Where t is the date index, q is the date of the most recently observed 10-Q report, and $t \geq q$

‡ The average of the return spreads between the 'long' and 'short' portfolios. The exact computation of the Fama and French factors is provided in Appendix B

Most importantly, Fama and French find that ordering companies by these firm characteristics yields a positive return spread between the top-ranking and the bottom-ranking portfolios. In line with the theoretical risk-return trade-off, Fama and French interpret these positive returns as a compensation for risk. However, this interpretation is debatable. For instance, any such finding of excess return could also be due to a (temporary) market anomaly.

Applicability in Practice

A company's five Betas in equation (2) can be estimated using daily data of a company's return and the daily data provided by Kenneth R. French for the Market, SMB, HML, RMW, and CMA factors.[4] Once the company's exposure to these factors is estimated, the expected return depends on the expected values of the market, SMB, HML, and CMA factors. This approach, however, is problematic. First, a company's betas in equation (2) cannot be reasonably expected to remain constant over time. Secondly, the future values of the Market, Size, Value, Profitability, and Investment factors are unknown. Hence, this article focuses on an alternative method of applying this research in practice, which is outlined in the Methodology-section.

Reproducibility

Reproducing positive portfolio return spreads based on the company characteristics identified by Fama and French is challenging. First, conditions under which the stock market values companies might be constantly changing. Fama and French investigated a dataset spawning from 1963–1991. It is at least questionable whether their findings are reproducible in modern times. Secondly, the company universe from which companies are ordered from is not disclosed.[5] If it includes illiquid companies and/or companies that are not subject to generally accepted accounting practices and/or regulatory oversight, or are in any other way disqualifying for investment, any findings of excess returns might only be reproducible *in theory*.

2 Research Objective

The goal of this paper is to obtain statistically sound *approval* or *disapproval* of specific investment beliefs to support financial decision-making in modern practice. More specifically, I am empirically evaluating investment hypotheses formed by popular belief and implied by the most prominent academic research in the field of factor investing. To maximize applicability in practice, I am restricting this study to highly investable S&P 500® companies and most-recent data spawning from January 2010 to December 2020.

The hypothesized investing beliefs under investigation are defined a priori and are listed in table 2.

[4] Daily data for the 5 factors for North American stocks is availble on the website of Kenneth R. French

[5] Only geographic regions of companies are provided, e.g. "North America". A list of actual companies included in the analysis is not disclosed.

Table 2: Model Implied Asset Pricing Beliefs

Company Characteristic	Investment Hypothesis
Capital Asset Pricing Model	
Beta	High Beta outperforms low Beta
Fama & French 5 Factor Model	
Market Cap	Small outperforms Big
Book-to-Market	High Value outperforms low Value
EBT-to-Book	High Profitability outperforms low Profitability
YoY Assets Growth	Low Investment outperforms high Investment
Common Price Multiples*	
Price-to-Earnings (PE)	Low PE outperforms high PE
Price-to-Revenue (PR)	Low PR outperforms high PR
Price-to-Free Cash Flow (PCF)	Low PCF outperforms high PCF
Sharelab Limited†	
GrowthSpread® (GS)	High GS outperforms low GS

* No underlying theoretical model. For ranking consistency, price multiples are computed in terms of their reciprocal values.

† No underlying theoretical model. GrowthSpread® is a 'fair-value' indicator developed by Sharelab Limited in 2018.

The methods and success measures to statistically evaluate the aforementioned investment hypotheses are outlined in the next section.

Note: In this study, findings of statistical support in favor of any of the aforementioned asset pricing beliefs, and findings of a *lack* of statistical evidence for any such beliefs, are considered to be of equal importance to the industry practitioner.

3 Methodology

The asset pricing beliefs under investigation are listed in table 2. All nine investment hypotheses are defined a priori, i.e. prior to conducting research. For each investment hypothesis, a long and a short portfolio is built. I construct these two portfolios by ranking S&P 500® constituents by a specific company characteristic. From this list of companies, the *long* portfolio includes 30% of the top-ranking companies, and the *short* portfolio includes 30% of the bottom-ranking companies.

For example, when it comes to the "high Value outperforms low Value"-hypothesis, S&P 500® companies are ordered by *descending* values of the Book-to-Market ratio. The long portfolio includes the top 30% of companies, i.e. those companies having the highest Book-to-Market ratio. The short portfolio includes the bottom 30% of companies, i.e. those companies having the lowest Book-to-Market ratio. The portfolios' holdings are updated on the first U.S. trading day of each month.

For each company characteristic in table 2, the performances of the respective long and short portfolios are recorded from from 4 January 2010 to 31 December 2020. This data forms the basis for statistical evaluation of the investigated asset pricing beliefs. In general, for an investment hypothesis to be credible, the long portfolio should positively divert from the short portfolio.

3.1 Statistical Evaluation

There are two methods for evaluation of the statistical significance of the investigated investment hypotheses:

1. The long and the short portfolios' performances are compared to the performances of 100,000 random portfolios. This provides a measure of percentile rank of the long and short portfolios against the estimated population of possible portfolio returns.
2. A spread portfolio is built by investing in the long portfolio and short-selling the short portfolio: $\mathrm{PF}_{\mathrm{ls}} = \mathrm{PF}_{long} - \mathrm{PF}_{short}$. The expected monthly change in the spread portfolio is tested against the null $H_0 : \mathbb{E}[\Delta\mathrm{PF}_{\mathrm{ls}}] \leq 0$.

When it comes to (1), the performances of the 100,000 simulated portfolios are fair to compare to the performances of the long and short portfolios of each hypothesis. They are constructed as follows:

- On the first trading day of each month, each of the 100,000 simulated portfolios picks 30% of S&P 500® companies uniformly at random;
- In case a company characteristic requires the company universe to be filtered, the simulated portfolios are also restricted to the same available companies to choose from;[6]
- Dividends (if any) accrue over the monthly holding period and are reinvested evenly into the portfolio in the following month.

3.2 Success Measures

An investment hypothesis in table 2 is deemed *true*, if either or both of the following statements are true:

- The long portfolio ranks in the 90^{th} percentile or higher in comparison to the 100,000 simulated portfolio returns, *and* the short portfolio ranks in the 10^{th} percentile or lower in comparison to the 100,000 simulated portfolio returns;
- The expected change in the spread portfolio is significantly greater than zero, considering a significance level of $\alpha \leq 0.05$.

[6] For example, the Book-to-Market ratio requires a company's Book Equity to be greater than zero, and hence only these companies are included

3.3 Research-Specfific Challenges

In table 3 I list challenges of empirical research in asset pricing and my approach to eliminate these problems in this study.

Table 3: Common Problems with Research in Empirical Asset Pricing

Issue	Problem	Elimination Approach
1	No reproducibility and lack of transparency	Full reproducibility of all (interim-) results through accessible data and programming code*
2	Investigation of time periods far in the past	Investigation of most-recent time period (Jan-2010 to Dec-2020)
3	Practically non-investable company universe	Highly investable and liquid company universe (S&P 500® constituents)
4	Look ahead bias	Avoidance of all identified look-ahead biases
5	Data dredging / 'p-hacking'	1. All 9 firm characteristics under investigation are defined a priori 2. Full disclosure of all data related investigations
6	Lack of measures to evaluate statistical significance	Evaluation of statistical significance by:† 1. Monte Carlo simulation estimated population of possible return paths 2. Hypothesis test on the expected change in the long/short spread portfolio
7	Non-disclosure of implicit and explicit assumptions	Full disclosure of known implicit and explicit assumptions‡
8	Non-disclosure of limitations of research	Disclosure of limitations of research§

* Contact: c.satzky@gmail.com

† See the Statistical Evaluation-section

‡ See Appendix C

§ See the Limitations and Outlook-section

4 Results

For the CAPM-Beta portfolios, i.e. the "high Beta outperforms low Beta"-hypothesis, I discuss the plots and results in great detail. The methodology for all other hypotheses is analogous to the Beta portfolios. Hence, I am providing a shortened summary of results for all other investment hypotheses.

4.1 Beta Portfolios

This section discusses the results for the hypothesis: *"High Beta companies outperform low Beta companies"*. A company's CAPM-Beta is estimated as follows:

$$\hat{\beta}_{i,t} = \frac{\sum r_{i,t} r_{m,t}}{\sum r_{m,t}^2}$$

Where:

$r_{i,t}$:	CAPM-Beta estimate for company i on day t
$r_{m,t}$:	Return of company i on day t
$\hat{\beta}_{i,t}$:	Return of the S&P 500® index on day t

In this study, a company's Beta is estimated on a rolling basis using 252 past, daily returns. This reflects a calendar year's worth of daily data. By using this procedure, no look-ahead bias is introduced: If the portfolio is built on date t, Beta is estimated using returns spawning from $t - 251$ to t.

The underlying assumptions of the Beta estimates are:

1. Daily returns of the S&P 500® index proxy the overall market's return;
2. The daily return of the risk-free asset is zero.

The programming code used to estimate a company's Beta is disclosed in Appendix A.

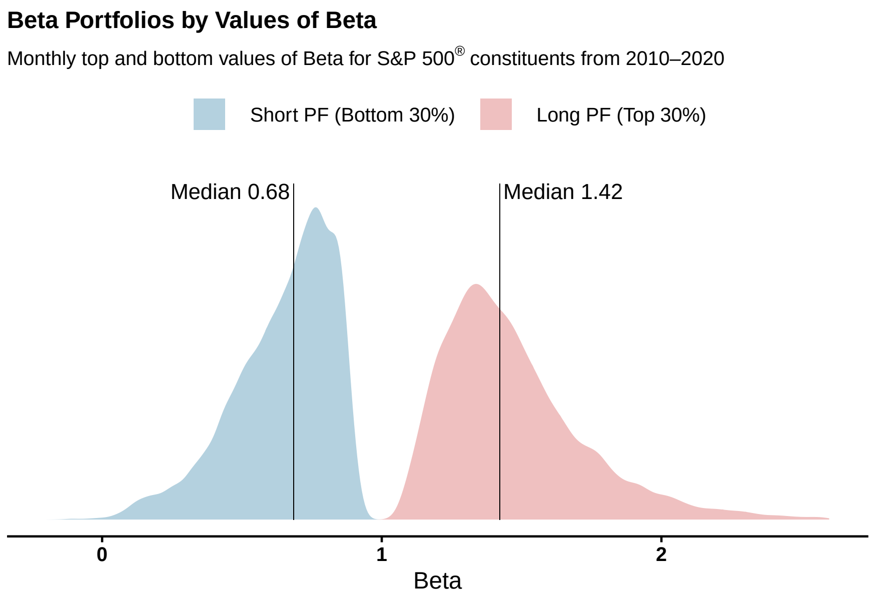
In the following, Beta is computed for each company in the dataset and for each first trading date of the month from 2010–2020. On each first trading date of the month, two portfolios are built:

- The **long portfolio** contains 30% of S&P 500® companies with the highest values of Beta;
- The **short portfolio** contains 30% of S&P 500® companies with the lowest values of Beta.

Thus, each portfolio contains roughly 150 companies at a time. For more details, see the Methodology-section.

Pooling the Beta estimates for each time a portfolio is built from 1/2010–12/2020, the Beta estimates are distributed as follows:

Figure 1: Distribution of CAPM-Beta Across Long and Short Portfolios

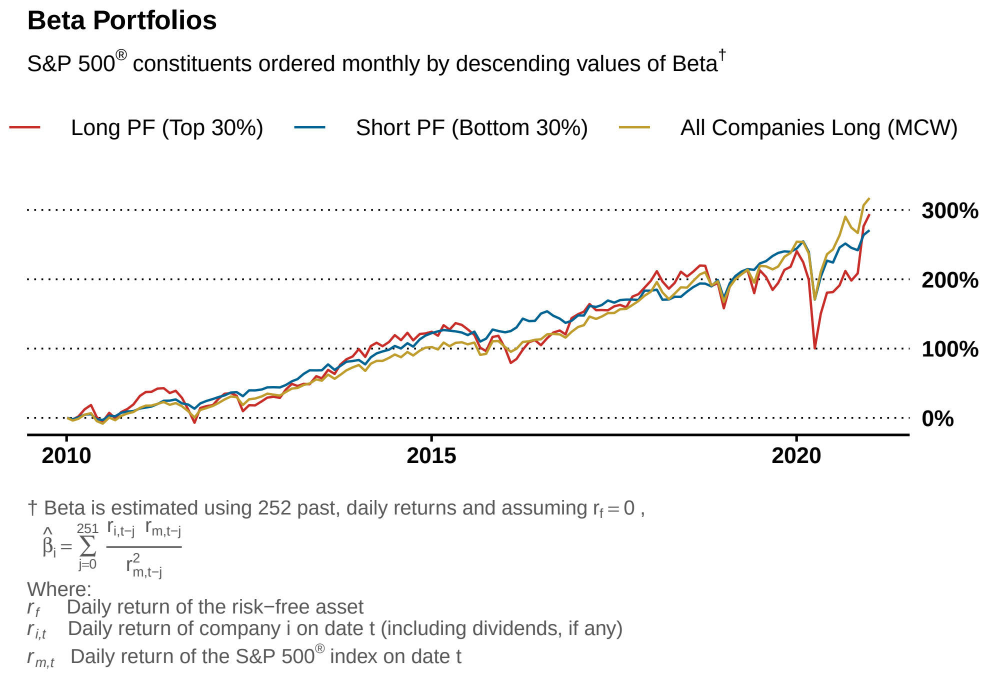


The median value of Beta in the long portfolio is 2.08 times greater than the median Beta value in the short portfolio, i.e. median$(\hat{\beta}_{long}) = 1.42$ versus median$(\hat{\beta}_{short}) = 0.68$. Hence, the resulting long and short portfolios are suitable to investigate the hypothesis: *"High Beta companies outperform low Beta companies."*

The long and short portfolios' cumulative returns (with reinvested dividends, if any) are plotted in figure 2.[7]

[7] For comparison, the plot includes the performance of a hypothetical portfolio including *100*% of S&P 500® constituents that have Beta estimates. This benchmark portfolio is Market Capitalization weighted (MCW) and hence proxies the performance of the S&P 500® index. This portfolio can be interpreted as the Market Capitalization weighted total return of the complete stock universe, from which the long and short portfolios are subsets of. It includes roughly 500 companies at a time.

Figure 2: Cumulative Return of the Beta Portfolios

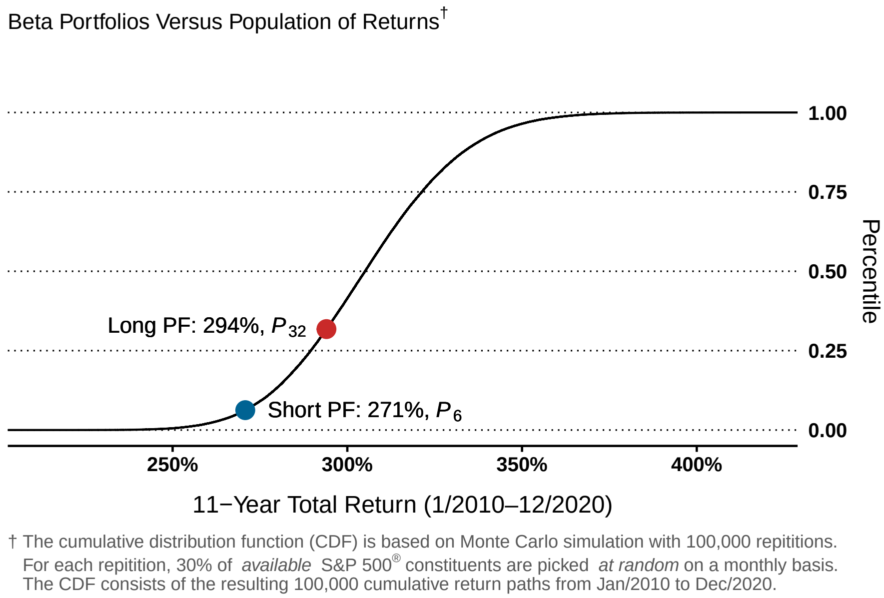


Subjectively, the long portfolio does not appear to divert enough from the short portfolio to reasonably support the hypothesis under investigation. To obtain objective judgment, I carry out two methods of statistical evaluation.

Monte Carlo Simulation: Beta Portfolios Versus 100,000 Random Portfolios

To evaluate whether the return of the Beta portfolios significantly differs from portfolios drawn at random, their performances are compared to the population of possible returns. The population of possible returns is estimated by simulating the performances of 100,000 portfolios. On a monthly basis, each of these portfolios draws 30% of S&P 500® constituents uniformly at random. The company universe includes only those companies having a Beta estimate. The long Beta portfolio is deemed significantly *outperforming* if it ranks in the 90^{th} or greater percentile. The short Beta portfolio is considered significantly *underperforming* if it ranks in the 10^{th} or smaller percentile. For more details, see the Statistical Evaluation-section.

Figure 3: Beta Portfolios Versus Estimated Population of Returns



The long portfolio (high Beta) ranks in the 31.79^{th} percentile. Hence, out of 100 portfolios, the high Beta portfolio, on average, outperforms 32 portfolios. Since $P_{31.79} \not> P_{90}$, the long portfolio does not significantly outperform portfolios drawn at random.

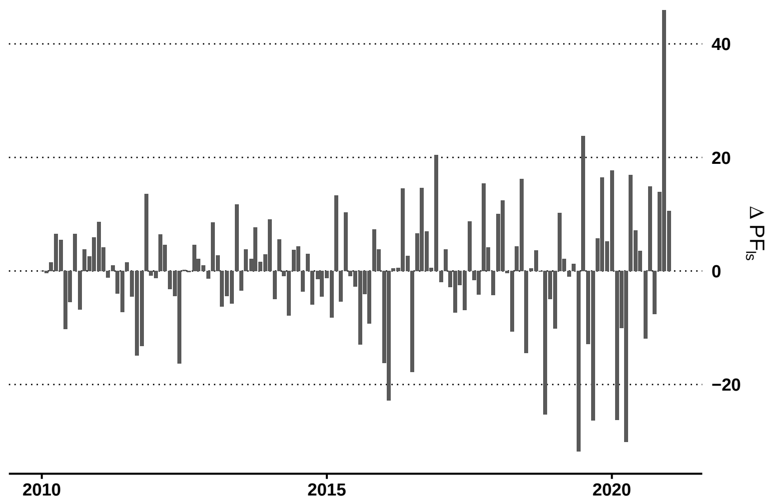
The short portfolio (low Beta) ranks in the 6.26^{th} percentile. Hence, out of 100 portfolios, the low Beta portfolio, on average, underperforms 94 portfolios. Since $P_{6.26} < P_{10}$, the short portfolio **significantly underperforms** portfolios drawn at random.

Albeit the low Beta portfolio is significantly underperforming portfolios drawn at random, this is not sufficient statistical evidence in support of the investment belief: *"High Beta companies outperform low Beta companies"*. For more information, see the Success Measures-section.

Change in the Long/Short Portfolio Spread

If the *"high Beta companies outperform low Beta companies"*-hypothesis is true, we expect the long portfolio to positively divert from the short portfolio. Judging from figure 2, this does not seem the case. More formally, consider the long/short spread portfolio PF_{ls}, yielding the performance of investing into the high Beta portfolio, and short-selling the low Beta portfolio. For a hypothetical $100 attributed with each portfolio, the monthly gains/losses of such spread portfolio is shown in figure 4.

Figure 4: Monthly Gains/Losses the Spread Portfolio $\mathrm{PF}_{\mathrm{ls}}$



Judging from figure 4, the monthly change in the long/short spread portfolio ($\Delta\mathrm{PF}_{\mathrm{ls}}$) appears to be randomly fluctuating around zero, with elevated standard deviation towards the Covid-19 crisis in 2020.

For the investment hypothesis to be true, I am interested in whether the *expected change* in the long/short spread portfolio is positive. Hence, I am testing the null hypothesis: $H_0 : \mathbb{E}[\Delta\mathrm{PF}_{\mathrm{ls}}] \leq 0$. If the null hypothesis can be rejected for a significance level of $\alpha = 0.05$, I conclude that the expected change in the long/short spread portfolio is significantly positive, i.e. $H_1 : \mathbb{E}[\Delta\mathrm{PF}_{\mathrm{ls}}] > 0$. In this case, the investment hypothesis is *true*. The result for this right-tailed hypothesis test is summarized below.[8]

$$\mathbb{P}\left(\overline{\Delta\mathrm{PF}_{\mathrm{ls}}} = 0.176, \mathrm{SD}_{\Delta\mathrm{PF}_{\mathrm{ls}}} = 10.79 \middle| H_0 : \mathbb{E}[\Delta\mathrm{PF}_{\mathrm{ls}}] \leq 0\right) = 0.43$$

Where:

$\overline{\Delta\mathrm{PF}_{\mathrm{ls}}}$	The mean value of $\Delta\mathrm{PF}_{\mathrm{ls}}$
$\mathrm{SD}_{\Delta\mathrm{PF}_{\mathrm{ls}}}$	The sample standard deviation of $\Delta\mathrm{PF}_{\mathrm{ls}}$
$\Delta\mathrm{PF}_{\mathrm{ls}}$	Monthly, absolute (additive) change in the long-short-spread portfolio

Since p-value $= 0.43 > \alpha = 0.05$, the null hypothesis of negative or zero expected gains/losses in the Beta spread portfolio cannot be rejected. Thus, there is no statistical evidence that the expected change in the Beta spread portfolio is positive.

[8]**Note:** This frequentist hypothesis test gives the probability ("p-value") of observing the data $\overline{\Delta\mathrm{PF}_{\mathrm{ls}}} = 0.18$ and $\mathrm{SD}_{\Delta\mathrm{PF}_{\mathrm{ls}}} = 10.79$, while assuming that H_0 is *true*. For a small enough p-value, we reject H_0 and conclude that H_1 is true.

Verdict

The low Beta portfolio **significantly underperforms** portfolios drawn at random. However, there is no statistical evidence in support of the hypothesis: *"High Beta companies outperform low Beta companies"*.

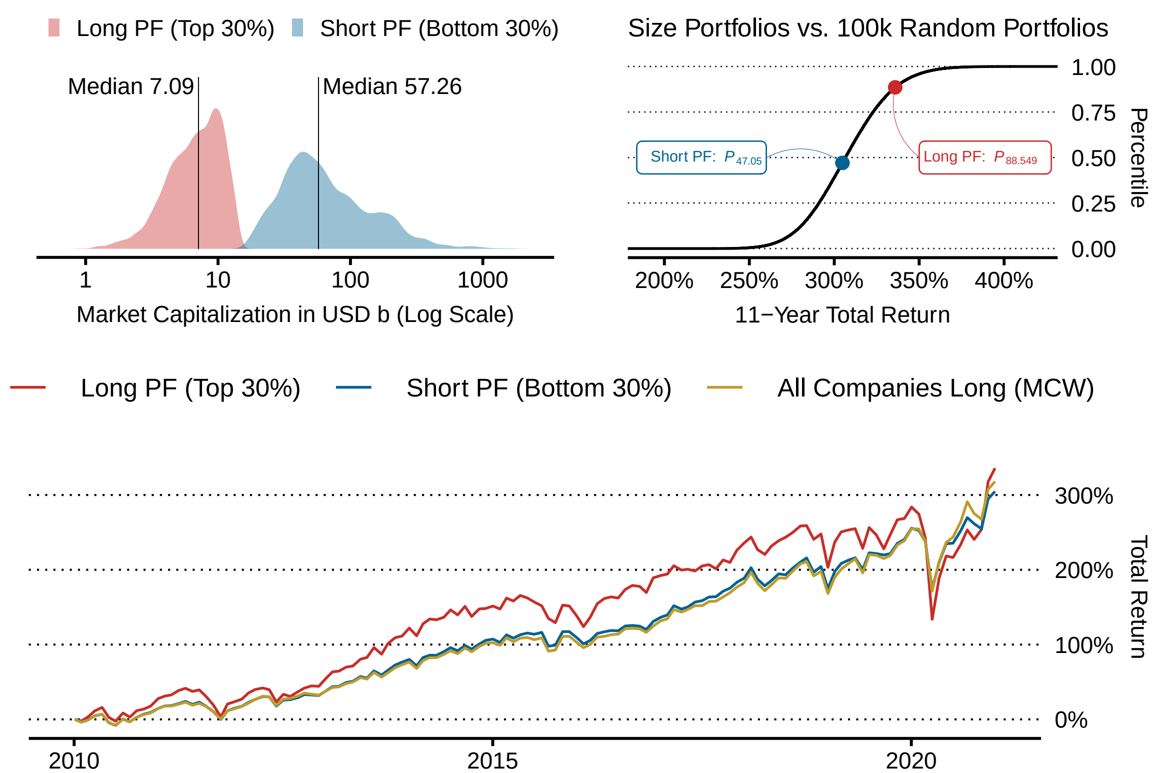
4.2 Size Portfolios

This section presents the results for the hypothesis: *"Small companies outperform big companies."* The Size portfolios under investigation are constructed as follows:

Long Portfolio:	30% of S&P 500® companies having the lowest market capitalization
Short Portfolio:	30% of S&P 500® companies having the highest market capitalization

The long and short portfolios' company holdings are updated on the first U.S. trading day of each month from January 2010 to December 2020. For details, see Methodology.

Figure 5: Size Portfolios Results



Size Portfolios Versus Random Portfolios

The upper right plot in figure 5 compares the long and short portfolios' performances to the estimated population of possible returns.

- The long portfolio (small companies) ranks in the 88.55^{th} percentile. Since $P_{88.55} \not> P_{90}$, the Small portfolio does not significantly outperform portfolios drawn at random;
- The short portfolio (big companies) ranks in the 47.05^{th} percentile. Since $P_{47.05} \not< P_{10}$, the Big portfolio does not significantly underperform portfolios drawn at random.

For details, see the Statistical Evaluation-section.

Portfolio Spread Evaluation

There is no statistical evidence that the expected change in the long/short spread portfolio is positive.

$$\mathbb{P}\left(\overline{\Delta \mathrm{PF}_{\mathrm{ls}}} = 0.002, \mathrm{SD}_{\Delta \mathrm{PF}_{\mathrm{ls}}} = 0.071 \middle| H_0 : \mathbb{E}[\Delta \mathrm{PF}_{\mathrm{ls}}] \leq 0\right) = 0.353$$

Where:

$\overline{\Delta \mathrm{PF}_{\mathrm{ls}}}$	The mean value of $\Delta \mathrm{PF}_{\mathrm{ls}}$
$\mathrm{SD}_{\Delta \mathrm{PF}_{\mathrm{ls}}}$	The sample standard deviation of $\Delta \mathrm{PF}_{\mathrm{ls}}$
$\Delta \mathrm{PF}_{\mathrm{ls}}$	Monthly, absolute (additive) change in the long-short-spread portfolio

Verdict

There is no statistical evidence to support the claim: *"Small companies outperform big companies"*. Noteworthy, the long portfolio (small companies) outperforms 88 out of 100 portfolios drawn at random.

For details, see Success Measures.

4.3 Value Portfolios

This section presents the results for the hypothesis: *"High Value companies outperform low Value companies."* A company's Value is estimated in terms of the Book-to-Market ratio (BtM):

$$\mathrm{BtM}_{i,t} = \frac{\text{Book Equity}_{i,q}}{\text{Market Capitalization}_{i,t}}$$

Where:

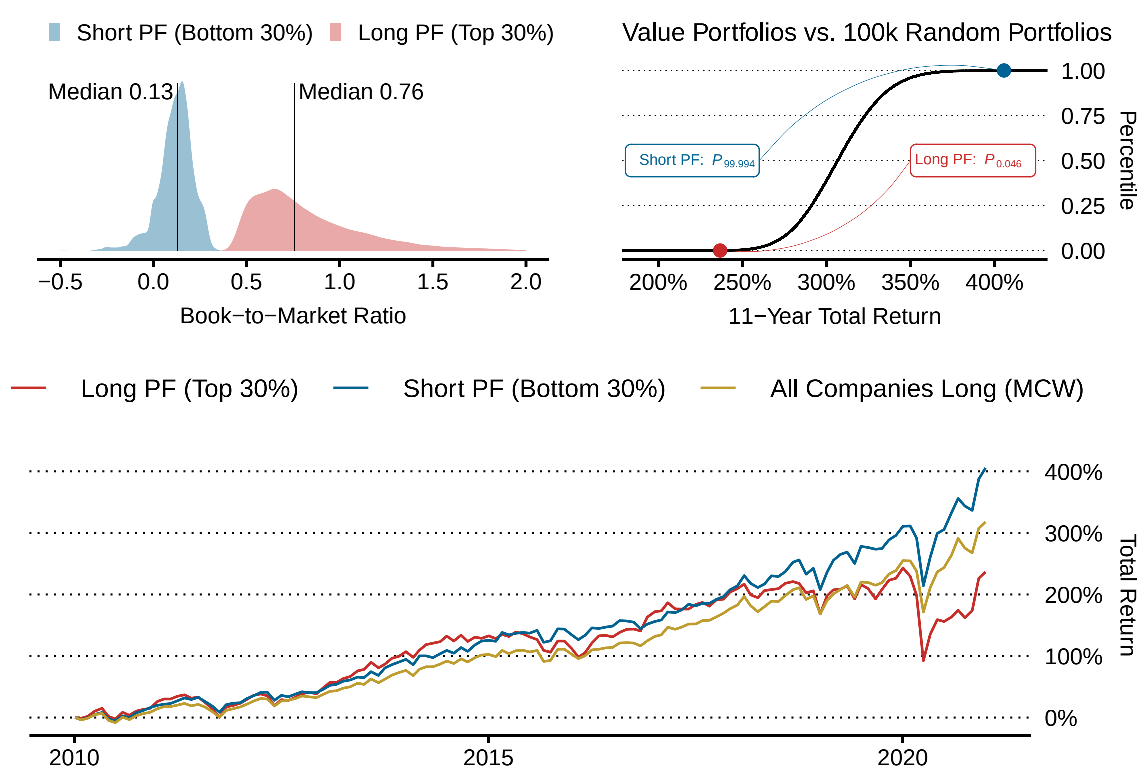
Book Equity$_{i,q}$	Book Equity for company i, observed on the most-recent quarter's publication date q
Market Capitalization$_{i,t}$	Market capitalization for company i on date t

The Value portfolios under investigation are constructed as follows:

Long Portfolio:	30% of S&P 500® companies having the highest Book-to-Market ratio (BtM)
Short Portfolio:	30% of S&P 500® companies having the lowest Book-to-Market ratio (BtM)

The long and short portfolios' company holdings are updated on the first U.S. trading day of each month from January 2010 to December 2020. For details, see Methodology.

Figure 6: Value Portfolios Results



Value Portfolios Versus Random Portfolios

The upper right plot in figure 6 compares the long and short portfolios' performances to the estimated population of possible returns.

- The long portfolio (high Value) ranks in the 0.05^{th} percentile. Since $P_{0.05} \not> P_{90}$, the high Value portfolio does not significantly outperform portfolios drawn at random;
- The short portfolio (low Value) ranks in the 99.99^{th} percentile. Since $P_{99.99} \not< P_{10}$, the low Value portfolio does not significantly underperform portfolios drawn at random.

For details, see the Statistical Evaluation-section.

Portfolio Spread Evaluation

There is no statistical evidence that the expected change in the long/short spread portfolio is positive.

$$\mathbb{P}\left(\overline{\Delta\mathrm{PF}_{\mathrm{ls}}} = -0.013, \mathrm{SD}_{\Delta\mathrm{PF}_{\mathrm{ls}}} = 0.064 \middle| H_0 : \mathbb{E}[\Delta\mathrm{PF}_{\mathrm{ls}}] \leq 0\right) = 0.989$$

Where:

$\overline{\Delta\mathrm{PF}_{\mathrm{ls}}}$	The mean value of $\Delta\mathrm{PF}_{\mathrm{ls}}$
$\mathrm{SD}_{\Delta\mathrm{PF}_{\mathrm{ls}}}$	The sample standard deviation of $\Delta\mathrm{PF}_{\mathrm{ls}}$
$\Delta\mathrm{PF}_{\mathrm{ls}}$	Monthly, absolute (additive) change in the long-short-spread portfolio

Verdict

There is no statistical evidence to support the claim: *"High Value companies outperform low Value companies"*. Noteworthy, the contrarian hypothesis would have been statistically significant.

For details, see Success Measures.

4.4 Profitability Portfolios

This section presents the results for the hypothesis: *"High Profitability outperforms low Profitability."* A company's Profitability is estimated in terms of the EBT-to-Book Equity ratio. Companies are filtered for Profitability within 1.5 standard deviations around the average Profitability.

$$\text{Profitability}_{i,t} = \frac{\sum_{j=0}^{3} \text{EBT}_{i,q-j}}{\text{Book Equity}_{i,q}}$$

Where:

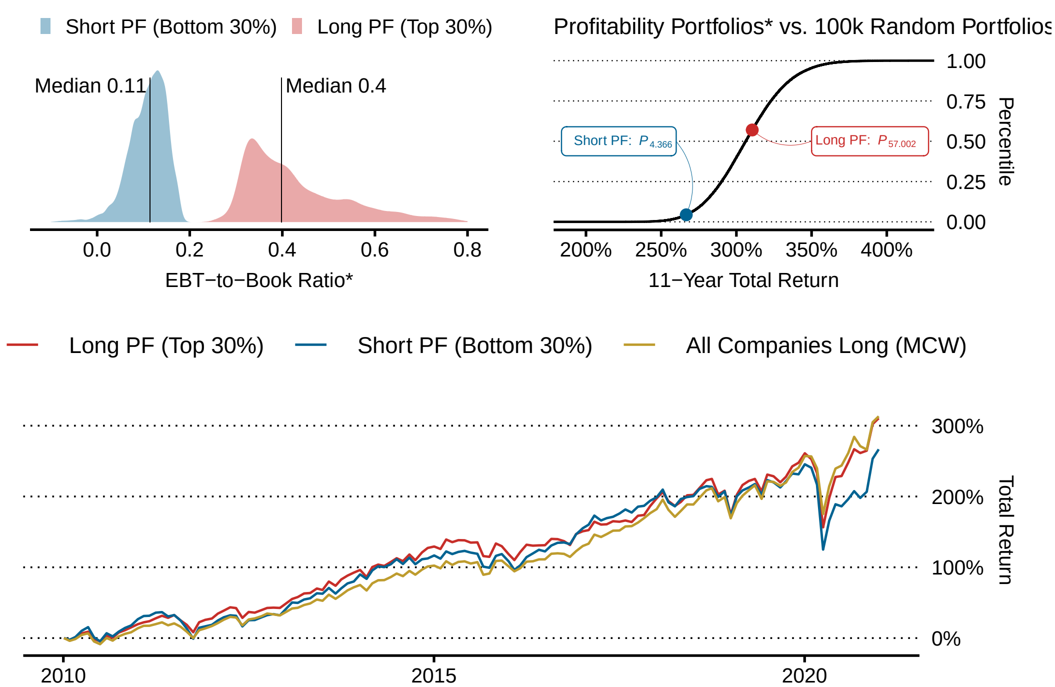
t	Daily point in time (date index)
q	Publication date of most-recently published quarterly report
i	Company i, where $i \in$ S&P 500® Constituents

The profitability portfolios under investigation are constructed as follows:

Long Portfolio:	30% of S&P 500® companies having the highest EBT-to-book ratio
Short Portfolio:	30% of S&P 500® companies having the lowest EBT-to-book ratio

The long and short portfolios' company holdings are updated on the first U.S. trading day of each month from January 2010 to December 2020. For details, see Methodology.

Figure 7: Profitability Portfolios Results



Profitability Portfolios Versus Random Portfolios

The upper right plot in figure 7 compares the long and short portfolios' performances to the estimated population of possible returns.

- The long portfolio (high Profitability) ranks in the 57^{th} percentile. Since $P_{57} \not> P_{90}$, the high Profitability portfolio does not significantly outperform portfolios drawn at random;
- The short portfolio (low Profitability) ranks in the 4.37^{th} percentile. Since $P_{4.37} \not< P_{10}$, the low Profitability portfolio does not significantly underperform portfolios drawn at random.

For details, see the Statistical Evaluation-section.

Portfolio Spread Evaluation

There is no statistical evidence that the expected change in the long/short spread portfolio is positive.

$$\mathbb{P}\left(\overline{\Delta\mathrm{PF}_{\mathrm{ls}}} = 0.003, \mathrm{SD}_{\Delta\mathrm{PF}_{\mathrm{ls}}} = 0.036 \middle| H_0 : \mathbb{E}[\Delta\mathrm{PF}_{\mathrm{ls}}] \leq 0\right) = 0.144$$

Where:

$\overline{\Delta\mathrm{PF}_{\mathrm{ls}}}$	The mean value of $\Delta\mathrm{PF}_{\mathrm{ls}}$
$\mathrm{SD}_{\Delta\mathrm{PF}_{\mathrm{ls}}}$	The sample standard deviation of $\Delta\mathrm{PF}_{\mathrm{ls}}$
$\Delta\mathrm{PF}_{\mathrm{ls}}$	Monthly, absolute (additive) change in the long-short-spread portfolio

Verdict

The short portfolio (low Profitability) **significantly underperforms** portfolios drawn at random. However, there is no statistical evidence to support the claim: *"High Profitability outperforms low Profitability"*.

For details, see Success Measures.

4.5 Investment Portfolios

This section presents the results for the hypothesis: *"Low Investment outperforms high Investment."* A company's Investment is estimated in terms of Total Assets growth:

$$\text{Investment}_{i,t} = \frac{\text{Assets}_{i,q}}{\text{Assets}_{i,q-4}} - 1$$

Where:

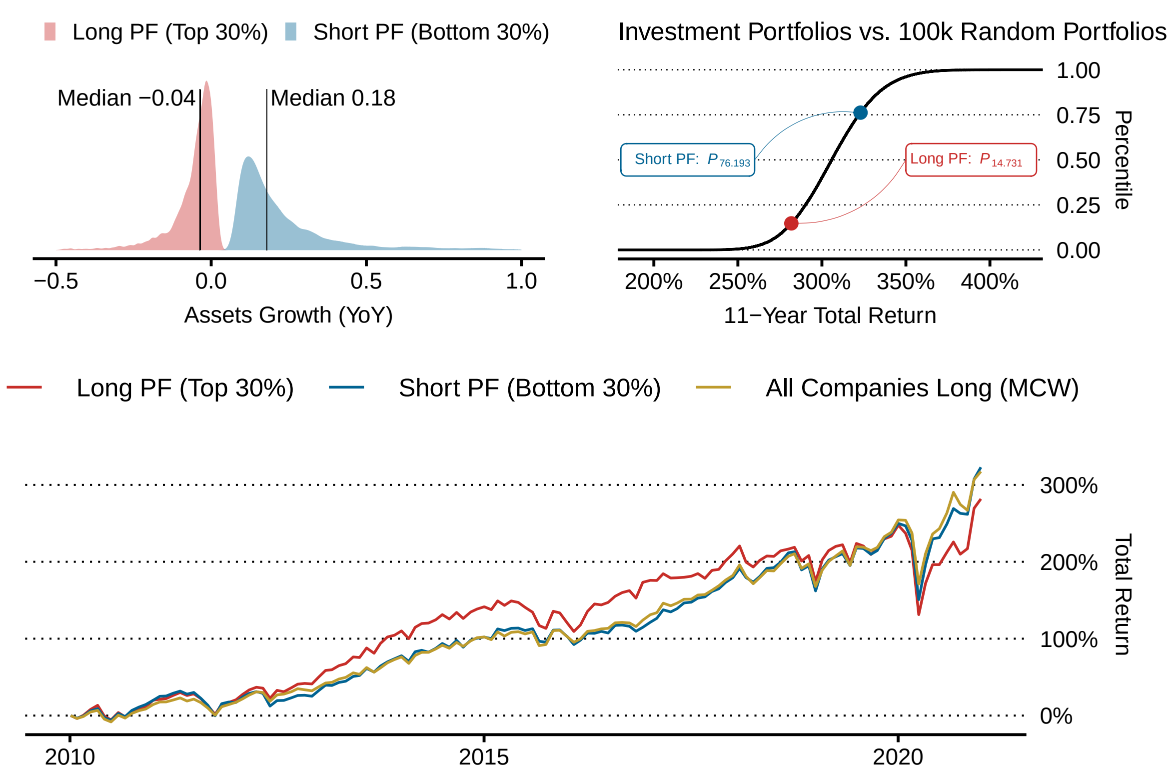
t	Daily point in time (date index)
q	Publication date of most-recently published quarterly report
i	Company i, where $i \in$ S&P 500® Constituents

The investment portfolios under investigation are constructed as follows:

Long Portfolio:	30% of S&P 500® companies having the lowest Assets growth
Short Portfolio:	30% of S&P 500® companies having the highest Assets growth

The long and short portfolios' company holdings are updated on the first U.S. trading day of each month from January 2010 to December 2020. For details, see Methodology.

Figure 8: Investment Portfolios Results



Investment Portfolios Versus Random Portfolios

The upper right plot in figure 8 compares the long and short portfolios' performances to the estimated population of possible returns.

- The long portfolio (low Investment) ranks in the 14.73^{th} percentile. Since $P_{14.73} \not> P_{90}$, the low Investment portfolio does not significantly outperform portfolios drawn at random;
- The short portfolio (high Investment) ranks in the 76.19^{th} percentile. Since $P_{76.19} \not< P_{10}$, the high Investment portfolio does not significantly underperform portfolios drawn at random.

For details, see the Statistical Evaluation-section.

Portfolio Spread Evaluation

There is no statistical evidence that the expected change in the long/short spread portfolio is positive.

$$\mathbb{P}\left(\overline{\Delta\mathrm{PF}_{\mathrm{ls}}} = -0.003, \mathrm{SD}_{\Delta\mathrm{PF}_{\mathrm{ls}}} = 0.039 \middle| H_0 : \mathbb{E}[\Delta\mathrm{PF}_{\mathrm{ls}}] \leq 0\right) = 0.819$$

Where:

$\overline{\Delta\mathrm{PF}_{\mathrm{ls}}}$	The mean value of $\Delta\mathrm{PF}_{\mathrm{ls}}$
$\mathrm{SD}_{\Delta\mathrm{PF}_{\mathrm{ls}}}$	The sample standard deviation of $\Delta\mathrm{PF}_{\mathrm{ls}}$
$\Delta\mathrm{PF}_{\mathrm{ls}}$	Monthly, absolute (additive) change in the long-short-spread portfolio

Verdict

There is no statistical evidence to support the claim: *"Low Investment outperforms High Investment"*.

For details, see Success Measures.

4.6 Price-to-Earnings Ratio (Earnings Yield) Portfolios

This section presents the results for the hypothesis: *"Low Price-to-Earnings outperforms high Price-to-Earnings"* or, equivalently, *"High Earnings Yield outperforms low Earnings Yield."*

Note: When ensuring ranking consistency across several companies;

- Price-to-Earnings ratio has the domain: $\{\text{Net Income: Net Income} > 0, \text{Market Cap: } \mathbb{R}\}$
- Earnings Yield has the domain: $\{\text{Net Income: } \mathbb{R}, \text{Market Cap: Market Cap} > 0\}$

Hence, the investment hypothesis is investigated in terms of Earnings Yield (EY), i.e. the reciprocal of the Price-to-Earnings ratio. For more information, see Appendix D.

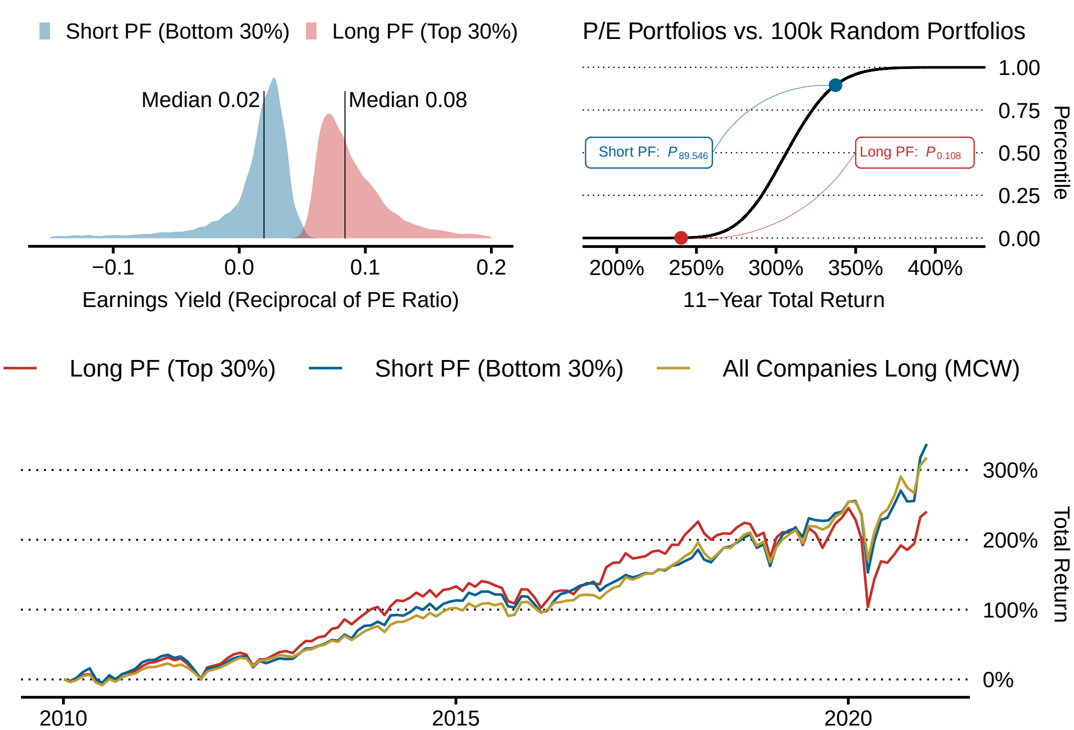
$$\text{PE}^{-1} = \text{EY} = \frac{\sum_{j=0}^{3} \text{Net Income}_{i,q-j}}{\text{Market Cap}_{i,t}}$$

The portfolios under investigation are constructed as follows:

Long Portfolio:	30% of S&P 500® companies having the highest Earnings Yield
Short Portfolio:	30% of S&P 500® of companies having the lowest Earnings Yield

The long and short portfolios' company holdings are updated on the first U.S. trading day of each month from January 2010 to December 2020. For details, see Methodology.

Figure 9: Earnings Yield Portfolios Results



Earnings Yield Portfolios Versus Random Portfolios

The upper right plot in figure 9 compares the long and short portfolios' performances to the estimated population of possible returns.

- The long portfolio (high EY) ranks in the 0.11^{th} percentile. Since $P_{0.11} \not> P_{90}$, the high EY portfolio does not significantly outperform portfolios drawn at random;
- The short portfolio (low EY) ranks in the 89.55^{th} percentile. Since $P_{89.55} \not< P_{10}$, the low EY portfolio does not significantly underperform portfolios drawn at random.

For details, see the Statistical Evaluation-section.

Portfolio Spread Evaluation

There is no statistical evidence that the expected change in the long/short spread portfolio is positive.

$$\mathbb{P}\left(\overline{\Delta \mathrm{PF}_{\mathrm{ls}}} = -0.007, \mathrm{SD}_{\Delta \mathrm{PF}_{\mathrm{ls}}} = 0.058 \middle| H_0 : \mathbb{E}[\Delta \mathrm{PF}_{\mathrm{ls}}] \leq 0\right) = 0.926$$

Where:

$\overline{\Delta \mathrm{PF}_{\mathrm{ls}}}$	The mean value of $\Delta \mathrm{PF}_{\mathrm{ls}}$
$\mathrm{SD}_{\Delta \mathrm{PF}_{\mathrm{ls}}}$	The sample standard deviation of $\Delta \mathrm{PF}_{\mathrm{ls}}$
$\Delta \mathrm{PF}_{\mathrm{ls}}$	Monthly, absolute (additive) change in the long-short-spread portfolio

Verdict

There is no statistical evidence to support the claim: *"Low Price-to-Earnings companies outperform high Price-to-Earnings companies"*.

4.7 Price-to-Revenue Ratio (Revenue Yield) Portfolios

This section presents the results for the hypothesis: *"Low Price-to-Revenue outperforms high Price-to-Revenue"* or, equivalently, *"High Revenue Yield outperforms low Revenue Yield."*

Note: When ensuring ranking consistency across several companies;

- Price-to-Revenue ratio has the domain: $\{\text{Revenue: Revenue} > 0, \text{Market Cap: } \mathbb{R}\}$
- Revenue Yield has the domain: $\{\text{Revenue: } \mathbb{R}, \text{Market Cap: Market Cap} > 0\}$

Hence, the investment hypothesis is investigated in terms of Revenue Yield (RY), i.e. the reciprocal of the Price-to-Revenue ratio. For more information, see Appendix D.

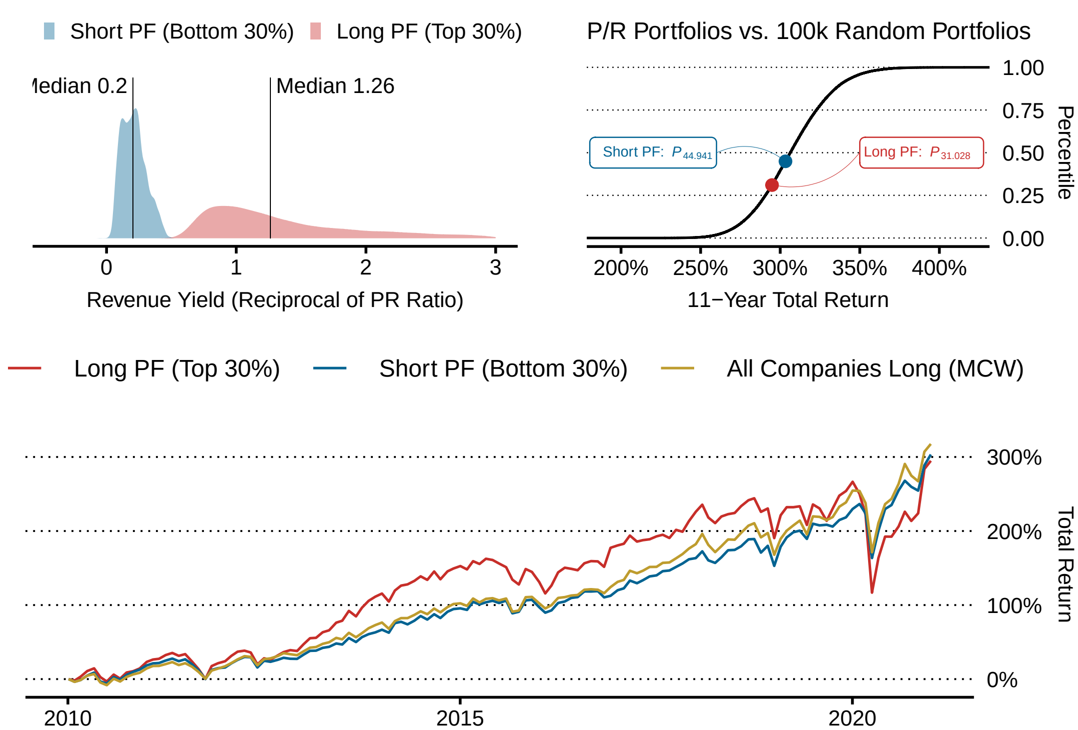
$$\mathrm{PR}^{-1} = \mathrm{RY} = \frac{\sum_{j=0}^{3} \text{Revenue}_{i,q-j}}{\text{Market Cap}_{i,t}}$$

The portfolios under investigation are constructed as follows:

Long Portfolio: 30% of S&P 500® companies having the highest Earnings Yield
Short Portfolio: 30% of S&P 500® of companies having the lowest Earnings Yield

The long and short portfolios' company holdings are updated on the first U.S. trading day of each month from January 2010 to December 2020. For details, see Methodology.

Figure 10: Revenue Yield Portfolios Results



Revenue Yield Portfolios Versus Random Portfolios

The upper right plot in figure 10 compares the long and short portfolios' performances to the estimated population of possible returns.

- The long portfolio (high RY) ranks in the 31.03^{th} percentile. Since $P_{31.03} \not> P_{90}$, the high RY portfolio does not significantly outperform portfolios drawn at random;
- The short portfolio (low RY) ranks in the 44.94^{th} percentile. Since $P_{44.94} \not< P_{10}$, the low RY portfolio does not significantly underperform portfolios drawn at random.

For details, see the Statistical Evaluation-section.

Portfolio Spread Evaluation

There is no statistical evidence that the expected change in the long/short spread portfolio is positive.

$$\mathbb{P}\left(\overline{\Delta\mathrm{PF}_{\mathrm{ls}}} = -0.001, \mathrm{SD}_{\Delta\mathrm{PF}_{\mathrm{ls}}} = 0.075 \middle| H_0 : \mathbb{E}[\Delta\mathrm{PF}_{\mathrm{ls}}] \leq 0\right) = 0.539$$

Where:

$\overline{\Delta\mathrm{PF}_{\mathrm{ls}}}$	The mean value of $\Delta\mathrm{PF}_{\mathrm{ls}}$
$\mathrm{SD}_{\Delta\mathrm{PF}_{\mathrm{ls}}}$	The sample standard deviation of $\Delta\mathrm{PF}_{\mathrm{ls}}$
$\Delta\mathrm{PF}_{\mathrm{ls}}$	Monthly, absolute (additive) change in the long-short-spread portfolio

Verdict

There is no statistical evidence to support the claim: *"Low Price-to-Revenue companies outperform high Price-to-Revenue companies"*.

4.8 Price-to-Cash Flow Ratio (Cash Flow Yield) Portfolios

This section presents the results for the hypothesis: *"Low Price-to-Cash Flow outperforms high Price-to-Cash Flow"* or, equivalently, *"High Cash Flow Yield outperforms low CF Yield."*

Note: When ensuring ranking consistency across several companies;

- Price-to-Cash Flow ratio has the domain: $\{\text{Free CF: Free CF} > 0, \text{Market Cap: } \mathbb{R}\}$
- Cash Flow Yield has the domain: $\{\text{Free CF: } \mathbb{R}, \text{Market Cap: Market Cap} > 0\}$

Hence, the investment hypothesis is investigated in terms of Cash Flow Yield (CFY), i.e. the reciprocal of the Price-to-Cash Flow ratio. For more information, see Appendix D.

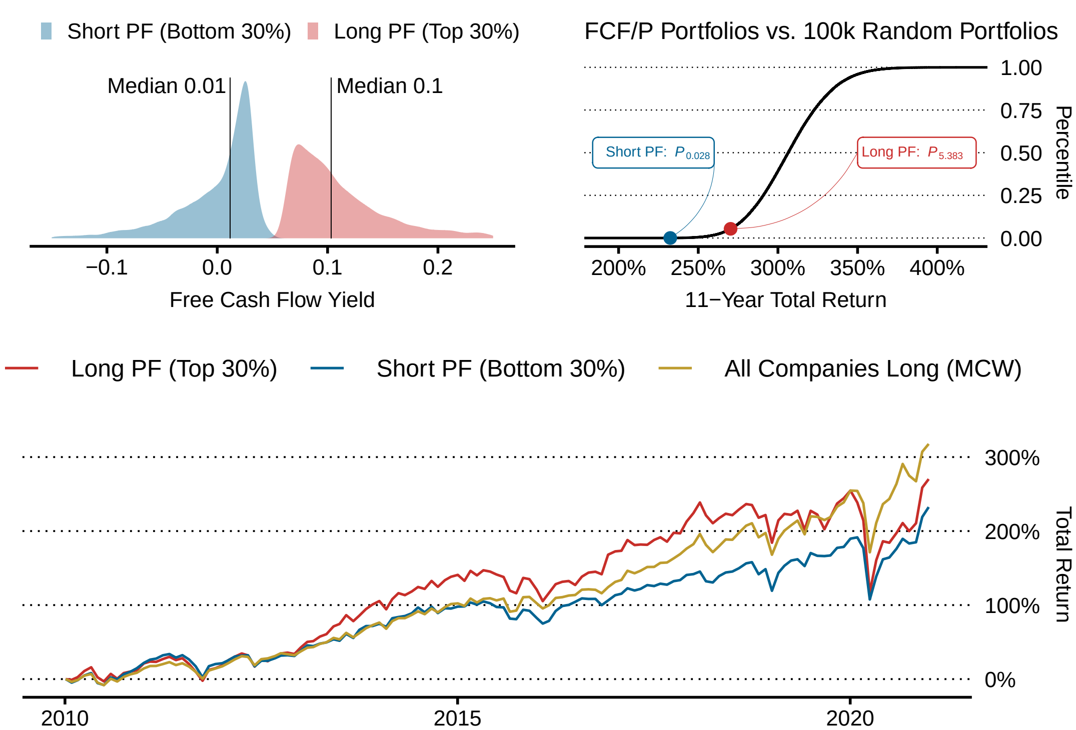
$$\text{PCF}^{-1} = \text{CFY} = \frac{\sum_{j=0}^{3} \text{Free Cash Flow}_{i,q-j}}{\text{Market Cap}_{i,t}}$$

The portfolios under investigation are constructed as follows:

Long Portfolio:	30% of S&P 500® companies having the highest Cash Flow Yield
Short Portfolio:	30% of S&P 500® of companies having the lowest Cash Flow Yield

The long and short portfolios' company holdings are updated on the first U.S. trading day of each month from January 2010 to December 2020. For details, see Methodology.

Figure 11: Cash Flow Yield Portfolios Results



Cash Flow Yield Portfolios Versus Random Portfolios

The upper right plot in figure 11 compares the long and short portfolios' performances to the estimated population of possible returns.

- The long portfolio (high CFY) ranks in the 5.38^{th} percentile. Since $P_{5.38} \not> P_{90}$, the high CFY portfolio does not significantly outperform portfolios drawn at random;
- The short portfolio (low CFY) ranks in the 0.03^{th} percentile. Since $P_{0.03} < P_{10}$, the low CFY portfolio **significantly underperforms** portfolios drawn at random.

For details, see the Statistical Evaluation-section.

Portfolio Spread Evaluation

There is no statistical evidence that the expected change in the long/short spread portfolio is positive.

$$\mathbb{P}\left(\overline{\Delta \mathrm{PF}_{\mathrm{ls}}} = 0.003, \mathrm{SD}_{\Delta \mathrm{PF}_{\mathrm{ls}}} = 0.062 \middle| H_0 : \mathbb{E}[\Delta \mathrm{PF}_{\mathrm{ls}}] \leq 0\right) = 0.299$$

Where:

$\overline{\Delta \mathrm{PF}_{\mathrm{ls}}}$	The mean value of $\Delta \mathrm{PF}_{\mathrm{ls}}$
$\mathrm{SD}_{\Delta \mathrm{PF}_{\mathrm{ls}}}$	The sample standard deviation of $\Delta \mathrm{PF}_{\mathrm{ls}}$
$\Delta \mathrm{PF}_{\mathrm{ls}}$	Monthly, absolute (additive) change in the long-short-spread portfolio

Verdict

The short portfolio (low Cash Flow Yield) **significantly underperforms** portfolios drawn at random. However, there is no statistical evidence to support the claim: *"Low Price-to-Cash Flow companies outperform high Price-to-Cash Flow companies"*.

4.9 GrowthSpread® Portfolios

This section presents the results for the hypothesis: *"High GrowthSpread® companies outperform low GrowthSpread® companies."*

Sharelab Limited developed GrowthSpread® (GS) in 2018, and obtained the trademark in early 2019. Sharelab Limited describes GS as "fundamental fair-value indicator". The computation of GS is performed by Sharelab Limited, using only information that had been publicly available at each point in time. GS is computed without estimation and does not bear any model risk.

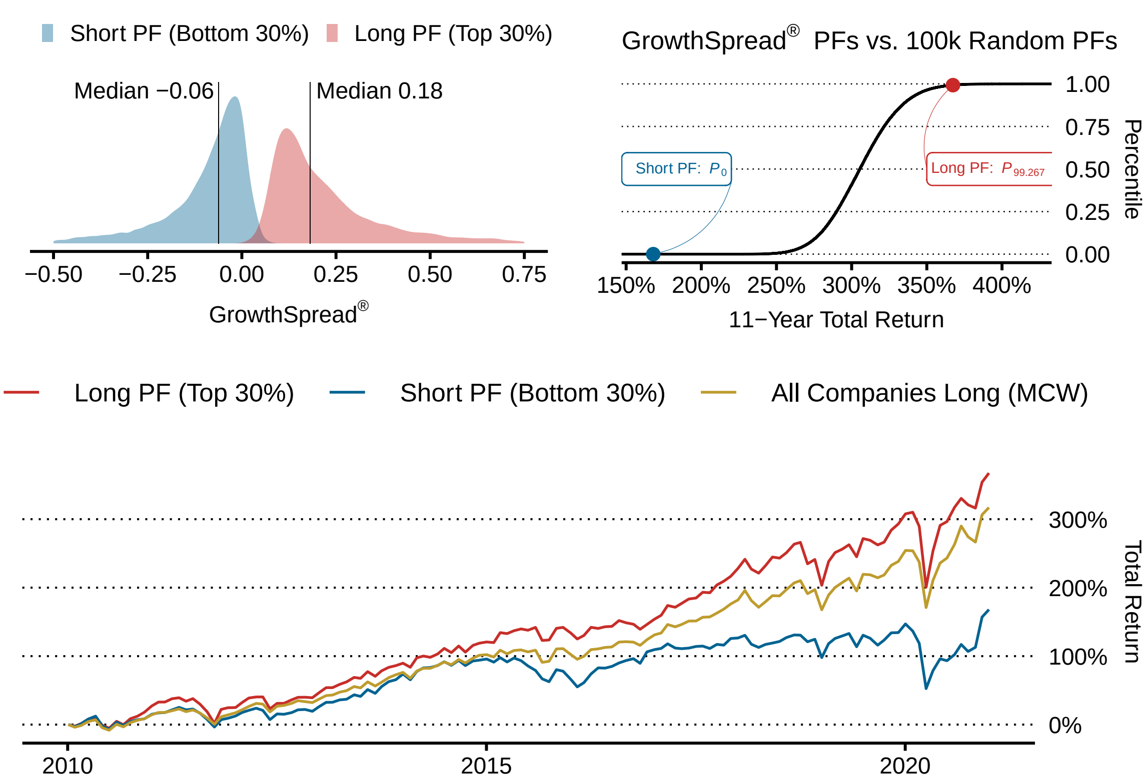
A positive value of GS indicates undervalued stock. Negative GS indicates overvalued stock. Hence, long and short portfolios are created using *descending* values of GS.

The portfolios under investigation are constructed as follows:

Long Portfolio:	30% of S&P 500® companies having the highest GrowthSpread®
Short Portfolio:	30% of S&P 500® of companies having the lowest GrowthSpread®

The long and short portfolios' company holdings are updated on the first U.S. trading day of each month from January 2010 to December 2020. For details, see Methodology.

Figure 12: GrowthSpread® Portfolios Results



GrowthSpread® Portfolios Versus Random Portfolios

The upper right plot in figure 12 compares the long and short portfolios' performances to the estimated population of possible returns.

- The long portfolio (high GS) ranks in the 99.27^{th} percentile. Since $P_{99.27} > P_{90}$, the high GS portfolio **significantly outperforms** portfolios drawn at random;
- The short portfolio (low GS) ranks in the 0^{th} percentile. Since $P_0 < P_{10}$, the low GS portfolio **significantly underperforms** portfolios drawn at random.

For details, see the Statistical Evaluation-section.

Portfolio Spread Evaluation

The expected change in the long/short spread portfolio is **significantly positive**.

$$\mathbb{P}\left(\overline{\Delta \mathrm{PF}_{\mathrm{ls}}} = 0.015, \mathrm{SD}_{\Delta \mathrm{PF}_{\mathrm{ls}}} = 0.061 \middle| H_0 : \mathbb{E}[\Delta \mathrm{PF}_{\mathrm{ls}}] \leq 0\right) = 0.003$$

Where:

$\overline{\Delta \mathrm{PF}_{\mathrm{ls}}}$	The mean value of $\Delta \mathrm{PF}_{\mathrm{ls}}$
$\mathrm{SD}_{\Delta \mathrm{PF}_{\mathrm{ls}}}$	The sample standard deviation of $\Delta \mathrm{PF}_{\mathrm{ls}}$
$\Delta \mathrm{PF}_{\mathrm{ls}}$	Monthly, absolute (additive) change in the long-short-spread portfolio

Verdict

There is **significant statistical evidence** to support the claim: *"High GrowthSpread® companies outperform low GrowthSpread® companies"*. In addition, the short portfolio (low GS) **significantly underperforms** portfolios drawn at random, and the long portfolio (high GS) **significantly outperforms** portfolios drawn at random.

5 Conclusion

This results of the statistical evaluation of all investigated investment beliefs are summarized in table 4. This research found credible statistical evidence in support of the *"high GrowthSpread® outperforms low GrowthSpread®"*-hypothesis. There is no statistical evidence in favor of any of the remaining eight hypotheses under investigation.

When it comes to one-directional portfolios, I found evidence in support of the following statements:

1. The high GrowthSpread® portfolio significantly *outperforms* portfolios drawn at random;
2. The low GrowthSpread® portfolio significantly *underperforms* portfolios drawn at random;
3. The low Price-to-Cash Flow portfolio significantly *underperforms* portfolios drawn at random;
4. The low Profitability portfolio significantly *underperforms* portfolios drawn at random;
5. The low Beta portfolio significantly *underperforms* portfolios drawn at random.

Noteworthy, the small Size portfolio (i.e. low Market Capitalization) ranked in the 88.55^{th} percentile in comparison to 100,000 portfolios drawn it random. The cut-off point for statistical significance is set at $P_x \geq P_{90}$. Yet, out of 100 portfolios drawn at random, the low Market Capitalization portfolio, on average, *outperforms* 88 portfolios.

On a side note, the investment belief *"high Value outperforms low Value"* had been statistically significant for the contrarian hypothesis. Low Value companies are typically synonymous with 'Growth' companies. Thus, Growth companies are found to outperform high Value companies.

Table 4: Statistical Results for Hypothesized Investment Beliefs

Investment Hypothesis	Statistical Evaluation: Percentile Rank: Long PF*	Statistical Evaluation: Percentile Rank: Short PF*	Statistical Evaluation: $H_1 : \mathbb{E}[\Delta \text{PF}_{\text{ls}}] > 0$: P-Value†	Hypothesis Verdict‡
Capital Asset Pricing Model				
High Beta outperforms low Beta	31.79^{th}	$\mathbf{6.26^{th}}$	0.426	false
Fama & French Firm Characteristics				
Small outperforms big	88.55^{th}	47.05^{th}	0.353	false
High Value outperforms low Value	0.05^{th}	99.99^{th}	0.989	false§
High Profitability outperforms low Profitability	57^{th}	$\mathbf{4.37^{th}}$	0.144	false
Low investment outperforms high investment	14.73^{th}	76.19^{th}	0.819	false
Common Price Multiples				
Low Price-to-Earnings (PE) outperforms high PE	0.11^{th}	89.55^{th}	0.926	false§
Low Price-to-Revenue (PR) outperforms high PR	31.03^{th}	44.94^{th}	0.539	false
Low Price-to-Cash Flow (PCF) outperforms high PCF	5.38^{th}	$\mathbf{0.03^{th}}$	0.299	false
Sharelab Limited				
High GrowthSpread® (GS) outperforms low GS	$\mathbf{99.27^{th}}$	$\mathbf{0^{th}}$	**0.003**	**true**

* Percentile of the portfolio's total return versus the estimated population of possible returns, see Statistical Evaluation

† For p-value ≤ 0.05, the expected change in the long-short spread portfolio is significantly positive

‡ true: Significant statistical evidence in support of the investment hypothesis. For details, see Success Measures

§ The contrarian hypothesis would have been statistically significant

6 Limitations and Outlook

Conceptionally and intentionally, this research is limited to S&P 500® companies from 1/2010–12/2020. It is unknown, whether the results of this study are reproducible in other geographic equity markets and across different time periods.

All investigated company characteristics are computed using historical information only. For example, the Price-to-Revenue estimate is based on the most-recently observed, trailing twelve months Revenue. It might be beneficial to use the analysts' consensus on the future Revenue estimate instead. Using such forward-looking estimates instead of historical estimates might generally improve the performance of the investigated investment beliefs. Adequate datasets of historical consensus estimates are available to institutional investors.

When it comes to the investment belief derived by the Capital Asset Pricing Model (CAPM), the Beta estimate is based on 252 daily returns. This Beta estimate might be a poor estimate for a company's future Beta. Using high-frequency data e.g. at the 10-minute interval should reasonably reduce the variance of the Beta estimate. Furthermore, methods such as heterogeneous autoregression (HAR) could be applied to model a forward-looking Beta estimate.

Finally, empirical research generally imposes limitations. These are not study-specific and can be summarized by:

> "The Only Constant in Life Is Change"
>
> — *Heraclitus*, ~ 500 B.C.

More specifically, due to an ever growing information flow and continuing advances in technology to make use of such information, market participants should be able to make better-informed investment decisions as time goes by. Empirical research, which is inevitably based on past data, fixes the level of technology and effective information flow to one specific period in time. Thus, whenever a market anomaly is published, and under the assumption that markets are efficient enough to absorb such information, any such finding might not be reproducible in time periods that follow.

To mitigate this general problem of empirical research, any empirical anomaly should be re-investigated frequently and across different markets, to ensure that such belief is robust across different time periods and market conditions.

Appendix A: Estimating Beta Efficiently in Large Datasets

Assuming that for the analysis period 2010–2020 the return of the risk-free asset is zero, the CAPM reduces to:

$$E[r_i] = \beta_i \, E[r_m] \tag{3}$$

To estimate a company's β, we can use past returns of the company's stock, $r_{i,t}$, and the stock market, $r_{m,t}$. Typically, the estimation is based on a rolling window of 252 most-recent and consecutive, daily returns (one calendar year). Going forward, daily returns of the S&P 500® index (symbol: ˆGSPC) are used to proxy the general stock market's return, $r_{m,t}$. The estimation of β for company i is then found by the univariate regression model:

$$r_{i,t} = \hat{\beta}_i \, r_{m,t} \tag{4}$$

Where:

$r_{i,t}$	Return of company i on day t
$r_{m,t}$	Return of the S&P 500® index on day t
$\hat{\beta}_i$	Beta estimate for company i

For large datasets, estimating $\hat{\beta}$ using pre-built functions to fit regression models (e.g. `lm` in *R* or `sklearn.linear_model` in *Python*) is computationally costly. To radically improve computational efficiency, it is sensible to look into the specifics of ordinary least squares (OLS) estimation. Under OLS, $\hat{\beta}_i$ in equation (3) is found by *minimizing* the error term ϵ_i:

$$\epsilon_{i,t} = \sum_{j=0}^{252} (r_{i,t-j} - \hat{\beta}_i r_{m,t-j})^2 \tag{5}$$

Note that $\epsilon_{i,t}$ in (4) is minimized by setting its first derivative with respect to $\hat{\beta}_i$ to zero:

$$\begin{aligned}
\frac{\partial \epsilon_i}{\partial \hat{\beta}_i} &= -2 \sum (r_i - \hat{\beta}_i r_m) r_m \\
0 &= -2 \sum (r_i r_m - \hat{\beta}_i r_m^2) \\
0 &= -2 \sum r_i r_m + 2 \sum \hat{\beta}_i \cdot r_m^2 \\
2 \sum \hat{\beta}_i r_m^2 &= 2 \sum r_i r_m \\
\hat{\beta}_i &= \frac{\sum r_i r_m}{\sum r_m^2} \qquad (6)
\end{aligned}$$

Hence, to estimate β_i in equation (3), one can simply divide the sum of the product between the company's stock return and the market's return by the sum of squared market returns.

Computing (5) efficiently on a rolling basis of 252 observations for hundreds of different companies and decades of daily returns, we can use the *R* libraries `data.table` and `RcppRoll` to vectorize the $\hat{\beta}_i$ computation.

```r
# laod libraries
library(data.table)
library(RcppRoll)

# compute the daily return product
data[, product_return := return_adj_div * return_sp500]

# compute the daily squared return of the S&P500 index
data[, sq_return_sp500 := return_sp500**2]

# for each company ticker, compute the 252-observations rolling sum of the return product
data[, product_return_rolling_sum := c(rep(NA,251), roll_sum(product_return,n=252)), ticker]

# for each company ticker, compute the 252-observations rolling sum of the squared market's return
data[, sq_return_sp500_rolling_sum := c(rep(NA,251), roll_sum(sq_return_sp500,n=252)), ticker]

# compute beta for all companies and all time periods
sep_data[, beta_1yr := product_return_rolling_sum/sq_return_sp500_rolling_sum]
```

Appendix B: Fama and French Factors Computation

The factor construction example provided in the Fama and French Model chapter simplifies and roughly approximates the exact computation of the original Fama and French factors. This section provides a more accurate description.

HML, RMW, and CMA Factors

Prior to ordering companies by a specific firm characteristic, Fama and French split the company universe into 'Small' and 'Big' companies. The ordering in terms of a firm characteristic takes place within each of these two super-portfolios. This results in two return spreads, and the resulting 'factor' is the average of these. More specifically, the construction of the Value, Profitability, and Investment factors follows the following process:

1. Choose a 'region' of public companies, e.g. North American stocks
2. By the end of June of each year, order the company universe by *ascending* market capitalization
3. Form a 'Small' portfolio from the 10% top-ranking companies and a 'Big' portfolio from the 10% bottom-ranking companies
4. Reorder the companies within each portfolio by a firm characteristic of interest:
   - High Minus Low (HML): Descending values of book-to-market ratio
   - Robust Minus Weak (RMW): Descending values of EBT-to-book ratio[9]
   - Conservative Minus Aggressive (CMA): Ascending values of YoY total assets growth
5. For each Small and Big portfolio, form 'long' and 'short' sub-portfolios:
   - $\text{pf}_{\text{Small}}^{long}$: 30% of top-ranking companies from the Small portfolio
   - $\text{pf}_{\text{Small}}^{short}$: 30% of bottom-ranking companies from the Small portfolio
   - $\text{pf}_{\text{Big}}^{long}$: 30% of top-ranking companies from the Big portfolio
   - $\text{pf}_{\text{Big}}^{short}$: 30% of bottom-ranking companies from the Big portfolio
6. Compute the average of the return spreads between the 'long' and 'short' portfolios:

$$`HML`,`RMW`,`CMA` = \frac{(\text{pf}_{\text{Small}}^{long} - \text{pf}_{\text{Small}}^{short}) + (\text{pf}_{\text{Big}}^{long} - \text{pf}_{\text{Big}}^{short})}{2} \tag{7}$$

[9] Approximate formula

Small Minus Big (SMB) Factor

When it comes to the computation of the Size-factor, i.e. Small Minus Big (SML), 18 portfolios are involved. These are comprised of 6 sub-portfolios per each Value, Profitability, and Investment characteristic. For further details, please visit the website of Kenneth R. French.

Appendix C: Explicit and Implicit Assumptions

All identified explicit and implicit assumptions are disclosed below.

Long and Short Portfolios

On each monthly portfolio reconstruction day,

1. All companies are assumed to be bought and sold at the "close" price of that date;
2. The portfolio's total liquidation value is invested in equal amounts into each company.

Furthermore,

3. Dividends (if any), accrue over the monthly holding period and are reinvested evenly into the next months' portfolio holdings.

Note that points (2) and (3) implicitly assume fractional investments. For example, if a fictional portfolio's value is USD 100, and the two companies to invest in trade at USD 20 and USD 60, respectively, it is assumed that 2.5 shares of company A, and 5/6 shares of company B are bought. This results in equal investment of USD 50 into each company. For net asset values of at least USD 5m, this restriction should be negligible in practice.

Point-in-Time Information

All firm characteristics (e.g. Book-to-Market ratio) are estimated using past information that had been publicly available up to the closing time of the day a portfolio's holdings are updated.

For example, if the next portfolio cycle starts at market close on 1. May 2010, then, at maximum, the indicator includes information until this time/date. All information used had been published at this time. This ensures that there is no introduction of look-ahead bias.

CAPM-Beta Estimation

When it comes to the long and short Beta portfolios, a company's Beta is estimated on a rolling basis using 252 past, daily returns. This reflects a calendar year's worth of daily data. The underlying assumptions of the Beta estimates are:

1. Daily returns of the S&P 500® index are assumed as the market's return
2. The daily return of the risk-free asset is assumed to be zero

The programming code used to estimate a company's Beta is disclosed in the appendix.

Missing Values

While there are generally 500 companies in the S&P 500® index, not *all* companies might have information required to compute the firm characteristic necessary for investigation of a specific asset pricing hypothesis. For consistency and to maintain comparability, these companies are excluded from the selection window for all:

1. The hypothesis' long and short portfolios;
2. The 100,000 simulated portfolios, each picking 30% of *available* stocks uniformly at random;
3. The market-cap-weighted, all-companies-long benchmark

On the first trading day of each month, all S&P 500® constituents are re-checked for the data availability of the company characteristic in question. As soon as a company has sufficient public information to compute the characteristic, it is included into the analysis.

For example, Facebook Inc ($FB) joined the S&P 500® index on 20 December 2013. On the first trading day of January 2014, $FB had less than 4 published quarterly Earnings reports. Hence, the company characteristic Price-to-Earnings (PE) ratio could not be computed on that day. Thus, for the month January 2014, $FB is excluded from the analysis of the *"Low Price-to-Earnings outperforms high Price-to-Earnings"*-hypothesis. Subject to data availability, Facebook is added to the analysis of this hypothesis later in 2014.

Appendix D: Price Multiples Transformation

When it comes to the investment hypotheses regarding common price multiples, the Price-to-Earnings (PE), Price-to-Revenue (PR), and Price-to-Cash Flow (PCF) ratios are typically computed as follows.

$$\mathrm{PE}_{i,t} = \frac{\text{Market Cap}_{i,t}}{\sum_{j=0}^{3} \text{Net Income}_{i,q-j}}$$

$$\mathrm{PR}_{i,t} = \frac{\text{Market Cap}_{i,t}}{\sum_{j=0}^{3} \text{Revenue}_{i,q-j}}$$

$$\mathrm{PCF}_{i,t} = \frac{\text{Market Cap}_{i,t}}{\sum_{j=0}^{3} \text{Free Cash Flow}_{i,q-j}}$$

Where:

t Daily point in time (date index)
q Publication date of most-recently published quarterly report, where $q \leq t$
i Company i, where $i \in$ S&P 500® Consituents

Note that Net Income, Free Cash Flow and even Revenue can be zero or negative. In large datasets with many companies, this creates two problems:

1. PE, PR and PCF are not defined whenever the denominator is *zero*
2. The ranking of companies is inconsistent whenever the denominator is *negative*

To illustrate (2), let's consider ranking many companies in terms of share price and Earnings. Suppose there are four companies A, B, C, and D, *each* trading at a share price of $100. Further assume that company A has Earnings of $20, company B has Earnings of $10, company C has Earnings of -$10 and company D has Earnings of $-20. Clearly, the ranking order of these companies should be A, B, C, D. However, according to ascending values of PE ratio, the ranking order would be D, C, A, B.

To solve these problems, the *reciprocal* of each measure is used. This ensures that for all cases of positive, negative, or zero Earnings, Cash Flow, or Revenue, the ranking among companies remains consistent. I am referring to these transformations as Earnings Yield (EY), Revenue Yield (RY) and Cash Flow Yield (CFY), respectively.

$$\text{EY} = \text{PE}^{-1} = \frac{\sum_{j=0}^{3} \text{Net Income}_{i,q-j}}{\text{Market Cap}_{i,t}} \tag{8}$$

$$\text{RY} = \text{PR}^{-1} = \frac{\sum_{j=0}^{3} \text{Revenue}_{i,q-j}}{\text{Market Cap}_{i,t}} \tag{9}$$

$$\text{CFY} = \text{PCF}^{-1} = \frac{\sum_{j=0}^{3} \text{Free Cash Flow}_{i,q-j}}{\text{Market Cap}_{i,t}} \tag{10}$$

Consequently, the companies are ordered by *descending* values of EY, RY, and CFY.

References

Achim Zeileis. 2021. *Package "Zoo"*. https://cran.r-project.org/web/packages/zoo/.

Arnold, J. B. et al. 2019. *Package "Ggthemes"*. https://cran.r-project.org/web/packages/ggthemes/ggthemes.pdf.

Auguie, B. 2017. *Package "gridExtra"*. https://cran.r-project.org/web/packages/gridExtra/gridExtra.pdf.

Dowle, M. et al. 2020. *Package "Data.table"*. https://cran.r-project.org/web/packages/data.table/data.table.pdf.

E. F. Fama and K. R. French. 1993. *Common Risk Factors in the Returns on Stocks and Bonds*. Journal of Financial Economics 33. https://rady.ucsd.edu/faculty/directory/valkanov/pub/classes/mfe/docs/fama_french_jfe_1993.pdf.

Hadley Wickham. 2019. *Package "Stringr"*. https://cran.r-project.org/web/packages/stringr/.

Hadley Wickham, Dana Seidel, RStudio. 2020. *Package "Scales"*. https://cran.r-project.org/web/packages/scales/.

Hao Zhu. 2021. *Package "kableExtra"*. https://cran.r-project.org/web/packages/kableExtra/.

Henrik Bengtsson. 2021. *Package "matrixStats"*. https://cran.r-project.org/web/packages/matrixStats/.

Irizarry, Rafael A. 2020. *Introduction to Data Science*. https://rafalab.github.io/dsbook/.

Jarek Tuszynski. 2021. *Package "caTools"*. https://cran.r-project.org/web/packages/caTools/.

Jeffrey A. Ryan. 2020. *Package "Xts"*. https://cran.r-project.org/web/packages/xts/.

Kevin Ushey. 2018. *Package "Rcpproll"*. https://cran.r-project.org/web/packages/RcppRoll/.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org.

Sharelab. 2021. A platform for US equity analytics: SHARELAB LIMITED. https://www.sharelab.com.

Simon Urbanek. 2016. *Package "Fasttime"*. https://cran.r-project.org/web/packages/fasttime/.

Spinu, V. 2020. *Package "Lubridate"*. https://cran.r-project.org/web/packages/lubridate/lubridate.pdf.

Wickham, H. 2019. *Package "Tidyverse"*. https://cran.r-project.org/web/packages/tidyverse/tidyverse.pdf.

William F. Sharpe. 1964. *Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk*. The Journal of Finance Volume 19, Issue 3. https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.1964.tb02865.x.

Winston Chang. 2014. *Package "Extrafont"*. https://cran.r-project.org/web/packages/extrafont/.

Xie, Y. et al. 2020. *Package "Knitr"*. https://cran.r-project.org/web/packages/knitr/knitr.pdf.