P-Hacking in Asset Pricing;

Portfolio construction techniques in the spotlight

ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Abstract:

Researchers face many decisions when constructing portfolios. These degrees of freedom in choice may lead to p-hacking. This paper identifies which portfolio construction techniques have been used through time, the options that were presented, and how they used them in their work. The foundation of potential future guidelines is written and provides simple suggestions on how to avoid potential p-hacking with portfolio construction in the entire empirical asset pricing field

Master Thesis Financial Economics

Name student: Wouter Lapré Student ID number: 412874wl

Supervisors: drs. A. Soebhag, drs. B. van Vliet

Second assessor: prof dr. P. Verwijmeren

Date final version: 23 October 2022

Table of contents

1 Introduction

In 2017, Campbell R. Harvey, the President of the American Finance Association, published an article in the Journal of Finance (Harvey, 2017). Not very peculiar for a highly decorated researcher and President of the organization that owns the Journal of Finance to let his article be published. However, this paper was different. It gave an overview of the current state of empirical research in financial economics. Essentially, it stated critical views on his very own organization and the entire field of research. Harvey (2017) opens the article with an empirical example to show where the problems lie. In short, he gave his research assistant the instructions to find a new factor based on equity returns from CRSP data. The instructions were: "(1) form portfolios based on the first, second, and third letters of the ticker symbol; (2) show results for 1926 to present and 1963 to present; (3) use a monthly, not daily, frequency; (4) rebalance portfolios monthly and once a year; (5) value weight and equally weight portfolios; (6) make a choice on delisting returns; and (7) find me the best long-short portfolio based on the maximum t-statistic" (Harvey, 2017). This would lead to 25.280 possible long-short portfolios based on these choices. Harvey (2017) even argues in the article that several researchers would present more portfolio choices for the research assistant for various reasons. To clarify why this can cause implications, a recent example in statistics will be given. When people massively started COVID-19 testing, several users received false positive or negative tests (Bentley, 2021). Public opinion on the trials started fading into distrust. It started questioning the reliability of the test. Mina and Andersen (2020) stated that, indeed, loss in sensitivity of individual tests, within reason, can be compensated for by frequency of testing and wider dissemination of tests. In addition, public health messaging should ensure appropriate expectations of screening, particularly around sensitivity and specificity, so that false negatives and false positives do not erode public trust. After many media appearances by doctors and medical researchers who explained the tests are not 100% reliable. Luckily, public trust was not lost. They achieved this by explaining the fundamental statistical theory. Various factors could lead to false negatives or positive tests. Mayers and Baker (2020) listed possible causes of false negatives and negatives for the COVID-19 tests. This list leads up to nine probable causes of false positive or negative results. This shows that many factors, caused by human interference or random effect, can influence results. The factors that influence COVID-19 testing are very different from factors in Financial Economics. In asset pricing, researchers are on the hunt for those factors that can explain the cross-section of expected stock returns. This has produced hundreds of potential factors. Papers that have investigated this matter are (Cochrane, 2011), (Harvey & Liu, 2015), (McLean & Pontiff, 2016), and (Feng, Giglio, & Xiu, 2020) and they all agreed on the same thing. The factor "zoo" should be disciplined and guidelines have to be created. This is a fundamental task that has to be faced in today's asset pricing literature. Mainly, the current challenge is to distinguish functional factors from factors that are only positive due to data mining or design choices. Choosing which factor and how it will be presented in the paper may lead to p-hacking. Harvey (2017) recognizes this and states that many empirical design choices may be crucial for the results. This, combined with an insufficient replication culture, leads to many false positives within the field of financial economics.

In this paper, the empirical design choices within asset pricing literature are investigated and documented. We looked at frequently used portfolio construction techniques and recorded which decision was made. The portfolio construction techniques are based on asset pricing theory. We will state in more detail in Section II why these portfolio construction techniques are used and how they were developed since the start of asset pricing literature. These techniques apply to any data restrictions, design choices on sorting methods, and how sorting factors were calculated. For example, when portfolios are constructed, in almost all asset pricing papers, they use value or equal weighting. When a paper uses value weighting (hereafter VW), a “1” was documented. For equal weighting (hereafter EW), we write down a “0”. When they use alternative weighting or the weighting method was unidentified, we record it in the commentary section. We gratefully use the data set provided by Harvey and Liu (2019) to explore our research question. Based on a representative set of 323 empirical asset pricing studies, eleven different portfolio construction decisions were investigated and analyzed based on what researchers have to face. These choices are (1) 30/70 or 20/80 breakpoints, (2) NYSE or NYSE-AMEX-Nasdaq (NAN) breakpoints, (3) including or excluding firms with a negative book equity value, (4) including or excluding microcaps, (5) a price filter is used or not, (6) including or excluding utility firms, (7) including or excluding financial firms, (8) industry neutralization or not, (9) value-weighting or equal-weighting, (10) independent or dependent sorts, and (11) sorting on the market capitalization based on June or the most recent. Factors can be constructed using two possible decisions for each portfolio construction technique, which leads to 2048 (211) portfolio construction combinations. With this in mind, this paper focuses on the following research questions: Which portfolio construction techniques are used by researchers and why? The results identify the options presented to researchers and show if they are preferred options if the decisions are made randomly. There are options that have a higher probability of being chosen than others but nevertheless are there no 100% preferred options. The relationship between combinations of decisions is not very strong and is even very weak in particular cases. The researchers who pioneered the particular decision and the researcher who used this approach the most recent often reframe from giving explanations on why they chose this option. This indicates clear incentives that guidelines should be created in order to compare models with each other. We find that excluding negative book equity and financial firms has a significant and considerable impact on the t-statistic of the model used by researchers in our sample.

This study gives explanations and information on the current situation with portfolio construction within asset-pricing literature and calls for other researchers to build on this theoretical framework. We encourage future researchers to use the insights presented in this paper for creating guidelines for portfolio construction in the asset-pricing literature. These recommendations are given for informational purposes. These guidelines can be the key to preventing p-hacking within asset pricing literature. As stated by Soebhag, van Vliet and Verwijmeren (2022) seemingly small differences in design can significantly impact the resultant portfolio's performance. section I is the introduction, and section II is the literature review. In section III the methodology is described. In section IV, the results are presented, and in section V the conclusion and section VI recommendations are stated.

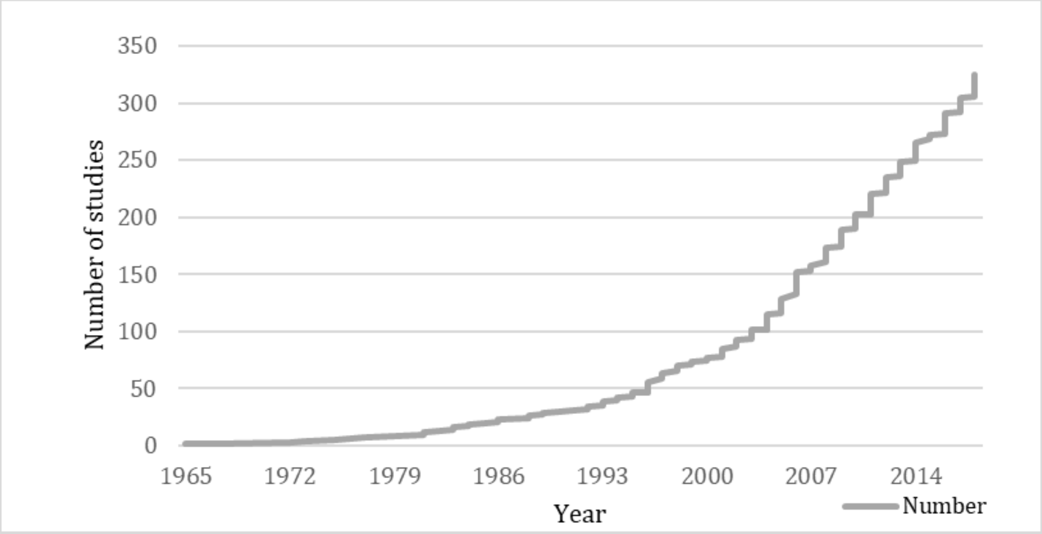
2 Empirical Review

So, what are the causes for the current situation in the asset pricing literature called the Factor Zoo? To understand this, we are going to look at fundamental asset pricing theory. Here, we observe and analyze the portfolio construction decisions made throughout the past and the present. Questions like: When were certain portfolio construction techniques developed and did they affect the papers written from that time onwards are asked to gain a clear insight into understanding the Factor Zoo. But first, we examine the field of asset pricing literature and how it was formed to the state it is in now.

As you can see in graph A, the number of papers in asset pricing literature that were published in top journals was exponentially increasing from 1964 onwards. This implies that more papers have been written and deemed important for editors to publish the paper in recent times. This means that for a researcher to get published in a top journal in the 2010s, it must compete with more papers for a spot compared to the 1970s. Competition drives efficiency and thus can be seen as a good aspect of the increasing number of papers written.

Graph 1

Number of studies from the asset pricing literature sample ranging from 1965 until 2018



Note. Number of papers N = 325

However, developments in the research field were investigated and were not promising. As stated in Harvey (2017) journals contribute to data mining through their focus on publishing papers with the most significant results. The reason is that journal editors often compete for citation-based impact numbers. Fanelli (2010) shows that papers with no significant results are less likely to be published. Sutton et al (2000) disclosed that it is well-documented that editors and reviewers are more likely to reject negative findings that do not support the hypothesis tested in the paper. These studies imply that researchers ought to seek significant results to have a higher probability of being published in one of the top journals. Researchers allocate their time and effort to potential studies where it is more likely to generate significant results.

This publication bias is thus kept alive by publishers, editors, and researchers.

To understand why the environment addressed by Harvey (2017) in asset pricing literature is a growing concern, this study is going to look at fundamental financial asset pricing theory. When acknowledging and describing the theory, one can realize why the current environment has developed into this.

It all starts with the model of portfolio choice developed by Markowitz (1959). The most impactful assumption in this model is that investors are risk-averse. When looking at their one-period investment return they only consider the variance and the mean. Here, the investor chooses the mean-variance efficient portfolio. This tells us that the investor chooses a portfolio that gives 1) the minimum variance, given an expected return 2) the maximum expected return, given the variance. Sharpe (1954) and Lintner (1965) created a testable prediction for the relation between expected return and risk based on the mean-variance model of Markowitz (1959). The CAPM was created and was constructed to identify an efficient portfolio if asset prices are to clear the market of all assets (Fama & French, 2004). Sharpe (1954) and Lintner (1965) included two different assumptions to the portfolio model of Markowitz (1959). Fama and French described the CAPM model as follows:

> The first assumption is a complete agreement given market-clearing asset prices at *t — 1*, investors agree on the joint distribution of asset returns from *t — 1* to *t*. And this distribution is the true one—that is, it is the distribution from which the returns we use to test the model are drawn. The second assumption is that there is borrowing and lending at a risk-free rate, which is the same for all investors and does not depend on the amount borrowed or lent. We can then see that all efficient portfolios are combinations of risk-free assets (either risk-free borrowing or lending) and a single risky tangency portfolio. Since all investors hold the same portfolio of risky assets, it must be the value-weight market portfolio of risky assets. Specifically, each risky asset's weight in the tangency portfolio, which we now call M (for the "market"), must be the

> total market value of all outstanding units of the asset divided by the total market value of all risky assets, in addition, the risk-free rate must be set (along with the prices of risky assets) to clear the market for risk-free borrowing and lending" (Fama & French, 2004, p. 34).

The main takeaway from the CAPM model is that the expected returns are only related to the risk relationship. The risk-return relationship gives us the opportunity to find the most efficient portfolio for maximum return as this is the only factor that explains return. One can imagine that the assumptions made cannot hold out in practice Assuming that you can unrestrictedly borrow and lend risk-free is not realistic. Roll (1977) investigated the model and come to the conclusion that the theory is not testable unless the exact composition of the true market portfolio is known and used in the tests. This implies that the theory is not testable unless all individual assets are included in the sample. As this is unrealistic it means that the CAPM model is only a theory and can't be used by practitioners. Still, the model is used to teach students the fundamentals of portfolio theory and asset pricing and is used in some models, e.g. to calculate the cost of equity capital in the Weighted average cost of capital calculation. The model is easy to use and therefore is still valid if the assumptions are, almost, met.

In the late 1970s, researchers began to uncover that return can be explained not only by Beta but by other factors. Variables like size, price momentum, and various price ratio's caused serious implications in the validation of the use of the CAPM. This led to the alternative approach to the CAPM by Ross (1976) called the Arbitrage Pricing Theory (APT). The APT is based on the intuition behind the CAPM. It is consistent with every other instruction for portfolio diversification but it doesn't rely on identifying the market portfolio as with the CAPM. As a matter of fact, no particular portfolio plays a role with the APT. The market portfolio doesn't have to be mean-variance efficient as one of the pillars of CAPM is built upon. Ross and Roll (1980) identified two major differences between the APT and the CAPM. First, and most simply, the APT allows more than just one generating factor. Second, the APT demonstrates that since any market equilibrium must be consistent with no arbitrage profits, every equilibrium will be characterized by a linear relationship between each asset's expected return and its return‘s response amplitudes, or loadings, on the common factors. Basically, this says, that we are confronted with the task of identifying the relevant factor structure instead of the true market portfolio. Shanken (1982) identified various factor structures but the APT lacks in explaining what the uniquely relevant factor structure is and if it even exists. Shanken (1982) therefore concluded that Ross's theory does not imply an exact linear risk-return relation. The factor model can be manipulated rather arbitrarily by repackaging a given set of securities. A new set of returns and a corresponding factor model can be produced, with virtually any prespecified random variables as the factors. By itself, therefore, factor analysis is not an adequate tool for identifying the random components of returns that should be relevant to asset pricing.

In the late 1980s and the beginning of the 1990s, there was a trend among researchers to declare anomalies. Patterns in average stock returns are not explained by the CAPM of Sharpe and Lintner and the APT of Ross. Examples are (De Bondt & Thaler, 1985) finding a reversal in long-term returns. Jegadeesh and Titman (1993) declare that stocks with higher returns in the previous twelve months tend to have higher future returns. Other researchers find different factors that influence the firm's stock return. Factors like size (ME, stock price times the number of shares), book-to-market- equity (BE/ME, the ratio of the book value of common equity to its market value), earnings/price (E/P), cash flow/price (C/P), and past sales growth. (Banz, 1981), (Basu, 1983), (Rosenberg, Reid, & Lanstein, 1985), (Lakonishok, Shleifer, & Vishny, 1994). Researchers were searching for explaining factors. What could cause the firm's abnormal stock return? Fama and French (1993) constructed the three-factor model and Fama and French (1996) say it explains that the anomalies are related. The model says that the expected return on a portfolio in excess of the risk-free rate [E(Ri) - Rf] is explained by the sensitivity of its return to three factors: (i) the excess return on a broad market portfolio (RM - Rf); (ii) the difference between the return on a portfolio of small stocks and the return on a portfolio of large stocks (SMB, small minus big); and (iii) the difference between the return on a portfolio of high-book-to-market stocks and the return on a portfolio of low-book-to-market stocks (HML, high minus low) (Fama & French, 1993). Fama and French aimed to answer any significant amount of variation in stock returns using a risk-based model where size and bookto-market factors can explain this variation. Whereas Fama and French (1996) say that almost all anomalies can be explained by their Fama and French (1993) model. Jegadeesh and Titman (1993) findings on the momentum effect couldn't be explained. So, the search continues for explaining factors. This results in the creation of Carhart's (1997) four-factor model. Carhart extends the three-factor model of Fama and French by adding a fourth risk factor. This riskfactor is capturing the momentum effect called WML. WML is the yield difference between the winner's stock portfolio and the loser's stock portfolio based on the anomaly Jagadeesh and Titman found in 1993. His method was not widely adopted by researchers although it offered better explanations as it captures the momentum effect. Still, researchers were not convinced and were searching for explaining factors. Novy-Marx (2013) and Titman et al (2004) found relations for variation in average returns in profitability and investment. This created evidence that the three-factor model by Fama and French (1993) didn't explain all variation in average returns. Fama and French (2015) therefore made the five-factor model by adding the profitability and investment effect to the existing three-factor model. These models created by Fama and French and Carhart were developed to theoretically and consistently explain the stock anomaly theory. As many researchers, (Özkan, 2020), (Huynh, 2018) and (Lin, 2017), found positive findings in using the model for their domestic markets, there is also still room for discussion. Dutta (2019) is skeptical about the ability of the FF5F model to detect longterm anomalies. According to Huynh (2018), three-factor and five-factor models did not pass the GRS test (Gibbons, Ross, and Shanken's test). This test indicates whether the model is a good asset pricing model. The emergence of price anomalies is not consistent with the efficient market hypothesis and, therefore, could explain these anomalies. The search continues, and researchers are keener to finding these explaining factors. As explained in the introduction, this can cause incentives for researchers but also for editors to publish significant results. Significant factors are controversial and generate discussions throughout the entire assetpricing world. As more and more papers are written and published, significant factors are harder to 'find.' Due to this, researchers have started altering methodological decisions. One of these altercations is how portfolios are constructed and techniques used.

Referring back to Fama and French (1993), this model was the beginning of a standard within asset pricing literature for portfolio construction that was to be divided into data restrictions and sorting techniques. For constructing their three-factor and five-factor models, Fama and French used the same method for portfolio construction. As a result, they set a precedent for other researchers to follow. However, there was still freedom of choice regarding portfolio construction decisions. This led to various decisions that researchers could deem debatable and is often unexplained.

As this paper focuses on portfolio construction decisions, Harvey et al (2016) showed that to identify the cross-section of expected returns, many studies base their statistical analysis on a single test. This opens the discussion from a methodological standpoint on whether this is valid or not. When a single hypothesis test is used a value α (statistical significance) stands for the chance that the factor is significant or not. This is the type I error that is accounted for. In a multiple testing framework, the chance of incorrect discoveries is high when only restricting each individual's test's type I error rate at α. The rationale behind it is that when many factors need to be tested, one of the factors will eventually be significant. In multiple hypothesis testing, the outcomes of the various individual tests need to be measured. They investigated 313 papers and found 316 different significant factors. To test all these factors, they suggest using multiple testing. There are three ways to deal with the bias associated with multiple testing. Out-of-sample validation, a statistical framework that allows for multiple testing and looking across multiple asset classes. Engelberg et al (2015) show that with the out-of-sample validation approach, 12 of the 97 anomalies tested were dropped because they couldn't replicate the in-sample performance. Asness et al (2013) followed the across multiple asset classes approach and find significant return premia to value and momentum in every asset class and strong comovement of their returns across asset classes, both of which challenge existing theories for their existence. Harvey et al (2016) elaborated on the statistical framework for multiple testing. The paper highlights that despite the rapid development of multiple testing methods, asset pricing literature didn't catch on. Most research focuses on the Bonferroni adjustment, which is deemed too strict. Soebhag et al (2022) find that under the Bonferroni adjustment, which is the significant level divided by the number of hypotheses, fewer discoveries are made than under the Holms adjustment which is more lenient. Harvey et al (2016) suggest Benjamini, Hochberg, and Yekutieli's adjustment besides the Bonferroni and Holm adjustment. To find more on the testing approaches methodology, see (Shanken, 1990), (Schweder & Spjøtvoll, 1982), (Benjamini & Hochberg, 1995). Harvey et al (2016) looks at various hypotheses under multiple hypothesis testing and find that 3-6 factors wouldn't be discovered, whereas under single hypothesis testing 10 factors were discovered.

Whether it is portfolio construction techniques or multiple hypothesis testing, decisions made during the research alter the outcome. To identify which portfolio construction techniques are used in our sample, we analyzed all 323 studies which are described in the next section, section III.

3 Portfolio construction techniques

To see which portfolio construction choices were made by the researchers, we analyzed and surveyed 325 empirical articles in top finance journals between 1965 and 2018. The 325 empirical articles are all from the field of asset pricing and provided by Harvey and Liu (2019). Where they used it for their census of the "Factor Zoo", in this paper the list is used to document the methodological choices made on portfolio sorting. In this section, which portfolio construction choices are identified, and how we documented these choices will be described. Hereafter, we describe why the choices are made by the researchers.

3.1 Identifying and documentation of the portfolio choices

After analyzing all the papers on the list, eleven different portfolio construction choices were often specified. This is based on examining the data and methodology sections of these asset pricing papers. A total set of 2048 choices can be made from the eleven different portfolio construction choices. This means 2048 different versions for each factor and thus 2048 different factor models. The methodology for documenting the choices will be explained in the next section.

3.1.1 Portfolio construction choices – Data restrictions

First, we looked at data restriction choices. Choices that have been made in asset pricing literature regarding data restrictions. As presented in Table 1, when the researcher decides to exclude firms that have a negative Book-to-Equity (BE) ratio, we mark the paper with a "1" in the Negative BE column. When the researcher includes firms with a negative BE ratio, we mark the paper with a 0 in the corresponding column. This also applies to the restriction choices

Price Filter, Micro, Financials, and Utilities. When an asset price filter is present with the purpose to exclude small firms, we mark a “1”. If not present, a “0” is marked. An example of a price filter is “We use only stocks with a price greater than $5 a share” (Jones & Lamont, 2002)

Table 1

Portfolio construction choices – Data restrictions

Choices	Included	Excluded
Negative BE	0	1
Price Filter	1	0
Micro	1	0
Financials	1	0
Utilities	1	0

For the Micro variable, we investigated whether researchers exclude small firms or in terms of market capitalization, ‘Micro Caps’. As in this paper, micro caps are defined as stocks that are smaller than the 20th percentile of market equity for NYSE stocks. This implies that when the price filter is present in the paper, Micro caps are excluded, and we will mark it with a 0. However, as stated in Donangelo (2014) “ We exclude the lowest 20th size quantile, that is, 5%, of the sample of firms to avoid anomalies driven by microcap firms, as discussed in Fama and French.” or in Titman et al (2004) “Its annual total net sales should be no less than U.S.$10 million to exclude firms at their early stage of development” implies the exclusion of small firms without the asset price filter present. In these examples, the Price filter is marked with a “0” and Micro with a “1”. If all stocks are included in the portfolio or there is no mention of excluding small firms/micro caps, then we mark the corresponding column with a “1”. Regarding the portfolio construction choices for Financials and Utilities, these imply that when a “1” or a “0” is documented in the field, Financial and/or Utility firms are included or excluded from the sample. An example showing these choices in the papers is as stated in Whited and Wu (2006), “We omit all firms 4900 and 4999 or between 6000 and 6999 regulated or financial firms”[1].

3.1.2 Portfolio construction choices – Sorting techniques and others

Secondly, sorting techniques are investigated, and the choices made when sorting portfolios have been documented. As mentioned in Section II, Fama and French (1993) created a benchmark for the construction of portfolios. In the academic finance literature, the search for explaining factors has led to creating portfolios by sorting on characteristics positively associated with expected returns. The standard procedure for creating long-short portfolios by Fama and French (1993) was sorting stocks by their market capitalization and, independently, on their characteristic. When sorted by size, the NYSE median breakpoint is the guideline.

Stocks were split into “small and “big” classifications. Hereafter, the stocks are, independently, sorted into “high and “low” classifications which are based on the 30th and 70th percentile of the characteristic. The high-minus-low portfolio is derived via the crossing of these classifications, which are 6 portfolios. In Table 2 are the sorting techniques identified from the dataset. We also classify other portfolio construction choices like Value-Weighted (VW) and Industry neutrality. These are grouped under others.

Table 2

Portfolio construction choices – Sorting techniques and Others

Sorting techniques	Implemented	Not Implemented
3070	1	0
Independent	1	0
NYSE	1	0

[1] Standard Industrial Classification (SIC) codes are four-digit numerical codes assigned by the U.S. government to business establishments to identify the primary business of the establishment. 4900-4999 is classified as

Electric Gas and Sanitary services and is grouped under Transportation and Public Utilities. 6000-6999 is classified as Finance, Insurance, and Real Estate. Per SICCODE.com

June size	1	0
Others	Implemented	Not implemented
VW	1	0
Industry Neutrality	1	0

As stated in the previous Alinea, the 30/70 portfolio decision is built on the foundation of Fama and French (1993). However, the alternative is the choice of the 20th and 80th percentile and is also often used by researchers e.g., Luo and Balvers (2017), Jegadeesh et al (2004) and and Baltussen et al (2018). Researchers thus have two choices that have been identified. When researchers make a different choice than the two choices presented, we leave the column blank. For example, when choosing the percentiles, a researcher can decide to choose a 25/75 percentile classification like in Naranjo et al (1998). In this case, I reframe from noting a 0 or 1 and leave the column blank. The comments column has then mentioned the reason why it is left blank.

With the sorting procedure of Fama and French (1993), the first sorting is based on size and then independently sorted on a characteristic. This is based on a 2x3 sorting. However, researchers also choose dependent sorting in the 2x3 or other sorting procedures. In papers like Vassalou and Xing (2004)and Bollerslev (2016) dependent sorting is applied. I document a "1" when independent sorting is used and a "0" for dependent sorting. When a paper has both tests with independent and dependent sorting, the conservative method, which is dependent, is noted as the primary sorting technique. Blanks are left in the column when there is no portfolio sorting method identified or mentioned. Furthermore, NYSE breakpoints are commonly used by researchers. However, the calculation of breakpoints over the NYSE-AMEX-Nasdaq was also frequently used by researchers. The "1" is noted when NYSE breakpoints are used and a "0" is documented when breakpoints over the NYSE-AMEX-Nasdaq are implemented. The column is left blank when there is no sorting method identified or mentioned. Lastly, for the sorting techniques, the size breakpoints are based on the market capitalization of firms are the end of June of that corresponding year. This is a common practice among the researchers following Fama and French (1992). Another popular decision was for some papers to use the market capitalization in the previous month in their size sort. The documentation of these decisions was that a "1" was noted when the studies use the end-of-June method and a "0" for the previous month's method. All other practices like beginning-of-the-year methods were documented as blank and added to the commentary section for clarification.

Furthermore, the description of the "Other" section in table 2 is discussed. These consists of Value-Weighted and industry-neutrality portfolio construction techniques. Valueweighting and equal-weighting are weighting schemes that are primarily used when constructing a portfolio. When the VW technique is used, a "1" is noted in the corresponding column, and a "0" when EW is predominately used. When the weighting method is unidentified or different from the two choices presented, the column is left blank for the analyzed paper. This decision is then clarified in the commentary section. Lastly, the final portfolio construction decision highlighted in this paper is industry hedging. The researcher can decide to construct industry-hedged factors to mitigate exposure to specific industries. When the researcher applies industry neutrality, a "1" is noted, and when absent a "0" suffices. When in robustness checks industry neutrality is taken into consideration, it is mentioned in the comments section. However, in the corresponding column, a "0" is noted. This is due that in the primary testing industry neutrality is not taken into consideration.

3.2 Descriptions of the portfolio construction choices

Last, we´ll describe the portfolio construction choices one by one. The remaining question is what the arguments behind the choices by researchers are.

3.2.1 Negative Book-to-Equity ratios

The choice to exclude firms with negative book-to-equity ratios is one that is backed up by several academics. As stated in Brown and Kapadia (2007) academics and practitioners exclude negative BE stocks from analysis, arguing that they are high default risk. Vassalou and Xing (2004) and Griffin and Lemmon (2002) began instating this data restriction choice. However, the arguments are unclear whether this choice is valid or not. “But here is a paradox: if we exclude negative BE stocks then continuity arguments suggest that we should also consider excluding high-growth stocks, which constitute the smallest category of BE relative to ME” (Brown et al, 2007). Regularly, this is not the case when researchers exclude the negative BE firms. Hereby, most academics and practitioners reframe from excluding negative BE firms and decide to exclude small firms which lower the probability of financial distress with firms.

3.2.2 Price Filter and Micro

The data restriction choices of price filtering and excluding micro-caps will be described combined. This is due to the fact, that the aim is for the same result. Fama and French (2008) have found that microcaps capture only 3% of the total market capitalization but are gauged as 60% of all stocks. Amihud (2002) states that illiquid effects are stronger with small stocks and Novy-Marx & Velikov (2016) indicate that small stocks have higher transaction costs. However, small firms are usually included in the data set. For reasons like in Cakici and Zaremba (2021), they control by using a price filter for penny stocks. However, unlike De Moor and Sercu (2013) they do not exclude micro caps because they can control for and explore the role of firm size in a separate test. The importance of whether to include microcaps is crucial as because academics find that “When excluding microcaps from the tests, many discovered anomalies would not exist” (Hou, Xue, & Zhang, 2020). The effect of excluding market caps to the size dimension of the CRSP universe is captured by Soebhag et al (2022). Excluding microcaps increases the median market capitalization. Furthermore, the market share of stocks below the median increases. Lastly, the number of stocks in the small segment decreases under the NYSE breakpoint when microcaps are excluded.

3.2.3 Financials

Fama and French (1992) stated that financial firms have high leverage which is an indication of distress. This created the rationale behind the exclusion of financial firms' choices for other researchers to pursue. The main argument is that the financial services industry is fundamentally different, and therefore reason enough to exclude. Most literature hereafter like Hirshleifer et al (2018) followed Fama and French (1992) in excluding financial firms from their dataset. Soebhag et al (2022) also stated that the reason for excluding financial firms is based on the that including financials may impact factor returns, and some factors more than others. Especially, when factors are not hedged against industry exposure, financial companies may be overweighted or underweighted. However, still some researchers decide to include financial firms like Cohen and Lou (2012) without giving arguments why they included them.

3.2.4 Utilities:

Whereas excluding financial firms is a far more common practice, excluding utility firms is not. In our dataset, mostly the utility firm is excluded when also the financial firm is excluded from the dataset. This is a data restriction that is usually not described when chosen in academic asset pricing literature. In Balachandran and Mohanram (2011) the argument for excluding utility firms was that firms in regulated industries are likely to have guaranteed rates of return on invested capital. In other papers, like Ortiz-Molina and Phillips (2014) and Loughran and Ritter (1995) which exclude utility firms, no explanation is given why they are left out.

3.2.5 Breakpoints 30/70 or 20/80

Creating portfolios sorted by characteristics positively associated with expected returns is customary within academic finance literature. When creating the long-short portfolios various breakpoints have been designed and used by researchers. As mentioned earlier, the 2x3 sorting procedure by Fama and French (1993) is a common practice. It begins with sorting stocks by their market capitalization, where small and big classifications are created based on the NYSE median break-point. Stocks are then classified and assigned into two separate groups. After this procedure, stocks are independently sorted on their characteristic which is investigated in the paper. Stocks are split into three different classifications based on the 30th and 70th percentile. High is based on the 70th percentile and low on the 30th percentile. These procedures combined create six different portfolios, which the high-minus-low portfolio is derived from. Fama and French (2018) and Hirshleifer et al. (2018) are recent examples that use this procedure. One other commonly used practice for breakpoints is the 20th and 80th percentile to sort portfolios. This can be found in Engelberg et al (2018) and Baltussen et al (2018). The consequences of using different breakpoints are such that they cannot be compared to each other Hollstein et al (2021) investigated the impact of different breakpoints (Fama & French, 1993) and found that there is a trade-off between specification versus diversification. More centered breakpoints tend to result in less (idiosyncratic) risk. More extreme sorts create stronger exposures to the underlying anomalies and, thus, higher average returns. Following this, we differentiate between the two in our model.

3.2.6 Independent or dependent

Dependent or independent sorting is the most commonly used procedure within asset pricing literature. Where independent sorting is more often implemented by researchers, dependent sorting has been used more frequently recently. We know that for independent sorting, the standard procedure is to use 5x5 or 2x3 sorting where size and other characteristic is sorted independently. Because they are sorted independently, it doesn't matter whether to start with size or another characteristic. With dependent sorting, this leaves room to change the order. One can imagine that when a 2x3x3 sorting is used, the interpretation of which order to use can be different resulting in different outcomes. However, Fays et al (2021) claim that when using independent sorting forming basis portfolios will lead to a biased allocation of stocks into style portfolios, stratification of the U.S. equity universe, and, therefore, a misleading optimization exercise. This bias is based on from January 1963 to December 2015, the market equity and book-to-market equity of a firm were, on average, negatively correlated (−5%); using independent sorting on negatively correlated variables can induce, by design, a strong tilt toward the extreme categories of inverse ranks, i.e., low-high and high-low. Whether using dependent or independent is, therefore still open for debate.

3.2.7 NYSE or NAN

The choice between NYSE or NYSE-AMEX-Nasdaq (NAN) breakpoints is the choice of whether the researchers want to emphasize smaller stocks. Under the NYSE breakpoints, the median market capitalization is higher than the NAN criteria. This will provide an overweight towards penny stocks and micro caps when using the NAN breakpoints. For example, where this can cause impact, is when measuring illiquidity characteristics, overweighting micro and small caps can result in more favorable results for the researcher to publish.

3.2.8 June size or last months

For the construction of size breakpoints based on the market capitalization of firms, it is common to follow Fama and French's (1992) method. This method is based on size breakpoints at the end of June of the current year t which are updated yearly. The alternative approach that researchers use is the market capitalization in the previous month. This method is updated every month instead of yearly. Examples that use this approach are Acharya and Pedersen (2005) and more recently Chemmanur and Yan (2019). Asset pricing is not concise on why these choices are made. As mentioned by Soebhag et al (2022) is that one argument in favor of using the most recent market capitalization might be to use timely information to construct the size sort. On the other hand, this may result in more turnover, since one rebalances the size sorts each month instead of each year.

3.2.9 Value Weighting (VW) or Equal Weighting (EW)

There is a substantial amount of literature that research the difference between the portfolio construction technique of Value Weighting against Equal Weighting. In conjunction, they agree on the fact that EW portfolios have higher returns than VW portfolios. Ohlson and Rosenberg (1982), Breen et al (1989), Canina et al (1998), Jegadeesh and Titman (1993) and DeMiguel et al (2009) show that the volatility of a VW portfolio is lower than that of an EW portfolio. This is caused by the exposures within the value-weighting procedure depending on the size of the companies in the portfolio. As stated by Bhattacharya and Galpi (2011) the weight of stock i in a value-weighted portfolio is, by definition, proportional to the market capitalization of stock i. This shows when smaller firms are equally weighted in the EW approach more emphasis will be set on the smaller firms which tend to be more volatile and can generate higher returns. The systematic risk factors determined by the initial portfolio weights explain the higher return for EW. It is common for researchers to show in the robustness test the alternative weighting to the main used in the tests. Even though the valueweighted approach is the most common approach, the equal-weighted is still often used in studies. Bhattacharya and Galpi (2011) highlight that in developed markets the value-weighted approach is more often used and in emerging markets the equal-weighted approach.

The study shows that this is caused by institutional shareholders and mutual funds demanding valueweighted portfolios as they use the value-weighted approach as a benchmark themselves.

3.2.10 Industry neutrality or unhedged

Lastly, the portfolio construction choice of neutralizing industry effects is highlighted. Stock characteristics predictive power can originate from sector-specific components or components across different industries. Moskowitz and Grinblatt (1999) find strong evidence that persistence in industry return components generates significant profits that may account for much of the profitability of individual stock momentum strategies. Also, Liu et al (2014)results show that the full-universe portfolios are significantly subject to the effects of the industries that perform badly during periods of market turbulence, and therefore suffer substantial downside risk. This shows that we can't neglect the possible impact of industry-neutral portfolios compared to unconditional sorting. However, even in the broadly accepted theory in Fama and French (1992) industry neutrality is not mentioned.

Now that we understand which decisions have been made in our sample, which represents roughly the broad asset-pricing literature, we are going to look at the frequency of these decisions and what this tells us. The next section is dedicated to the elaboration of these questions.

4 Results

Our sample consists of the 325 empirical articles identified for the list by Harvey and Liu (2019). In this section, we're going to document how frequent choices are being made and what this explains to us. If the construction choices were made randomly, the expectation would be that the studies pick 50% of the time one specific technique and 50% the alternative option. Some binary choices identified in this study could have a theoretical foundation and thus have good reasons to pick. We would expect that these options would that the frequency of the choices would be close to 100%. In table 3 the options are displayed with the percentage of studies that have picked this design choice.

Table 3

This table presents the methodological options identified and analyzed within our sample of 325 empirical studies between 1965-2018. The list of all the papers is constructed by Harvey and Liu (2019). The methodological options include the portfolio construction decisions and the data restriction decisions. Under the Options label, you can find the option presented for the researcher. To what quantity each option is picked is presented under the Distribution label

	Methodological options		**Distribution**	
	(1)	(2)	(1)	(2)
Choice 1	Exclude *BE* < 0	Include *BE* < 0	20.92 %	79.08 %
Choice 2	Place price filter	No price filter	18.15 %	81.85 %
Choice 3	Include microcaps	Exclude microcaps	87.69 %	12,31 %
Choice 4	Include financial firms	Exclude financial firms	71.38 %	28.62 %
Choice 5	Include utility firms	Exclude utility firms	90.46 %	9.54 %
Choice 6	Use 30/70 BP	Use 20/80 BP	46.90 %	53.10 %
Choice 7	Independent	Dependent	70.59 %	29.41 %
Choice 8	Use NYSE BP	Use NAN BP	41.26 %	58.74 %
Choice 9	June Size	Last month’s size	67.97 %	32.03 %
Choice 10	Value-Weighted	Equal-Weighted	58.74 %	41.26 %
Choice 11	Industry Neutrality	Unhedged	11.60 %	88.40 %

Note: BP = Breakpoints, BE = Book-to-Equity

When observing table 3 the results show that particular choices' pick probability is fairly equal. The preferences from researchers are clear as the deviation between specific choices is substantial. Excluding negative book-to-equity companies are, with 20.92%, far less often chosen than including these companies with researchers selecting 79.08% in the studies. Placing a price filter to eliminate certain, small, companies are less popular in the sample studies and is selected by 18.15%. Leaving this price filter out of the data selection criteria is preferred by 81.85% of the researchers. Excluding microcaps, firms with a market capitalization under a certain threshold are removed from the sample, are chosen within 12,31% of the studies. Including these microcaps is substantially more common by researchers with 87.69%. The decision to include financial firms within the sample (71.38%) is considerably more chosen than the excluding financial firms option (28.62%). There are no options that are 100% selected in all studies in the sample. The closest to 100% is the option to include utility firms with 90.46% of the researchers choosing this. Excluding utility firms are far less popular and are only determined by 9.54%. Using 30/70 breakpoints (46.90%) is close to the same amount that the 20/80 BP approach (53.10%) is chosen. The decision on independent or dependent sorting is less close, with 70.59% for independent and 29.41% for dependent sorting. Using NYSE breakpoints (41.26%) is not as popular to select as using the NAN breakpoints (58.74%). Sorting the size with the market capitalization from last June is used more in the studies than last month's market capitalization. Finally, weighting the portfolio using the value method (58.74%) is more often chosen than equal weighting (41. 6). Last, the decision to impose industry neutrality (11.60%) is far less commonly used than to leave them unhedged (88.40%). When observing the percentages reported in table 3 we find that none of the decisions made is counted under twenty studies. This means that at least twenty other studies pick the same methodological decision and can be cited by the researcher. We'll look at the studies that introduced an approach and who last used it in table 5.

First, the correlation matrix is used to determine whether combinations of the options presented are rare the correlation coefficients between the choices are not particular exceptionally see that the highest coefficient is 0.52, which indicates that there is an increased probability that when using independent sorting the June market capitalization is used to determine size instead of the last month's market capitalization. Similar, researchers that exclude negative Book to Equity firms are more likely to exclude financial firms (-0.51). The lowest correlation we find is 0.03 which explains that there is only a small favorable probability that researchers including microcaps would choose to enable 30/70 BP instead of 20/80 BP. We even examine that there is no correlation between excluding negative BE firms and using 30/70 BPs. This indicates that there is no relationship between these two options.

Furthermore, it would be interesting to examine which study is the first examined chronologically to choose an option. The researcher is the first within asset-pricing literature within our sample, to use this portfolio construction technique and therefore should explain the rationale behind this decision. In the standard of portfolio construction by Fama and French, explanations have been given on why financial firms are excluded. The exclusion of financial firms is due to their business model, which is highly different from other companies (Fama & French, 1992). Due to the nature of their business model, financial firms are highly levered. For non-financial firms, high leverage indicates financial distress and, therefore can't be compared to financial firms. In table 5 we examine the studies that pioneer using a particular portfolio design choice within asset-pricing literature. Besides reviewing the researchers who began using this approach, our interest also lies in the latest researchers that used the method. We aim to observe whether portfolio construction decisions are rationalized and theoretically explained. We exclude the Fama and French (1993) standard in the table, which consists of the following options: including BE < 0[1][2], No price filter, including micro caps, excluding financial firms, including utility firms, Use 30/70 BP[3], Independent sorting[3], Use of NYSE BP, June Size, Value-weighted and Unhedged.

Table 5 shows us that only six studies explain the rationale behind their particular choice. Four out of six studies introduced the approach within asset pricing literature and explained why. Two studies of the ´new´ group specified why they decided to use this choice. As the table reports, fifteen of the twenty-one studies do not give a rationale. This indicates that researchers often find it unnecessary to elaborate on their decision to use this method. It is expected that 'old' papers pioneer this approach and explains why they use it. This can be said for 'new' articles as well, where you'd expect that the researcher would describe the process of constructing the portfolio due to the many options and available knowledge. The results don't signal these expectations. Lastly, how does the decision made effect the t-statistic of the model is presented in table 6. Here the eleven construction choices that are faced by researchers are linear regressed against the t-statistics of the models used by the researchers. This tells what the impact of the decision is to the t-statistic.

[1] Excluded by Fama and French (1992) but not considered as the standard
[2] /80 is used by Fama and French in various tests but not considered as the standard
[3] Dependent sorting is also explained by Fama and French but not considered as the standard

Table 4

Correlation of methodological choices.

This table presents the correlation matrix between the eleven options within our sample.

Choice	Negative BE	Price Filter	Financials	Micro	Utilities	Industry Neutrality	NYSE	Independent	Jun_Size	BP3070
Negative BE	1.00									
Price Filter	-0.14	1.00								
Financials	-0.51	0.21	1.00							
Micro	-0.03	-0.39	-0.11	1.00						
Utilities	0.13	0.02	0.33	-0.10	1.00					
Industry Neutrality	-0.14	0.17	0.15	-0.17	-0.03	1.00				
NYSE	0.30	-0.10	-0.29	-0.05	0.18	-0.23	1.00			
Independent	0.27	-0.05	-0.30	-0.03	0.08	-0.06	0.14	1.00		
Jun_Size	0.36	-0.18	-0.42	0.14	0.06	-0.10	0.25	0.52	1.00	
BP3070	0.00	0.12	-0.15	0.03	-0.17	0.13	0.08	0.22	0.24	1.00

Table 5

List of studies using alternative portfolio techniques than the Fama and French standard*. It consists of the new and old studies chronologically that utilized this method from the sample provided by Harvey and Liu (2019). Including rationale, if present, on why this option has been chosen. The sample range is from 1965 until 2018.*

Option	Year	Researcher	Rationale
Exclude *BE* < 0	1992	Fama and French	Rare in data before 1980
	2018	Akbas, Jiang, and Koch	No rationale given
Place price filter	1997	Loughran and Vijh	Eliminates firms that are very small or in distress
	2018	Engelberg, Reed, and Ringgenberg	No rationale given
Exclude microcaps	1997	Loughran and Vijh	Eliminates firms that are very small or in distress
	2018	Andrade and Chhaochharia	No rationale given
Including financial firms	1965	Lintner	No rationale given
	2018	Da, Warachka, and Yun	No rationale given
Exclude utility firms	1995	Loughran and Ritter	Tend to be different from those of other operating companies
	2018	Hirshleifer, Hsu, and Li	No rationale given
Use 20/80 BP	1981	Banz	No rationale given
	2018	Engelberg, Reed, and Ringgenberg	No rationale given
Dependent	1981	Banz	No rationale given
	2018	Engelberg, Reed, and Ringgenberg	No rationale given
Use NAN BP	1983	Arbel, Carvell, and Strebel	No rationale given
	2018	Hirshleifer, Hsu, and Li	No rationale given

Last month's size	1981	Banz	No rationale given
	2018	Chemmanur and Yan	No rationale given
Equal-Weighted	1967	Douglas	No rationale given
	2018	Hollstein and Prokopczuk	Value-weighted portfolios can be dominated by a few big stocks
Industry Neutrality	1988	Jacobs and Levy	Utilized to purify anomaly return attributions from the impact of industry return comovement
	2018	Hirshleifer, Hsu, and Li	To make sure our results are not driven by any particular industry

Lastly, how does the decision made effect the t-statistic of the model is presented in table 6. Here the eleven construction choices that are faced by researchers are linear regressed against the t-statistics of the models used by the researchers. This tells what the impact of the decision is to the t-statistic. With significancy at the 5% level and 10% level, excluding negative BE firms and financial firms to your data set affects the t-statistic with 7.85 and -5.82 respectfully. This shows that these portfolio construction decisions have substantial impact to the t-value of your model and therefore the significancy of the model. Other big decisions that have big impacts are using NYSE breakpoints (4.61), using value-weighting (-3.51), neutralizing your portfolio's with an industry factor (-3.62), excluding microcaps (-2.07) and using 30/70 breakpoints (6.45). However, these coefficients are all insignificant and therefore only gives us insights on a possible impact. Decisions that have possible less impact to the t-value of the researchers model is excluding utility firms from the sample (-1.55), using a price filter to exclude very small firms (-1.24) and whether to use independent or dependent sorting (0.89). Same as other decisions, these findings are not significant and therefore has no explanatory value besides showing us insights.

Table 6

This table presents the methodological options identified within our sample of 325 empirical studies between 1965-2018 regressed against the t-statistic of the models used in the studies. Definitions of the construction choices are the same as mentioned in section III. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level

Portfolio construction decisions and T-statistics											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Negative BE	7.85** (2.26)										
Price Filter		-1.24 (-0..35)									
Exclude Microcaps			-2.07 (-0.54)								
Exclude Financials				5.82* (-1.89)							
Exclude Utilities					-1.55 (0.33)						
30/70						6.45 (1.05)					
Independent							0.89 (0.16)				
NYSE								4.61 (1.47)			
June size									4.62 (0.64)		
VW										-3.51 (-1.10)	
Industry Neutrality											-3.62 (-0.48)

5 Conclusion

In asset pricing literature there is a continuous search for explaining factors. Factors that are significant enough to cause discussion. Discussion leads to more citations and therefore is interesting for editors and publishers to issue the article in top journals. In this paper, we highlighted the search and how academics construct significant factors. As the number of asset pricing literature papers published in top journals is increasing, researchers have to dig deeper to find this significant factor. Character-based sorting is one of the popular methods to construct factors. This paper calls attention to the fact that many decisions need to be made when constructing factors. Since there are no guidelines or consensus on construction methods, researchers have degrees of freedom in their decisions. This allows for p-hacking when the outcome is impacted by the researcher's design choice. Researchers could design their portfolio construction choices in such a matter that it would overcome statistical barriers and produce significant factors. This paper doesn't investigate the impact of these construction choices on the performance of the portfolio. This is done by Soebhag et al (2022). They show that design choices substantially affect a model's probability of producing the highest Sharpe ratio. Construction choices, therefore, impact factor returns. This paper highlights the portfolio construction techniques observed in the last fifty years and documents their frequency and relevancy. We identified the main options researchers faced and discovered that the probability that an option is picked is fairly equal. There are also no strong correlations for the combinations of options. Excluding negative Book-to-Equity firms and financial firms from the sample has a significant and substantial impact to the significancy of the model. This indicates that there is no common consent among researchers on which option to choose. We searched for the rationale for why these decisions have been made by researchers. This concludes that researchers often reframe from giving explanations. Even studies that pioneer with a particular approach lack in clarification of what the rationale behind the decision is. The insights gained from the findings are important to acknowledge the deficiencies in the current and past state of asset pricing literature in terms of portfolio construction theory. This paper calls on future studies to elaborate on the insights given in this paper and build on the findings of the impact analysis by Soebhag et al (2022) In the recommendations section, the foundation of a potential future guideline is written and I encourage other researchers to assemble the remaining parts. In essence, to answer our research questions; What are the portfolio construction techniques and how are they used? There are a lot of various techniques used and are not given a rationale. The demand for guidelines hasn't been as big as now.

6 Recommendations

To write a uniform and clear guideline that applies to all conditions is difficult within empirical literature. However, insights gained by this paper and that of various others give reason to start. Mitton (2022) investigated the methodological variation in empirical corporate finance and found that the field could profit from narrowing down the freedom regarding the reporting of the robustness results. Brodeur et al. (2020) recommended that researchers graphically should show the distribution of their Sharpe ratios or other results. Soebhag et al (2022) findings report that the most important design choice around factor construction is those concerning NYSE or NAN breakpoints, micro stocks, industry-adjusted characteristics, and value-weighting. These papers all suggest tools to make sure analyzing papers will be easier when the papers have similar choices. After all, apples can only be compared to apples. Therefore, when construction options are different factors models shouldn't be compared with each other. The recommendations in this paper will be;

1) When constructing portfolios the various decisions should be documented and explained in the study
2) There should always be robustness tests with the alternative construction portfolio options documented and explained in the study
3) The Sharpe ratios of the chosen option and the alternative option should be visible and documented in the study

As stated by Soebhag et al (2022), NYSE or NAN breakpoints. Including or excluding microcaps, industry-adjusted characteristics or unhedged and value-weighting or equal weighting have the most impact on the performance of the portfolio. This paper aims for uniformity and therefore will suggest a standard based on rationales. Unless the researchers give a clear reason why to use an alternative approach, the standard will be applied. This is due to the following, NYSE breakpoints are widely accepted as size sorting because 60% of all the stocks in the three big exchanges are only 3% of the value. If you base your breakpoints on such values it will impact the sorting groups heavily. Therefore in this recommendation, NYSE breakpoints are the standard. For the decision to include or exclude microcaps, microcaps are often included although there is empirical consensus that small firms are illiquid, have higher transaction costs, are more volatile, and are out of reach for most institutional investors. Therefore, excluding microcaps from the sample as standard is recommended. When sorting unconditionally, factor portfolios gain distinctive exposure to certain industries. These portfolios pick up industry risks and are therefore not mean-variance efficient. To bring practitioners (who use it in their own models) and the empirical world more together, industry neutrality is considered the standard. When the conditional sorting approach is used, it is unclear what the effect is and this should be investigated further. In practice, Value-weighting is used as a benchmark to which portfolio managers are assessed. Equal-weighting gives more weight to small stocks and therefore equal-weighted portfolios are more illiquid. Plyakha et al (2015) identified the proportion of the excess return of the equal-weighted portfolio relative to the value- and price-weighted portfolios that come from differences in alpha and the proportion that comes from differences in systemic risk. Value-weighted is considered the standard by this paper. In our result section, we can also see that the insights given by the linear regression are that these choices may substantially impact the t-statistic of the model.

4) Standard portfolio construction techniques: NYSE BP, excluding microcaps, if unconditional industry neutral, value-weighted

With these recommendations for the guideline utilized, papers will be easier to compare and phacking shall be spotted. This will be a vital process for the future when written and published papers will be increasing more and more. To summarize, the guideline is crucial for the future of asset pricing literature.

7 Reference list

Acharya, V. V., & Pedersen, L. H. (2005). Asset pricing with liquidity risk. *Journal of financial Economics*, pp. 375-410.

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of financial markets*, pp. 31-56.

Asness, C. S., Moskowitz, T. J., & Pedersen, L. H. (2013). Value and momentum everywhere. *The Journal of Finance*, pp. 929-985.

Balachandran, S., & Mohanram, P. (2011). Is the decline in the value relevance of accounting driven by increased conservatism?. *Review of Accounting studies*,, pp. 272-301.

Baltussen, G., Van Bekkum, S., & Van Der Grient, B. (2018). Unknown unknowns: uncertainty about risk and stock returns. *Journal of Financial and Quantitative Analysis*, pp. 1615-1651.

Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of financial economics* , pp. 3-18.

Basu, S. (1983). The relationship between earnings' yield, market value and return for NYSE common stocks: Further evidence. . *Journal of financial economics*, pp. 129-156.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, pp. 289-300.

Bentley, P. (2021, April 1). Error rates in SARS-CoV-2 testing examined with Bayes' theorem. *Heliyon*.

Bhattacharya, U., & Galpin, N. (2011). The global rise of the value-weighted portfolio. *Journal of Financial and Quantitative Analysis*, pp. 737-756.

Bollerslev, T., Li, S. Z., & Todorov, V. (2016). Roughing up beta: Continuous versus discontinuous betas and the cross section of expected stock returns. *Journal of*

Financial Economics, pp. 464-490.

Breen, W., Glosten, L. R., & Jagannathan, R. (1989). Economic significance of predictable variations in stock index returns. *The Journal of finance*, pp. 1177-1189.

Brodeur, A., Cook, N., & Heyes, A. (2020). Methods matter: P-hacking and publication bias in causal analysis in economics. *American Economic Review*, pp. 3634-60.

Brown, G., & Kapadia, N. (2007). Firm-specific risk and equity market development. *Journal of Financial Economics*, pp. 358-388.

Cakici, N., & Zaremba, A. (2021). Size, Value, Profitability, and Investment Effects in International Stock Returns: Are They Really There? *The Journal of Investing, ,* pp. 65-86.

Canina, L., Michaely, R., Thaler, R., & Womack, K. (1998). Caveat compounder: A warning about using the daily CRSP equal-weighted index to compute long-run excess returns. *The Journal of Finance*, pp. 403-416.

Carhart, M. M. (1997). On Persistence in Mutual Fund Performance. *The Journal of Finance*, pp. 57-82.

Chemmanur, T. J., & Yan, A. (2019). Advertising, attention, and stock returns. *Quarterly Journal of Finance*.

Cochrane, J. H. (2011). Presidential address: Discount rates. *The Journal of Finance*, pp. 1047-1108.

Cohen, L., & Lou, D. (2012). Complicated firms. *s. Journal of financial economics,,* pp. 383-400.

De Bondt, W. F., & Thaler, R. (1985). Does the Stock Market Overreact? *The Journal of Finance*, pp. 793-805.

De Moor, L., & Sercu, P. (2013). The smallest firm effect: An international study. *Journal of International Money and Finance,,* pp. 129-155.

DeMiguel, V., Garlappi, L., Nogales, F. J., & Uppal, R. (2009). A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management science*, pp. 798-812.

Donangelo, A. (2014). Labor mobility: Implications for asset pricing. *The Journal of Finance*, pp. 1321-1346.

Dutta, A. (2019). Does the Five-Factor Asset Pricing Model Have Sufficient Power? *Global Business Review*, pp. 684-691.

Engelberg, J. E., Reed, A. V., & Ringgenberg, M. C. (2018). Short-selling risk. *The Journal of Finance,*, pp. 755-786.

Engelberg, J., McLean, R. D., & & Pontiff, J. (2015). Anomalies and News.

Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *The Journal of Finance*, pp. 427-465.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, pp. 3-56.

Fama, E. F., & French, K. R. (1996). Multifactor explanations of asset pricing anomalies. *The Journal of Finance*, pp. 55-84.

Fama, E. F., & French, K. R. (2004). The capital asset pricing model: Theory and evidence. *Journal of economic perspectives*, pp. 25-46.

Fama, E. F., & French, K. R. (2008). Dissecting anomalies. *The Journal of Finance*, pp. 1653-1678.

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of financial economics*, pp. 1-22.

Fama, E. F., & French, K. R. (2018). Choosing factors. *Journal of financial economics*, pp. 234-252.

Fanelli, D. (2010). Do pressures to publish increase scientists' bias? An empirical support from US States Data. *PloS one*.

Fays, B., Papageorgiou, N., & Lambert, M. (2021). Risk optimizations on basis portfolios: The role of sorting. *Journal of Empirical Finance,* pp. 136-163.

Feng, G., Giglio, S., & Xiu, D. (2020). Taming the factor zoo: A test of new factors. *The Journal of Finance*, pp. 1327-1370.

Griffin, J. M., & Lemmon, M. L. (2002). Book-to-market equity, distress risk, and stock returns. *The Journal of Finance*, pp. 2317-2336.

Harvey, C. R. (2017, July 8). Presidential Address: The Scientific Outlook in Financial Economics. *The Journal of Finance, 72*(4), pp. 1399-1440.

Harvey, C. R., & Liu, Y. (2015). Backtesting. *The Journal of Portfolio Management*, pp. 13-28.

Harvey, C. R., & Liu, Y. (2019). A census of the factor zoo. Opgehaald van Available at SSRN 3341728

Harvey, C. R., Liu, Y., & Zhu, H. (2016). … and the cross-section of expected returns. *The Review of Financial Studies*, pp. 5-68.

Hirshleifer, D., Hsu, P. H., & Li, D. (2018). Innovative originality, profitability, and stock returns. *The Review of Financial Studies*, pp. The Review of Financial Studies, 31(7), 2553-2605.

Hollstein, F., Prokopczuk, M., & Voigts, V. (2021). How Robust are Empirical Factor Models to the Choice of Breakpoints?. Opgehaald van Available at SSRN 3924821,

Hou, K., Xue, C., & Zhang, L. (2020). Replicating anomalies. *he Review of Financial Studies*, pp. 2019-2133.

Huynh, T. D. (2018). Explaining anomalies in Australia with a five-factor asset pricing model. *International Review of Finance*, pp. 123-135.

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers:

Implications for stock market efficiency. *The Journal of Finance*, pp. 65-91.

Jegadeesh, N., Kim, J., Krische, S. D., & Lee, C. M. (2004). Analyzing the analysts: When do recommendations add value?. *The Journal of Finance*, pp. 1083-1124.

Jones, C. M., & Lamont, O. A. (2002). Short-sale constraints and stock returns. *Journal of Financial Economics*, pp. 207-239.

Lakonishok, J., Shleifer, A., & Vishny, R. W. (1994). Contrarian Investment, Extrapolation, and Risk. *The Journal of Finance*, pp. 1541-1578.

Lin, Q. (2017). Noisy prices and the Fama–French five-factor asset pricing model in China. *Emerging Markets Review*, pp. 141-163.

Lintner, J. (1965). Security prices, risk, and maximal gains from diversification. *The Journal of Finance*, pp. 587-615.

Liu, X., Pong, E. S., Shackleton, M. B., & Zhang, Y. (2014). Option-implied volatilities and stock returns: Evidence from industry-neutral portfolios. *The Journal of Portfolio Management*, pp. 65-77.

Loughran, T., & Ritter, J. R. (1995). The new issues puzzle. *The Journal of finance*, pp. 23-51.

Luo, H. A., & Balvers, R. J. (2017). Social screens and systematic investor boycott risk. *Journal of Financial and Quantitative Analysis*, pp. 365-399.

Markowitz, H. (1959). *Portfolio Selection: Efficient Diversification of Investments.* New York: John Wiley & Sons.

Mayers, C., & Baker, K. (2020, June 3). Impact of false-positives and false-negatives in the UK's COVID-19 RT-PCR testing programme. United Kingdom.

McLean, R. D., & Pontiff, J. (2016). Does academic research destroy stock return predictability? *The Journal of Finance*, pp. 5-32.

Mina, M., & Andersen, K. (2020). COVID-19 testing: One size does not fit all. *Science*, pp. 126-127.

Mitton, T. (2022). Methodological variation in empirical corporate finance. *The Review of Financial Studies*, pp. 527-575.

Moskowitz, T. J., & Grinblatt, a. M. (1999). Do industries explain momentum? *The Journal of Finance*, pp. 1249-1290.

Naranjo, A., Nimalendran, M., & Ryngaert, M. (1998). Stock returns, dividend yields, and taxes. *The Journal of Finance*, pp. 2029-2057.

Novy-Marx, R. &. (2016). A taxonomy of anomalies and their trading costs. *The Review of Financial Studies*, pp. 104-147.

Novy-Marx, R. (2013). The other side of value: The gross profitability premium. *Journal of financial economics*, pp. 1-28.

Ohlson, J., & Rosenberg, B. (1982). Systematic risk of the CRSP equal-weighted common stock index: A history estimated by stochastic-parameter regression. *Journal of Business*, pp. 121-145.

Ortiz-Molina, H., & Phillips, G. M. (2014). Real asset illiquidity and the cost of capital. *Journal of Financial and Quantitative Analysis*, pp. 1-32.

Özkan, N. (2020). A Comparison of New Factor Models: Evidence From Turkey. *Ege Academic Review*, pp. 193-207.

Plyakha, Y., Uppal, R., & Vilkov, G. (2015). Why do equal-weighted portfolios outperform value-weighted portfolios. *SSRN Electronic Journal*.

Roll, R. (1977). A critique of the asset pricing theory's tests Part I: On past and potential testability of the theory. *Journal of Financial Economics*, pp. 129-176.

Roll, R., & Ross, S. (1980). An Empirical Investigation of the Arbitrage Pricing Theory. *The Journal of Finance*, p. 1073.

Rosenberg, B., Reid, K., & Lanstein, R. (1985). Persuasive Evidence of Market Inefficiency. *Journal of Portfolio Management*, pp. 9-17.

Ross, S. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, pp. 341-360.

Schweder, T., & Spjøtvoll, E. (1982). Plots of p-values to evaluate many tests simultaneously. *Biometrika*, pp. 493-502.

Shanken, J. (1982). The Arbitrage Pricing Theory: Is it Testable? *The Journal of Finance*, pp. 1129-1140.

Shanken, J. (1990). Intertemporal asset pricing: An empirical investigation. *Journal of Econometrics*, pp. 99-120.

Sharpe, W. F. (1954). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, pp. 425-442.

Soebhag, A., Vliet van, B., & Verwijmeren, P. (2022). Non-Standard Errors in Asset Pricing: Mind Your Sorts. Opgehaald van https://ssrn.com/abstract=4136672 or http://dx.doi.org/10.2139/ssrn.4136672

Sutton, A. J., Song, F., Gilbody, S. M., & Abrams, K. R. (2000). Modelling publication bias in meta-analysis: a review. . *Statistical methods in medical research*, pp. 421-455.

Titman, S., Wei, K. J., & Xie, F. (2004). Capital investments and stock returns. *Journal of financial and Quantitative Analysis*, pp. 677-700.

Vassalou, M., & Xing, Y. (2004). Default risk in equity returns. *The journal of finance*, pp. 831-868.

Whited, T. M., & Wu, G. (2006). Financial constraints risk. *The review of financial studies*, pp. 531-559.