JOHANNES KEPLER
UNIVERSITÄT LINZ

Author
Florian Halmdienst, BSc

Submission
Institute of Corporate Finance / Department of Asset Management

Thesis Supervisor
Assoz. Univ.-Prof. Dr. Johann Burgstaller

April 2020

Visiting the Factor Zoo: Theory Meets Practice

Master's Thesis

to attain the academic degree of

Master of Science

in the Master's Program

Finance and Accounting

JOHANNES KEPLER
UNIVERSITÄT LINZ
Altenberger Straße 69
4040 Linz, Österreich
www.jku.at
DVR 0093696

SWORN DECLARATION

I hereby declare under oath that the submitted Master's Thesis has been written solely by me without any third-party assistance, information other than provided sources or aids have not been used and those used have been fully documented. Sources for literal, paraphrased and cited quotes have been accurately credited.

The submitted document here present is identical to the electronically submitted text document.

Linz, April 2020

. . . .

Florian Halmdienst

TABLE OF CONTENTS

List of figures

List of tables

List of abbreviations

ACAPM	Anchoring-adjusted capital asset pricing model
AMEX	American stock exchange
APT	Arbitrage pricing theory
AQR	Applied Quantitative Research
AuM	Assets under Management
B/M	Book-to-market equity
BHY	Benjamini, Hochberg, and Yekutieli
BS6	Barillas and Shanken (2018) six-factor model
CAPM	Capital asset pricing model
CRSP	Center for Research in Security Prices
D3	Daniel et al. (2020) three-factor model
DDM	Dividend discount model
E.g.	Example gratia
Ed(s).	Editor(s)
EMH	Efficient market hypothesis
ER	Expense ratio
Et al.	Et alii
ETF	Exchange-traded fund
F.	Following
FF3	Fama and French (1993) three-factor model
FF5	Fama and French (2015) five-factor model
FF6	Fama and French (2018) six-factor model
GDP	Gross Domestic Product
H*Q*4	Hou et al. (2015) *q*-four-factor model
H*Q*5	Hou et al. (2018a) q^5-factor model
LASSO	Least Absolute Shrinkage and Selection Operator
MSCI	Morgan Stanley Capital International
NASDAQ	National Association of Securities Dealers Automated Quotations

NYSE	New York Stock Exchange
P.	Page
P.a.	Per annum
SSGA	State Street Global Advisors
SY4	Stambaugh and Yuan (2017) four-factor model
T-bill	U.S. treasury bill
Vs.	Versus

1. Introduction

1.1. Problem statement

Explaining the cross-section of stock returns has remained one of the key issues in finance for several decades. The first popular single-factor model was the capital asset pricing model (CAPM) of Sharpe (1964), Lintner (1965), and Mossin (1966), which explains the return of a stock as a linear function of its exposure to the market risk (beta factor). Over time, studies (e.g., Basu, 1977, 1983; Banz, 1981; Titman et al., 2004; Novy-Marx, 2013) have shown that, besides the beta factor, other factors such as value, size, investment, and profitability also contribute to explaining the cross-section of expected returns. These anomalies[1] violated the predominant CAPM and have led to the development of new asset pricing models, such as the q^5-factor model of Hou et al. (2018a) and the six-factor model of Fama and French (2018).

So far, academic research has detected over 400 potential factors that attempt to explain the cross-section of returns (Cochrane, 2011; Harvey et al., 2016; Hou et al., 2018b; Harvey and Liu, 2019). In particular, the majority of these factors have been discovered in the last 15 years (Harvey et al., 2016). Cochrane (2011, p.1047) describes this issue as "[…] a zoo of new factors." However, many researchers (e.g., Lo and MacKinlay, 1990; Harvey et al., 2016; Harvey, 2017; Hou et al., 2018b; Linnainmaa and Roberts, 2018; Chordia et al., 2019) argue that most factors are simply lucky findings due to extensive data mining or multiple testing.[2]

Nevertheless, some of these factors provide useful insights for investors. They capture empirical cross-sectional patterns of abnormal returns that could be translated into feasible investment strategies. Factor investing,[3] which strives to harvest these theoretical factor returns, can be retraced to the arbitrage pricing theory (APT) of Ross (1976) and the three-factor model of Fama and French (1993). It has attracted widespread attention in the investment community since the investigation of the Norwegian government pension fund by Ang et al. (2009). Ever since, studies (e.g., Ang et al., 2009; Eun et al., 2010; Van Gelderen and Huij, 2014; Blitz, 2012, 2015; Koedijk et al.,

[1] This paper uses the terms "anomaly," "return predictive signal," and "factor" interchangeably.

[2] See also McLean and Pontiff (2016) who demonstrate that returns of anomalies seem to disappear partially after their dissemination.

[3] This paper uses the terms "factor investing" and "smart beta" interchangeably.

2016; Dimson et al., 2017) have suggested that factor investing might be able to improve long-term portfolio returns. Nonetheless, this investment approach is a controversial one and its risks are often misunderstood (Malkiel, 2014; Arnott et al., 2016b; Glushkov, 2016; Arnott et al., 2017a, 2017b; Kalesnik and Linnainmaa, 2018; Arnott et al., 2019; Chen and Velikov, 2019; Patton and Weller, 2019).

Given the multitude of possible factors to explain and predict stock returns, the questions arise as to which factors actually provide independent information and how to choose them (Cochrane, 2011). Additionally, selecting among contending factor models is challenging (Fama and French, 2018). In practice, if thousands of factor strategies are backtested, some of them will look fantastic on paper purely by chance. But these backtested returns lead to exaggerated expectations about prospective factor performance. Indeed, as noted by Arnott et al. (2019), factor investing was mostly incapable of meeting its high expectations, especially over the last 15 years. Factor investing is a difficult business. Building a factor portfolio with ingenuous assumptions leads to disappointing results. Moreover, investment companies also employ many scientific authors. Prominent examples include AQR Capital Management, Robeco, Research Affiliates, BlackRock, and Lyxor. For one thing, practitioners search for factors and publish articles on them. On the other hand, it seems that the knowledge transfer between science and practice regarding factor selection does not run smoothly in both directions. To sum up, the finding of over 400 factors has led to more questions than answers in the literature. Furthermore, future performance of factor investing is often exaggerated, and the risks involved are often underestimated (Kalesnik and Linnainmaa, 2018; Arnott et al., 2019).

Issues surrounding the explanation of the cross-section of expected returns are a central theme in the finance literature. Therefore, research in the field of factors that independently predict and explain stock returns is still a highly relevant topic in the literature. However, research within this field is far from conclusive. To date, there is no consensus in the literature about which factors provide independent information about stock returns and whether factor returns are attributable to systematic risk, mispricing due to behavioral biases or data mining. In addition, factor investing is far off from just being a niche product. The world’s biggest asset managers such as BlackRock, Vanguard Group, and UBS (Switzerland) have widely adopted this investment approach. To illustrate, the whole factor industry already amounts about $2 trillion, yet it is expected to reach $3.4 trillion by 2022 (BlackRock, 2018). Hence, the implementation of factor investing and the risks involved are issues that are worth a look in more detail as huge amounts of capital flow into this investment approach.

1.2. Goal setting and research questions

Considering the growing number of factors that try to improve our understanding of the relation between returns and risk, and the rising importance of these factors in the investment practice, the purpose of this thesis is twofold. First, this thesis provides an overview of the academic discussion regarding the enormous number of factors and on how to determine which ones are true. Second, this paper aims to help investors develop more reasonable expectations about the expected returns of factor investing and the risks involved. In order to attain these objectives, I critically evaluate the existing literature on factors, factor models, and factor investing. Consequently, I aim to answer the following research questions:

(1) What are the reasons for the growing number of factors and which methods to choose for the reliable and independent factors are proposed by the academic literature?

The first research question addresses, among other things, the problem of data mining or multiple testing in financial economics. The main focus of attention lies on the reasons for the proliferation of the factor zoo, on how to deal with the issue of multiple testing, and on the discussion about different statistical approaches for identifying reliable and independent factors.

(2) Which factors have been included in asset pricing models over time and do practitioners actually implement the factors suggested by the academic literature?

The objective of this research question is to identify the factors that are included in asset pricing models. Also discussed is the theoretical motivation behind some of the well-known asset pricing models and performance comparisons between these models. Moreover, in order to evaluate the knowledge transfer between science and practice, factors applied in asset pricing models and factors suggested by practitioners are compared. Finally, this paper discusses possible reasons for the divergence between factors proposed by science and those implemented by practitioners.

(3) *How can we translate theoretical factor premiums into practicable investment strategies?*

The main purpose of this research question is to help investors understand the different possibilities on how to implement factors into workable strategies. In this context, this paper examines the debates about long-only vs. long-short and active vs. passive factor implementation. In that respect, systemic risks of factor-based exchange-traded funds (ETFs) are examined. Furthermore, I aim to explore which factors might be applicable and which factor premiums are likely to persist in the future. Therefore, Chapter 4.3.1 provides a comprehensive overview of the factors' potential drivers.

(4) *What are the risks of factor investing that are often ignored by investors?*

Finally, the examination of the risks of factor investing should help investors by developing more realistic expectations about future factor returns.

1.3. Course of action

This thesis has five chapters. The introduction contains the problem statement, the goal setting and the research questions. Chapter 2 clarifies the theoretical foundations. The third chapter examines the discussion on factors in the cross-section of expected returns from an academic point of view. In Chapter 4, implications for practice, such as the practical application of factor investing and the risks involved, are analyzed. The last chapter summarizes and discusses the findings.

2. Theoretical foundations

2.1. Efficient market hypothesis

The idea of efficient markets, which can be retraced to Fama (1970, p. 1),[4] assumes that "a market in which prices always fully reflect available information is called effi-

[4] See also Fama et al. (1969).

cient."[5] Consequently, it should be impossible to achieve long-term abnormal returns through individual stock picking based on fundamental analysis or technical analysis of stock trends because prices already reflect all available information. As a result, only tomorrow's news affects stock prices. Assuming that news is unforeseeable, the efficient market hypothesis (EMH) postulates that stock prices follow a random walk. This means that the market can only be outperformed due to luck and achieving excess returns is impossible in the long run. Hence, if markets are fully efficient, beating the market can only be achieved by adopting more risk.

The concept of the EMH has been contradicted by some famous economists. For example, Shiller (2003) argues that some but not all market participants behave irrationally, and that this behavior can lead to mispricing and herd behavior on stock markets. In this context, historical events that demonstrate such behavior include the dot-com bubble in the early 2000s or the financial crisis of 2007–2008. Additionally, many anomalies, which could not be explained through the EMH, emerged over time. Grossman and Stiglitz (1980) argue that in efficient markets, investors are not encouraged to spend money on collecting information in order to generate abnormal returns because prices already reflect all available information. However, investors obviously spend enormous amounts on research to gather information.

2.2. Arbitrage pricing theory

The APT of Ross (1976) is an extension of the CAPM framework that pursues a similar concept as it assumes a linear relation between risk and returns. Unlike the CAPM, it allows multiple factors to price systematic risk and to consequently explain stock returns. However, the APT does not explicitly state which factors affect stock returns. Another fundamental difference between the APT and the CAPM lies in the assumption of the APT that there is absence of possibilities to profit from arbitrage (in equilibrium). This assumption varies from the demand and supply equilibrium idea of the CAPM. The APT captures expected returns by the following equation (1):

$$\mathrm{E}(r_i) = r_f + \beta_{i1} f_1 + \beta_{i2} f_2 + \cdots + \beta_{in} f_n + \epsilon_i \tag{1}$$

[5] Fama (1970) redefines the word "fully" by distinguishing between three forms of market efficiency, namely weak, semi-strong, and strong.

The expected return $\mathrm{E}(r)$ on portfolio i is determined by the risk-free rate r_f, the risk premium of factor f_n, and the sensitivity β_n of portfolio i to risk factor f_n. ϵ_i is the error term whose expected value is zero and which represents the unsystematic risk of portfolio i. Possible factors include macroeconomic indicators (e.g., inflation, unemployment rate, interest rates, and economic growth; see, e.g., Chen et al., 1986) or firm-specific characteristics (e.g., firm size, book-to-market equity (*B/M*) and asset growth; see, e.g., Fama and French, 1993, 2015, 2018) (Khan, 2011). With the idea that multiple and interchangeable factors capture expected stock returns, the APT pioneered multi-factor thinking in asset pricing models and also represents the fundament of multi-factor investing or combining different factors within one portfolio.

2.3. Stock market anomalies

For many years, the CAPM with the beta factor was the predominant model to explain expected returns. The CAPM is consistent with the EMH as it assumes that most market participants act rationally. Over time, various researchers (e.g., Basu, 1977; Banz, 1981; Jegadeesh and Titman, 1993; Titman et al., 2004) demonstrated that firms with exposure to other factors[6] such as value, size, momentum, and investment generate abnormal returns relative to the CAPM. These empirical results, which are actually abnormal returns of long-short zero-cost portfolios, are referred to as anomalies in the academic literature because they cannot be explained by traditional theory of asset pricing (e.g., the CAPM). Anomalies indicate that either the stock market is inefficient (profit opportunities) or traditional asset pricing models such as the CAPM and the three-factor model of Fama and French (1993) are inadequate in explaining the cross-section of expected returns (Schwert, 2003). Meanwhile, there are hundreds of return predictive signals in the literature, as highlighted by Harvey et al. (2016) and Harvey and Liu (2019). Comprehensive empirical evidence of the cross-sectional patterns of these anomalies is the foundation of factor investing. However, after the anomalies are published and become known, many of them, especially those in U.S. stock markets, seem to disappear (e.g., Schwert, 2003; McLean and Pontiff, 2016; Jacobs and Müller, 2020).

[6] Unless otherwise stated, the factors and anomalies in this paper are referred to stock returns.

2.4. Behavioral finance

There is vast empirical evidence (e.g., Shefrin and Statman, 1985; Odean, 1998; Poteshman and Serbin, 2003) about irrational investor behavior that challenges the EMH.[7] Basically, the behavioral finance view on the occurrence of stock market anomalies builds upon investor psychology and limits to arbitrage (Barberis and Thaler, 2003). Investors' psychology includes different cognitive and behavioral biases, such as overconfidence, optimism, representativeness, anchoring, belief perseverance, availability biases, and ambiguity aversion.[8] In this sense, irrational behavior of investors in terms of processing information leads to under- or overreaction, generating mispricing (Barberis et al., 1998; De Bondt and Thaler, 1985; Lakonishok et al., 1994). Limits to arbitrage prevent rational investors from fully taking advantage of these over- or undervalued stocks, which could consequently generate rather persistent stock market anomalies. Examples of limits to arbitrage include fundamental risk, noise trader risk, and implementation costs (Barberis and Thaler, 2003).

2.5. Factor models

Factor models use one or more factors to predict returns, generate estimates of abnormal returns, estimate variance and covariance of returns, or calculate costs of equity (Zivot and Wang, 2003). A distinction can be made between single-factor models such as the CAPM and multi-factor models such as the Fama and French (1993, 2015, 2018) three-, five- and six-factor models, and the Hou et al. (2015, 2018a) *q*-four and q^5-factor models. Multi-factor models evolved as a reaction to stock market anomalies which contradict the CAPM. Thus, multi-factor models are another way to deal with anomalies. Furthermore, multi-factor models can be separated into macroeconomic (e.g., Chan et al., 1985; Chen et al., 1986), fundamental (e.g., Fama and French, 1993, 2015, 2018), and statistical models (Connor, 1995). In addition, there are factor models that cannot be assigned to one of these three categories as they contain characteristics from more than one category (Mondello, 2013). Risk factors in macroeconomic factor models constitute variables that capture, for example, changes in inflation or gross domestic product (GDP). Fundamental factor models include factors (e.g., firm size, *B/M*, asset growth) that describe firm-specific characteristics which may proxy for systematic risk and factors that affect stock returns. Statistical

7 See, in this respect, also the prospect theory of Kahneman and Tversky (1979).
8 See Barberis and Thaler (2003) for an explanation of these cognitive and behavioral biases.

factor models derive pervasive factors from cross-sectional or time-series panel data and are used to estimate the variance and covariance of historical returns (Connor, 1995).

2.6. Factor investing and smart beta

Theoretically, factor investing or smart beta might be retraced to the APT, which describes stock returns through multiple factors. However, the APT does not state how many and which factors are appropriate. Researchers who initially identified the first factors and hence factor-based investment strategies include, among others, Basu (1977), Banz (1981), Fama and French (1993), and Jegadeesh and Titman (1993). Therefore, factor investing can be seen as an investment approach that selects stocks based on certain characteristics or factors such as value (*B/M*), small size (market capitalization), momentum (price or earnings trends), and/or investment (asset growth) which are empirically associated with abnormal returns.[9] The term “risk factor” implies that the factor compensates investors for taking higher systematic risk. This paper also examines factors (e.g., momentum) which appear most likely due to behavioral biases from investors, instead of a compensation for systematic risk. This paper uses the terms “risk factor,” “factor,” and “factor premium” interchangeably. Moreover, factor investing is related to smart beta. Although there is a difference between factor investing and smart beta, these two terms are often used interchangeably. In contrast to smart beta, which mostly includes passively managed factor-based ETFs (single or multi-factor long-only strategies), factor investing also refers to actively managed long-only and long-short single or multi-factor strategies. Throughout this paper, the terms “factor investing” and “smart beta” are used interchangeably.

Given the plethora of potential factors that try to explain stock returns and which could simultaneously be exploited through investment strategies, the question arises as to why such a vast number of factors have actually emerged. Therefore, the next chapter begins by taking a step back and examining the reasons for the expansion of the factor zoo.

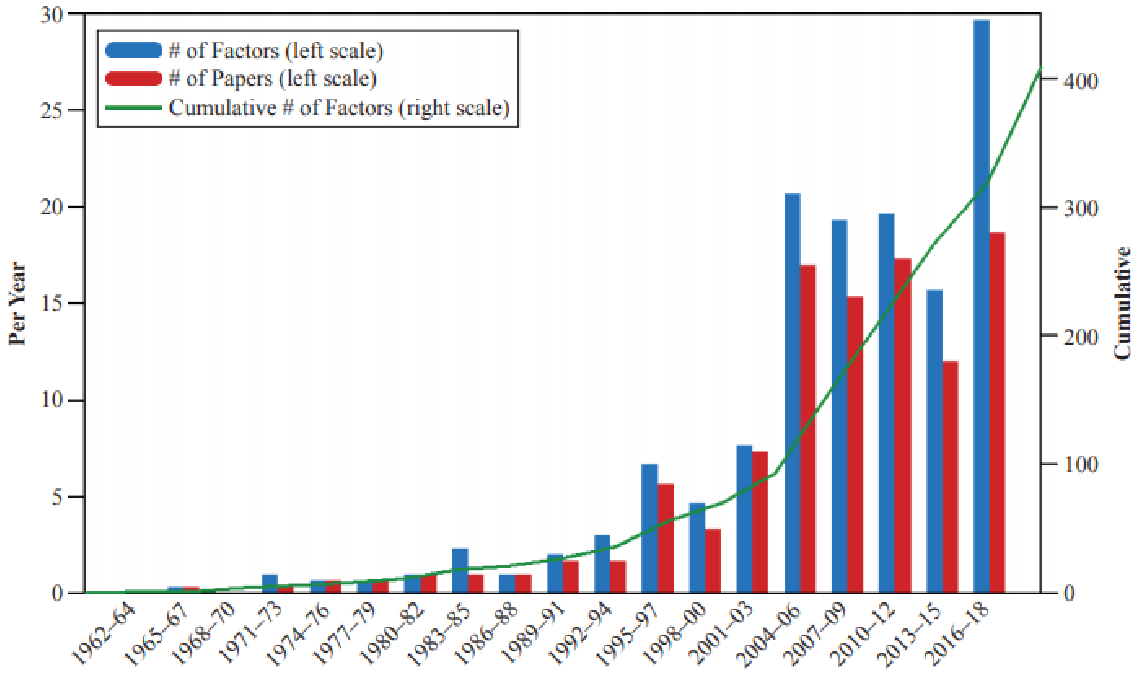
[9] More specifically, characteristics and factors are not exactly the same as characteristics are proxies for factors that drive stock returns.

3. The factor zoo

3.1. Reasons for the inflationary growth of factors

The probably most well-known study that has conducted a census of the factor zoo is by Harvey et al. (2016). They list 316 factors that were published in top academic journals.[10] Harvey and Liu (2019) update the list prepared by Harvey et al. (2016) and provide a census of 382 factors, as shown in Figure 1.[11] Between 1980 and 2003, when papers from Fama and French (1992), Carhart (1997), and Pastor and Stambaugh (2003) began to arouse interest in examining the cross-section of expected returns, only about three factors per year were published. In contrast, Figure 1 illustrates that the production of factors escalated in the last 15 years, when over 200 factors were "uncovered."

FIGURE 1
Factor production from top academic journals, through January 2019



Source: Harvey and Liu (2019, p. 2)

10 Green et al. (2013) describe and analyze 330 return predictive signals, McLean and Pontiff (2016) examine 97 anomalies, Yan and Zheng (2017) explore 18,000 signals from financial statements (they use 240 accounting variables and create 76 signals for each variable which results in such a high number), and Hou et al. (2018b) replicate 447 anomalies.

11 Note that Figure 1 does not include working papers, which at least document additional 70 factors (Harvey and Liu, 2019).

Given that hundreds of potential factors have been tested to explain the similar or identical cross-section of returns, the issue of multiple testing or data mining arises. However, this issue has been ignored in financial economics for many years. As a result, false cutoff levels for statistical significance have been used, as emphasized by Harvey et al. (2016) and Chordia et al. (2019). This has led to the detection of a huge number of factors many of which are most likely found by fluke and do not provide independent and reliable information.

As noted by Harvey (2017), other reasons for the strong rise of factors in the literature are different forms of *p*-hacking.[12] Academic journals tend to publish only the articles with positive or the most significant results (Fanelli, 2013). Consequently, academics are tempted to only report the most significant discovery of, for example, thousands of conducted correlations. This issue is also known as publication bias. The American Statistical Association (2016) has warned against such a procedure, which additionally contradicts the American Finance Associations' (2016) *Code of Professional Conduct and Ethics*. In this context, publishing only the correlations of 10 out of 11 examined variables is another form of *p*-hacking. Regrettably, such a practice is also common in financial economics, as noted by Harvey (2017).

Furthermore, *p*-hacking is an issue if different statistical approaches (e.g., Fama and MacBeth, 1973 vs. panel data regression or linear probability vs. logit or probit) are tried and only the one with the most significant result is reported. Additionally, Harvey (2017) identifies data manipulation and the exclusion of outliers or certain time periods when the results are not as significant as *p*-hacking. Another reason, which has similar consequences to *p*-hacking, is that many studies overweight microcaps, as detected by Hou et al. (2018b). Fama and French (2008) point out that microcaps comprise only 3% of the market capitalization of U.S. stocks,[13] but they concurrently constitute about 60% of all stocks. After excluding microcaps, Hou et al. (2018b) show that 64% of their 447 replicated anomalies are insignificant at the 5% level. These results are

[12] *P*-hacking (also known as data dredging, data fishing, data snooping, data mining, or multiple testing) is basically the use of data mining to find a pattern that can be presented as statistically significant although the result is likely to be a false discovery and thus not reliable. Therefore, this paper subsumes *p*-hacking as a form of data mining. Furthermore, this paper uses the terms "data mining" and "multiple testing" interchangeably.

[13] Refers to the market capitalization of all stocks from the stock exchanges NYSE, AMEX, and NASDAQ.

consistent with the concerns from Harvey (2017) that many factors have been *p*-hacked. Further studies that recognize the multiple testing bias in the literature of stock market anomalies are provided by Lo and MacKinlay (1990), Sullivan et al. (1999, 2001), White (2000), Linnainmaa and Roberts (2018), and Chordia et al. (2019). Finally, engaging data mining has become much easier in recent years as the costs for data collection and estimation drastically decreased (Harvey et al., 2016).

Apart from data mining or *p*-hacking, the disregard of transaction costs in academic studies has led to the identification of many false factors. Accounting for transaction costs leads to the conclusion that many factors do not produce significant results or do not generate abnormal returns anymore as highlighted by Novy-Marx and Velikov (2016), Chen and Velikov (2019), and Patton and Weller (2019).

To sum up, different forms of *p*-hacking or data mining and the resulting multiple testing bias have very likely led to the proliferation of the factor zoo and the discovery of many factors that are in fact false, as pointed out by Harvey et al. (2016), Harvey (2017), Hou et al. (2018b), and Linnainmaa and Roberts (2018).[14] If the issues of data mining or *p*-hacking were taken into account earlier, the factor zoo had most likely not grown to such a massive extent. For this reason, the next chapter discusses some approaches that help to deal with the multiple testing problem and to rule out spurious factors.

3.2. Ways to deal with multiple testing

3.2.1. Out-of-sample validation

Basically, there are two ways to deal with multiple testing: out-of-sample validation[15] and an adjustment of the statistical frameworks that takes multiple testing into account (Harvey et al., 2016). The cleanest way to address multiple testing and to detect spurious factor is to do an out-of-sample examination. The papers by, for example, Jaffe et al. (1989) and Wahal (2019) investigate the returns of anomalies prior to the initial sample. In particular, Jaffe et al. (1989) find significant size and value effects across

[14] Note that other researchers (e.g., Yan and Zheng, 2017; Engelberg et al., 2018; Wahal, 2019; Jacobs and Müller, 2020) contradict the argument that most anomalies are the product of data mining.

[15] Out-of-sample observations include tests within different time periods (e.g., before or after the sample of the initial study or post-publication), different regions (stock markets across the globe) or different asset classes (e.g., individual stocks, equity index futures, currencies, bonds or commodity futures).

their whole sample period from 1951 to 1986. Wahal (2019) observes a significant value and profitability premium in the 1940–1963 period. Contrarily, the investment premium does not appear in the sample from Wahal (2019), neither if the sample is extended back to 1926.

Jegadeesh and Titman (2001) study the returns of a long-short momentum strategy in the eight years after the sample from Jegadeesh and Titman (1993) who originally found the momentum anomaly and document similar significant momentum profits as in the earlier time period. McLean and Pontiff (2016) conducted a prominent study that explores 97 anomalies by using a slightly different approach than Jegadeesh and Titman (2001). They study the returns (long-short zero-cost portfolios) of 97 anomalies after the period from the initial sample but before and after publication. The findings of their study show that the average returns decline about 26% post-sample (after the initial sample, but before publication) and 58% post-publication. More specifically, they argue that about 26% of the post-publication effect is attributable to data mining and the remaining 32% of the initially found return might be explained by arbitrage trading.

In contrast to the previously discussed papers, Schwert (2003) and Linnainmaa and Roberts (2018)[16] examine both pre-sample (the time period before the one in the initial study) and post-sample data of anomalies. Schwert (2003) analyzes well-known anomalies (e.g., size, value and momentum) in different sample periods and finds that, in contrast to momentum, the size and value factor seem to have lost their explanatory power after their publication. However, Asness et al. (2018) find evidence that contradicts the results from Schwert (2003) concerning the size anomaly. They control for the quality and junk of a firm, subsequently finding a significant size effect which is persistent through time and across international stock markets. Linnainmaa and Roberts (2018) study 36 accounting-based anomalies in the U.S. Their findings indicate that a large number of the investigated anomalies, including the investment anomaly, are most likely statistical artifacts in consequence of data mining.

Another way to conduct out-of-sample testing is an examination of factors across international markets (particularly outside the U.S.) and different asset classes. Karolyi (2016) points out that only 16% of the empiric studies published in the top four finance journals investigate markets outside the U.S. Hence, Karolyi (2016) suggests the in-

[16] Linnainmaa and Roberts (2018) examine data of all CRSP firms that goes back to the year 1918.

vestigation of global financial data in order to address the problem of data mining or multiple testing. Studies that carry out examinations of anomalies in international markets are done by Chan et al. (1991), Fama and French (1998, 2012, 2017), Griffin (2002), Chui et al. (2010), Asness et al. (2013), Titman et al. (2013), Watanabe et al. (2013), and Frazzini and Pedersen (2014).[17] The overall finding of these papers is that prominent anomalies such as size, value, momentum, profitability, investment (or asset growth), and beta arise, with some exceptions, also in stock markets across the globe. Moreover, Moskowitz et al. (2012), Asness et al. (2013), Frazzini and Pedersen (2014), and Koijen et al. (2018) study the appearance of anomalies within various asset classes. They find empirical evidence for the patterns of their examined anomalies (momentum, value, beta, and carry[18]) across different asset classes.

Motivated by the study of McLean and Pontiff (2016), Jacobs and Müller (2020) use an out-of-sample approach that combines the test of 241 anomalies pre- and post-publication, and across 39 international stock markets. Their findings are only consistent with the results in McLean and Pontiff (2016) regarding the post-publication decline of anomalies in the U.S. However, the evidence in Jacobs and Müller (2020) suggests that the return predictability of anomalies in international markets is strong in both post-sample and post-publication periods, which contradicts the overall conclusion from McLean and Pontiff (2016) that anomalies in general lose explanatory power after dissemination.

Although the out-of-sample approach is a powerful tool to rule out spurious factors, it contains one major disadvantage, namely the non-applicability in real time. “Real time” evaluations could nonetheless be achieved by holding out some data. The prob-

[17] A global examination of the size anomaly is conducted by Chan et al. (1991), Fama and French (1998, 2012, 2017), Griffin (2002), and Asness et al. (2018).
A global examination of the value anomaly is conducted by Fama and French (1998, 2012, 2017), and Asness et al. (2013).
A global examination of the momentum anomaly is conducted by Chui et al. (2010), Fama and French (2012), Moskowitz et al. (2012), and Asness et al. (2013).
A global examination of the profitability anomaly is conducted by Fama and French (2017).
A global examination of the investment anomaly is conducted by Titman et al. (2013), Watanabe et al. (2013), and Fama and French (2017).
A global examination of the beta anomaly is conducted by Frazzini and Pedersen (2014).

[18] “Carry” has usually been a research object in currency markets. Koijen et al. (2018) define the carry of an asset as the futures (or synthetic futures, if not available) return with the assumption that the price remains unchanged. Thus, the expected return of an asset consists of its carry plus its anticipated price appreciation.

lem with such a procedure is that all the data are admittedly accessible for other researchers (Harvey et al., 2016). Therefore, future data is required for a genuine out-of-sample test. In order to conduct such an out-of-sample test, we need to wait years for the required data. For this reason, Harvey et al. (2016) and Green et al. (2017), among others, propose multiple testing frameworks that immediately help to find out whether an uncovered anomaly is true (statistically significant) or note.

3.2.2. Statistical multiple testing frameworks

Decades ago, the beta factor in the CAPM was found to be an explanatory variable of the cross-section of expected returns. The *t*-statistic of this factor was 2.57 and thus in excess of the usual cutoff level of 2.0, as reported by Fama and MacBeth (1973). Since then, hundreds of variables have been tried to explain average returns. Given this plethora of tests, the usual cutoff level of 2.0, which is actually appropriate for a single test, seems to be too low. In this context, the main issue is that all of these tests explore a similar cross-section of returns. Even though these studies are published at different times, the majority of the examined sample periods overlap each other, as highlighted by Harvey et al. (2016).

In statistics, researchers (e.g., Tukey, 1951; Scheffé, 1959) have recognized the problem of multiple testing over 65 years ago. In medical literature, multiple testing is also a vibrant field of research (e.g., Ioannidis, 2005; Farcomeni, 2007). However, in the finance literature, multiple testing has largely remained ignored for a long time. The research (e.g., Shanken, 1990; Ferson and Harvey, 1999; Boudoukh et al., 2007) that takes multiple testing into account mostly uses the adjustment of Bonferroni, which simply multiplies the *p*-value with the test tried.

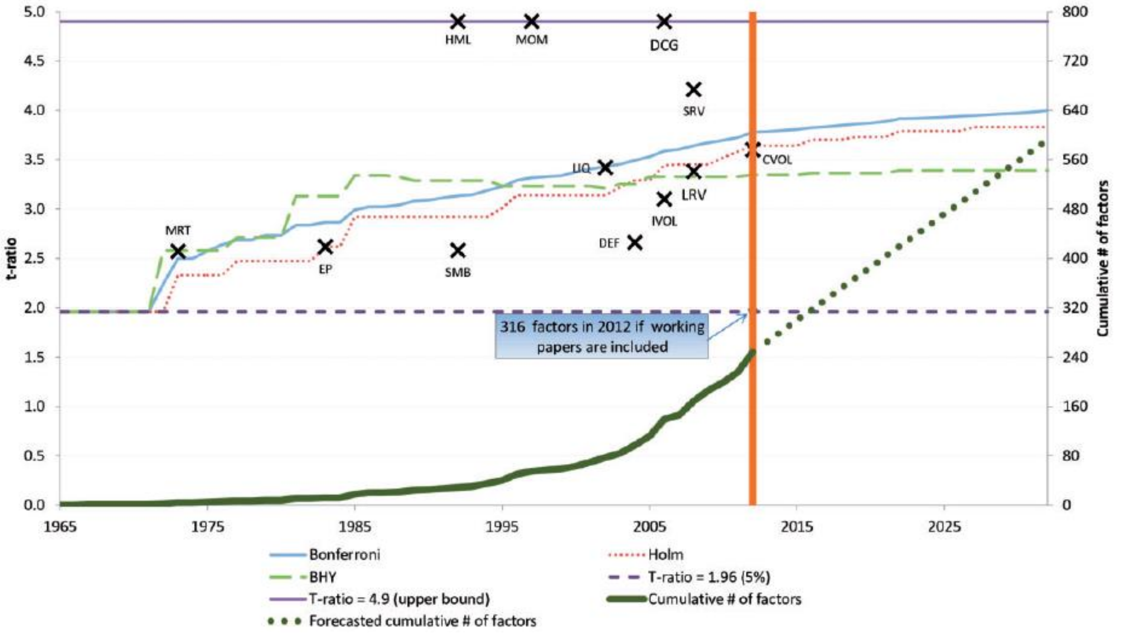
Harvey et al. (2016) fill this gap in the literature and discuss three possible (Bonferroni, Holm, and Benjamini, Hochberg, Yekutieli) *p*-value adjustments. Consequently, they propose a new statistical framework that takes multiple testing into account. They apply the three *p*-value adjustments to their gathered factors and assume that the total number of tried factor tests is available. As they note, this hypothesis is incorrect as they cannot account for all the insignificant factors that have been tried and never been published. However, they try to address this problem of missing data with different methods. In a next step, they transform the three obtained benchmark *p*-values (Bonferroni, Holm, and BHY) in *t*-statistics. Figure 2 visualizes their multiple testing framework which contains the three benchmark *t*-statistics. In contrast to BHY, the *t*-statistics which are adapted from Bonferroni and Holm consistently increase with the number of discovered factors. Bonferroni implied *t*-statistic starts at 1.96, reaches 3.78 in 2012, and ends at 4.00 in 2032. Holm's adjusted *t*-statistics are always a little lower than the Bonferroni *t*-statistics, although the differences are marginal. In comparison, the BHY-adjusted *t*-statistic does not consistently increase as it rises by 1985 and slightly falls afterward. After 2010 it stabilizes at 3.39.

Furthermore, Figure 2 re-evaluates the statistical significance (*t*-statistics) of some prominent factors compared to the benchmark *t*-statistics. Five of these factors (HML, MOM, DCG, SRV and MRT) are significant compared to all three benchmark *t*-statistics. EP, LIQ, and CVOL partly exceed the adjusted thresholds, and the other factors never surpass the suggested benchmark *t*-statistics.[19] Harvey et al. (2016) also apply their statistical framework to the whole sample of 296 factors collected in their paper. This examination leads Harvey et al. (2016, p. 37) to the conclusion that "[…] many of the factors discovered in the field of finance are likely false discoveries: of the 296 published significant factors, 158 would be considered false discoveries under Bonferonni, 142 under Holm, 132 under BHY (1%), and 80 under BHY (5%).

[19] According to Harvey et al. (2016, p. 25), the factors in Figure 2 include: "[…] MRT (market beta; Fama and MacBeth, 1973), EP (earnings-price ratio; Basu, 1983), SMB and HML (size and book-to-market; Fama and French, 1992), MOM (momentum; Carhart, 1997), LIQ (liquidity; Pastor and Stambaugh, 2003), DEF (default likelihood; Vassalou and Xing, 2004), IVOL (idiosyncratic volatility; Ang et al., 2006); DCG (durable consumption goods; Yogo, 2006), SRV and LRV (short-run and long-run volatility; Adrian and Rosenberg, 2008), and CVOL (consumption volatility; Boguth and Kuehn, 2012)."

FIGURE 2

Multiple testing framework: Adjusted *t*-statistics, 1965–2032



Source: Harvey et al. (2016, p. 25)

The overall conclusion of the previously discussed literature in Chapters 3.1 and 3.2 is that many detected factors are probably false due to data mining or multiple testing. Moreover, out-of-sample tests and an increase of the hurdle for statistical significance have been discussed as possible ways to deal with the issue of multiple testing. However, even after controlling for multiple testing, Harvey et al. (2016) still find approximately 150 factors out of about 300 to be significant predictors of expected returns. Moreover, about 70% of these factors only have a Sharpe ratio of less than 0.5 p.a. (Harvey et al, 2016). Therefore, the issue that arises is how to choose reliable and independent factors out of this subset, which constantly grows. The next chapter proposes some methods that attempt to give an answer to this question.

3.3. Methods for identifying independent factors

Motivated by the huge number of factors that potentially contribute to explain the cross-section of expected returns, controlling the proliferation of the factor zoo has become a fundamental subject in the asset pricing literature. The main task of this strand of the literature is to determine whether an existing or newly discovered factor possesses additional explanatory power for stock returns, taking into account the hundreds of factors that already contribute to explain the cross-section of expected returns (Feng et al., 2020). In other words, which factors provide reliable and independent information about stock returns, and which factors are redundant and just measure the same phenomenon in a different way (Cochrane, 2011)? To examine this topic, the literature suggests different statistical methods. Harvey and Liu (2018) propose a bootstrap method for identifying factors. Based on a performance metric suggested by Barillas and Shanken (2017), Barillas and Shanken (2018) and Fama and French (2018) examine the maximum squared Sharpe ratio to ascertain whether a new factor improves the performance of an existing asset pricing model. Freyberger et al. (2018) suggest a nonparametric method, namely a group least absolute shrinkage and selection operator (LASSO) procedure, for selecting independent return predictive characteristics. Kozak et al. (2020) also use LASSO-style estimation to construct characteristic-sparse stochastic discount factors (SDFs) and find that these SDFs are not adequately able to explain the cross-section of expected returns. Feng et al. (2020) suggest a combination of double-selection LASSO of Belloni et al. (2014) and two-pass regressions of Fama and MacBeth (1973) to systematically select a benchmark model for evaluating new factors. Furthermore, Hwang and Rubesam (2018) use a modified Bayesian variable selection method for identifying parsimonious factor models to explain stock returns. Finally, Gu et al. (2019) provide an evaluation of different machine learning methods to determine return predictive signals. Additionally, Kelly et al. (2019) propose an instrumented principal component analysis to study factors in the cross-section of expected returns. A comprehensive discussion of the aforementioned techniques is beyond the scope of this paper.[21]

After applying the methods from the above-mentioned papers on the factor zoo, it is revealed that these different techniques also detect different factors as independent and reliable return predictors. In general, the results may be surprising as these methods identify only a handful of factors which have mainly been discovered for decades. For example, Harvey and Liu (2018) find the original market factor of Sharpe (1964) as the most dominant factor, and size, profitability and value as further factors that marginally contribute to explain the cross-section of expected returns. Furthermore, Feng et al. (2020) test the estimates for SDF loadings for factors introduced from 2012 until 2016. They find profitability (the version of Fama and French, 2015

[21] Besides the abovementioned studies, I refer to Bryzgalova (2016), Green et al. (2017), and Pukthuanthong et al. (2019) for further studies.

and Hou et al., 2015), investment of Hou et al. (2015), quality minus junk of Asness et al. (2019), and the intermediary capital factor from He et al. (2017) as independent factors. Lastly, Gu et al. (2019) detect variations of momentum, liquidity, and volatility as dominant return predictors.

This chapter provided an overview of different statistical approaches for choosing reliable and independent factors. Moreover, the results of applying some of these techniques on a broad set of factors have been given. Parsimonious factor models, which are discussed in the next chapter, attempt to explain the cross-section of stock returns by combining as few as possible factors. Hence, these models also provide a subset of robust factors, which can be later compared to the factors proposed by practitioners to evaluate the knowledge transfer between science and practice.

3.4. Asset pricing models

3.4.1. Applied factors

In contrast to the previously discussed literature that seeks to reduce the enormous number of factors, asset pricing models somewhat experienced a slightly contrary development. Research within this field initiated with a single-factor model (CAPM) and currently evolved up to six factors that are combined in a model (e.g., Barillas and Shanken, 2018; Fama and French, 2018). The factor portfolios used in these factor models can be interpreted as factor investing strategies that are highlighted by researchers. However, the main goal of asset pricing models is to explain the cross-section of expected returns with as few factors as possible. Research has taken different approaches to develop parsimonious models with firm-specific factors.[22] One approach builds upon rational asset pricing and is consistent with the EMH. Prominent examples include the Fama and French (1993, 2015) three- and five-factor models (FF3, FF5), and the Hou et al. (2015, 2018a) *q*-four and q^5-factor models (H*Q*4, H*Q*5). Another approach develops models that are primarily statistical in nature, such as the six-factor model of Barillas and Shanken (2018) (BS6) and the six-factor model of Fama and French (2018) (FF6). Finally, a third approach constructs factor models with behavioral factors that capture mispricing. Examples include the Stambaugh and

[22] Besides asset pricing models with factors as proxies for firm-specific characteristics, another stream of the literature (e.g., Chan et al., 1985; Chen et al., 1986; Shanken and Weinstein, 2006) explores asset pricing models containing macroeconomic factors.

Yuan (2017) four-factor model (SY4) and the Daniel et al. (2020) three-factor model (D3).[23] Table 1 summarizes the different factors used in the aforementioned asset pricing models.

TABLE 1
Factors in asset pricing models

Based on rational asset pricing			
FF3	FF5	H*Q*4	H*Q*5
Market	Market	Market	Market
Size	Size	Size	Size
Value	Value	Investment	Investment
	Investment	Profitability	Profitability
	Profitability		Expected growth
Statistical in nature		Inclusion of behavioral factors	
BS6	FF6	SY4	D3
Market	Market	Market	Market
Size	Size	Size	Long-horizon mispricing
Value	Value	Two mispricing factors	Short-horizon mispricing
Investment	Investment		
Profitability	Profitability		
Momentum	Momentum		

Source: Author's own illustration

As demonstrated in this chapter, the literature is full of different factor models that attempt to explain stock returns. But which factor model performs best and how are such comparisons conducted in the literature? To answer this question, the next chapter at first investigates the economic foundation of the FF3, FF5, and H*Q*4. Subsequently, papers that conduct performance comparisons of the abovementioned asset pricing models are examined.

3.4.2. Theoretical foundation

Harvey et al. (2016) highlight that the multiple testing issue, which always clouds statistical inference, is also prevalent within the search for the best factor model out of a vast number of potential factor models and factors that could be included in a model.

[23] A further behavioral model that is able to explain most of the well-known stock market anomalies is the anchoring-adjusted CAPM (ACAPM) of Siddiqi (2018) which adapts the CAPM for the anchoring and adjustment heuristic of Tversky and Kahneman (1974). In contrast to the SY4 and D3, the ACAPM intrinsically includes behavior instead of indicators or factors that measure behavior.

In that respect, Fama and French (2018) suggest that the number of factor models compared has to be limited in order to perform reliable comparisons. Besides out-of-sample tests of factor models, they suggest that the set of considered models can be restricted to models that are motivated by theory. The theoretical rationale behind factor models is therefore an important topic and concurrently a possible explanation of why certain factors work. The CAPM is a perfect example for a model that is derived from theory. It describes a linear relation between expected returns and risk (beta). This implies that higher returns can only be obtained by taking more risk.

Fama and French (1993) develop the three-factor model by extending the CAPM with a value and a size factor. The explanatory power of these two factors, which have been empirically discovered by Basu (1977) and Banz (1981), was the starting point for the extension of the CAPM. Fama and French (1993) motivate the size and value factor with the theoretical argument that they are priced risk factors which capture systematic risk.[24] Thus, the three-factor model does not doubt the idea behind the CAPM that higher returns can only be achieved by taken higher systematic risk. Moreover, Fama and French (1993) argue that the size and value factors are related to fundamentals, such as earnings. Ball (1978) earlier stated that the relation from average returns and *B/M* exists due to the discount rate effect. Under the assumption that current fundamentals reasonably represent expected cash flows, low price-to-fundamental ratios (e.g., high *B/M*) should indicate higher expected returns. Fama and French (1993) argue that the positive *B/M*-return relation might be explained by a common risk factor which has its source in the relative profitability. As shown by Fama and French (1992), small firms are more fragile in certain market phases (e.g., in the 1980s). This suggests that size is the source of a common risk factor which might explain the negative firm size–return relation.

Various studies have detected further anomalies, in particular investment (Titman et al., 2004; Cooper et al., 2008; Aharoni et al., 2013) and profitability (Novy-Marx, 2013), that are left unexplained by the FF3. Therefore, Fama and French (2015) decided to augment the initial three-factor model with investment and profitability factors. In contrast to the justification of the new factors in the three-factor model, Fama and

[24] This perspective was challenged shortly thereafter, for example, by Lakonishok et al. (1994), who stated that the value anomaly not necessarily captures risk but arises instead due to behavioral biases such as extrapolating past growth into the future.

French (2015) do not anymore motivate the investment and profitability factor with risk-based explanations. The reason for this is probably the ambiguity whether the drivers of these two factors are risk-based or behavioral. On a risk-adjusted basis, high-profitability firms which are basically less risky generate higher returns than low-profitability firms which tend to be riskier (Asness et al., 2019). This fact seems to contradict risk-based explanations. Titman et al. (2004) find evidence that is consistent with the hypothesis that the negative capital investment-return relation occurs because of the underreaction to overinvestment and empire building. Cooper et al. (2008) also argue that the driver of the asset-growth (or investment) anomaly rather lies in behavioral biases of investors, who overextrapolate past performance into the future. However, Titman et al. (2013) and Watanabe et al. (2013) find evidence from international stock markets that is inconsistent with behavioral-based explanations of the asset growth effect.

Instead, by means of risk-based explanations, Fama and French (2015) use a rewritten dividend discount model (DDM)[25] as an umbrella theory to motivate why value, investment, and profitability are related to expected returns.[26] According to the DDM, the stock value M_t results through discounting the expected dividend E of the stock with r, which is the expected stock return or, more specifically, the rate of return on expected dividends. In equation (2), which is a slightly manipulated version of the DDM, $Y_{t+\tau}$ is earnings at time t and $dB_{t+\tau} = B_{t+\tau} - B_{t+\tau-1}$ is the change in book equity.[27]

$$\frac{M_t}{B_t} = \frac{\sum_{\tau=1}^{\infty} \mathrm{E}(Y_{t+\tau} - dB_{t+\tau})/(1+r)^{\tau}}{B_t} \tag{2}$$

With regard to expected stock returns, Fama and French (2015) derive three predictions from equation (2). Basically, we have to ask how the expected return r has to adjust in order to achieve equilibrium in equation (2), provided that everything else remains equal and that either *B/M* (value), expected earnings (profitability), or change

[25] Further information regarding the DDM and valuation theory is provided by Preinreich (1938), Gordon and Shapiro (1956), Miller and Modigliani (1961), and Ohlson (1995).

[26] However, Fama and French (2015) do not even attempt to declare whether the higher expected returns from firms with low investment or high profitability arise due to higher systematic risk or mispricing.

[27] Equation (2) is the DDM, in combination with clean surplus accounting divided by Bt, which is the book equity at time t. Under clean surplus accounting, the expected dividend E equals earnings at time t minus change in book equity (see Fama and French, 2006, 2015).

in book equity (investment) decreases or increases. First, if everything in equation (2) except *B/M* and expected return is fixed, a decrease of M_t (means a higher *B/M*) results in an increase of r (a higher expected return). Second, we fix everything in equation (2) except future earnings and expected return. Then an increase in expected earnings (higher profitability) implicates higher expected stock returns. Finally, controlling for expected earnings and *B/M* in equation (2), an increase in book equity (higher investment) implies a decrease in expected return.

Hou et al. (2019) admittedly express four concerns about the theoretical foundation behind the FF5. First, Fama and French (2015) assume that the rate of return r in equation (2), which is also the expected return, is equal for every period. This assumption is inconsistent with the concept of time-varying expected returns as there are differences between the rate of return and the expected return for the next period. This is particularly the case for the profitability factor. Second, equation (2) states that value, profitability and investment are three separate factors. However, after adding profitability and investment factors to the FF3, the value factor in the FF5 becomes empirically redundant, which contradicts the theoretical rationale derived from equation (2). Third, equation (2) predicts a negative expected investment–rate of return relation. However, Hou et al. (2019) find a positive relation between expected investment and the expected return for the next period. Fourth and finally, Fama and French (2015) use past investment to predict expected investment. According to Hou et al. (2019), past investment is inadequate in predicting expected investment. Additionally, they argue that asset growth is an insufficient proxy for future book equity growth.

As the name indicates, the H*Q*4 and H*Q*5 are theoretically motivated by investment-based asset pricing, which, in turn, is based on the *q*-theory (Tobin, 1969; Cochrane, 1991a). *Q* represents the investment rate as the ratio between the market value of the company's share capital and the replacement cost of the company's share capital. According to the *q*-theory, it is useful that companies invest if $q > 1$ as the capital is more valuable inside the company. Conversely, companies should sell their assets if $q < 1$ as the capital is more valuable outside the company. Companies invest or disinvest until they reach an equilibrium where *q* approximates to one. If $q = 1$, there is consequently no need for investing or disinvesting.

Hou et al. (2015) use equation (3) to predict that, all else equal, stocks with high investment should generate lower expected returns and stocks with high expected profitability should generate higher expected returns. E_0 is the expectation at date 0, r_{i1}^S is the stock return at date 1 (or the discount rate), $\prod_{i1}$ is the marginal benefit of investment or the marginal product of assets (expected profitability) at date 1, and $a(I_{i0}/A_{i0})$ is the marginal cost of investment at date 0. Thus, equation (3) connects variations in expected returns with adjustments in investment and profitability.

$$E_0[r_{i1}^S] = \frac{E_0[\prod_{i1}]}{1+a(I_{i0}/A_{i0})} \tag{3}$$

In line with the *q*-theory, companies invest more if their marginal *q* is high or > 1. Low discount rates for investments (or low stock returns) imply high marginal *q* (high investment), which could therefore explain the negative investment-return relation. Equation (3) also provides an explanation for the positive profitability-return relation. In equation (3), the expected return is obtained by dividing expected profitability by investment to assets. Assuming low investment, high profitability means higher discount rates (high expected returns), which maintains low investment. Discount rates (expected returns), which are not high enough for counteracting the high expected profitability, would lead to higher present values of investments and, consequently, to higher investments. However, this deduction contradicts the assumption of given low investment. Given high investment, low expected profitability implies low discount rates (low expected returns). Discount rates (expected returns), which are not low enough for counteracting low expected profitability, would lead to lower present values of investments and, consequently, to lower investment. This deduction is inconsistent with the assumption of high investment.

3.4.3. Performance comparison

Given that there are over 400 potential factors which could be included in a factor model, the question which model performs best in terms of explaining average stock returns arises. To compare models with traded factors, the most important thing is to what extent each factor model is able to explain the particular factor-premiums from another factor model, as highlighted by Barillas and Shanken (2017, 2018). In order to conduct such a comparison, Hou et al. (2019) rely on factor spanning tests. This empirical approach is mostly identical to the methodology in Barillas and Shanken (2017, 2018) and Fama and French (2015, 2018).[28]

In particular, Hou et al. (2019) regress the individual factors from their H*Q*4 and H*Q*5 on all factors from the FF5 and FF6, SY4, BS6, and D3. The remaining alpha is the part of the individual factor that cannot be explained by the overall factors from the other factor model. Additionally, they test the null hypothesis (alphas of the factors in the regressions are together zero) by means of the test developed by Gibbons et al. (1989). The overall conclusion of their factor spanning regressions is that the seemingly different factor models are closely related (correlated) on empirical grounds. In detail, the H*Q*4 and H*Q*5 largely explain the FF5, FF6, BS6, replicated SY4, and replicated D3 (except its earnings factor).

The investigation of methods for determining reliable factors and the identification of factors used in asset pricing models is now completed. In order to evaluate the exchange of knowledge between science and practice, the following chapter ascertains factors considered by practitioners. This is followed by a factor comparison between science and practice.

4. Factor investing – from theory to practice

4.1. Factors considered by practitioners

As previously discussed, hundreds of possible factors are suggested by the academic literature (Harvey and Liu, 2019). In practice, the first issue concerns determining which factors might be able to generate future excess returns and which are actually applicable (Cazalet and Roncalli, 2014). Many scholars have detected that most factors are likely statistical artifacts because of data mining (Harvey et al., 2016; Harvey, 2017; Hou et al., 2018b; Chordia et al., 2019). Subsequently, Harvey and Liu (2018) and others[29] propose statistical methods to identify reliable factors.[30] However, these methods might be technically intricate for many practitioners. For this reason, Hsu et al. (2015) developed a simple three-step approach for assessing robust factors. They

[28] A further study that compares factor models is conducted by Ahmed et al. (2019).

[29] See, for example, Bryzgalova (2016), Barillas and Shanken (2017), Fama and French (2018), Freyberger et al. (2018), Hwang and Rubesam (2018), Gu et al. (2019), Kelly et al. (2019), Pukthuanthong et al. (2019), Feng et al. (2020), and Kozak et al. (2020).

[30] See Chapter 3.3, for a rough overview of these methods.

maintain that firstly, factors must be the subject of several academic papers over many years. This means that scholars have examined the factor with different data, debated possible explanations for the factor premium, and attempted to refute its effect. If there exist no further studies about a certain factor, this will be a sign that the factor is actually not significant and hence not appropriable (Hsu et al., 2015).

Second, factors should be persistent across countries and periods. Most factors were initially discovered in data from U.S. stock markets. Karolyi (2016) shows that 84% of the empirical studies in the field finance solely explore U.S. data. Therefore, testing factors outside the U.S. and in different sub-periods has to be done in order to examine their robustness. The effect of factors should occur in most countries, whether or not it is driven by systematic risk or bias in the behavior of investors. Otherwise, there is a high probability that the factor is only an artifact of U.S. data. True risk factors should be compensated with higher returns in different stock markets and it would appear odd if investors exhibited irrational behavior only in the U.S. (Hsu et al., 2015).

Third, factors should be robust across changes in definitions. Studies that do not report positive results or significant *t*-statistics are rarely published because they are generating few citations (Fanelli, 2013). Scholars recognize this and might consequently cherry-pick the factor definitions that generate the highest alpha with the largest *t*-statistic. That is why the definition of published factors could be a result of cherry-picking and small changes in the definition of the factor might shrink or even erase its alpha. Thus, it is indispensable to take cherry-picking into account. Making small changes to the factors' definition can help to identify cherry-picked factors. If a marginal change in the definition of the factor leads to a remarkable deviation of the factor premium, it will be highly probable that the factor has been cherry-picked or overfitted. Therefore, the average return of several definitions from one factor might be a more realistic estimate for future factor premiums (Hsu et al., 2015).

After applying this three-step heuristic to six prominent factors (value, momentum, low beta, illiquidity, size, and quality), Hsu et al. (2015) find that only value, low beta, and perhaps momentum and illiquidity seem to be robust and investable factors. One year later, Beck et al. (2016) extended the three-step heuristic from Hsu et al. (2015) with a trading cost perspective in order to obtain a more realistic and practical estimation of future factor returns. They applied the trading cost model of Aked and Moroz (2015) to examine the implications of trading cost on the profitability of factor replicating indices. After considering transaction costs, Beck et al. (2016) merely determine low beta and value as robust factors that can be exploited through ETFs. Moreover, they maintain that skilled managers might be able to also exploit illiquidity and momentum, which require more frequent trading.

Blitz and Van Vliet (2018) propose a conservative investment formula that is based on only three factors, namely low return volatility, high net payout yield, and strong price momentum. Motivated by the book by Van Vliet and De Koning (2017), which originally proposes the concept of low risk investing, they attempt to test this investment strategy. In summary, they find that 100 liquid U.S. stocks selected on the basis of these three factors generate a compounded return of 15.1% p.a. from 1929 until the end of 2016. According to Blitz and Van Vliet (2018), the returns of this conservative investment formula are positive in every decade and robust across European, Japanese, and emerging stock markets. Furthermore, this investment strategy does not require accounting data as it only uses past return and net payout yield data.

A glance at the investment practice reveals that the factors in Table 2 are currently deployed by some of the prominent providers of factor-based investment products.

TABLE 2

Overview of factors considered by the key players in the factor industry

BlackRock	FTSE Russell	Invesco	Research Affiliates
Value	Value	Value	Value
Size	Size	Size	Size
Momentum	Momentum	Momentum	Momentum
Quality	Quality	Quality	Quality
Low Volatility	Low Volatility	Low Volatility	Low Volatility
	Dividend Yield	Dividend Yield	Income

Robeco	S&P Dow Jones Indices	SSGA	Vanguard
Value	Enhanced Value	Value	Value
Momentum	Momentum	Size	Momentum
Quality	Quality	Momentum	Quality
Low Volatility	Low Volatility	Quality	Min volatility
	Dividend Yield	Low Volatility	Liquidity
	Equal Weight		

Source: Author's own illustration based on Perrins (2019, internet)

4.2. Choosing factors: Science vs. practice

Chapter 3.3 discussed which factors remain after applying different statistical approaches to the factor zoo. In addition, the factors that are included in the current asset pricing models have been detected in Chapter 3.4.1. Finally, Chapter 4.1 identified factors that are proposed by practitioners. In order to understand whether the exchange of knowledge between science and practice takes place successfully, this chapter contrasts the different factors suggested by academics and practitioners. Furthermore, this chapter examines the justifications for the considered factors from both scholars and practitioners.

Table 3 shows the factors included in the current asset pricing models and factors proposed by practitioners.[31] According to Table 3, the knowledge transfer between science and practice does not seem to be straightforward as scholars and practitioners propose different factors. But let us start with the factors where there is consensus between science and practice. The well-known size and value factors, which have been applied in factor models for nearly 30 years, are also widely implemented in practice. Moreover, momentum and the more recently discovered quality or profitability factors are applied as well in asset pricing models as in the investment practice. However, a momentum factor has only recently and somewhat reluctantly been added to the FF6 and hitherto not been included in the *q*-factor models of Hou et al. (2015, 2018a). Fama and French (2018) express their concerns regarding the momentum factor as this factor certainly is robust in out-of-sample tests but has no explicit theoretical motivation. Furthermore, they argue that the issue with such factors is the difficulty of assessing whether the effect will exist in the future if there is no model that determines the drivers that cause the pattern (Fama and French, 2018). However, regarding the momentum effect, there are, for example, models of over- and underreaction that attempt to explain the patterns behind this anomaly (Barberis et al., 1998; Daniel et al., 1998; Hong and Stein, 1999; Hong et al., 2000).

[31] For simplification purposes, in Table 3, factor models represent the factors proposed by researchers as these models mainly include, with some exceptions, the independent factors that are identified by the statistical methods in Chapter 3.3. Hwang and Rubesam (2018) and Gu et al. (2019) are the few studies that identify significantly other relevant factors than those used in prominent asset pricing models.

A major difference between science and practice concerns the low volatility factor. This factor is recommended and seen as investable by various practitioners (e.g., Blitz and Van Vliet, 2007; Hsu et al., 2015; Beck et al., 2016; Blitz and Van Vliet, 2018; Blitz et al., 2020) and offered by all of the eight key players in the factor industry listed in Table 2. Blitz and Van Vliet (2007) support their recommendation for investing in the low volatility factor with comprehensive empirical evidence that documents persistent patterns of this effect. Haugen and Heins (1975) initially detected the low volatility effect in U.S data. Blitz and Van Vliet (2007) find the anomaly appearing also in Japan and Europe, and show that the anomaly has even become stronger in the recent years of their 20-year sample. In addition, Blitz et al. (2013) observe a low volatility effect within emerging markets, Chen et al. (2018) and Han et al. (2018) in the Chinese stock market, and Joshipura and Joshipura (2016) in the Indian stock market. Further studies that document a low volatility effect in international stock markets are conducted by Baker and Haugen (2012), Frazzini and Pedersen (2014), and Walkshäusl (2014). Finally, the low volatility effect has also been documented within other asset classes such as corporate bonds (Carvalho et al., 2014; Houweling and Van Zundert, 2017), the options market (Falkenstein, 2009), and the cross-section of mutual fund returns (Jordan and Riley, 2015). Blitz et al. (2020) interpret all these arguments as clear evidence for the low volatility factor being a persistent and reliable driver of stock returns and not being a data fluke.

In contrast, none of the asset pricing models in Table 3 uses a low volatility factor to describe stock return, although there is a stream of literature that empirically documents these patterns over the long term within stock markets across the globe, different measures and other asset classes. Fama and French (2015, 2018) and Hou et al. (2015, 2018a) do not explicitly declare why they do not include a low volatility or low beta factor in their models. However, Hou et al. (2015) state that their profitability factor largely explains the Ang et al. (2006) idiosyncratic volatility effect which is closely related to the low volatility anomaly.[32] In addition, Fama and French (2016) show that the profitability and investment factors of their five-factor model capture returns associated with the low beta anomaly in Black et al. (1972) and Frazzini and Pedersen (2014), and the idiosyncratic volatility effect in Ang et al. (2006). Furthermore, Arnott

[32] However, unlike low volatility, which is measured by beta or volatility, idiosyncratic volatility is measured by daily return data over the past one month. Hence, idiosyncratic volatility is not really ideal for investment purposes anyway as it requires a high amount of turnover.

et al. (2016a) find that low beta factors do not have a statistically significant correlation between their valuation levels and following performance in the period from 1967 to 2016 in the U.S. This may be because low beta stocks had low valuation levels until the last 20 years. Arnott et al. (2016a) suggest that low beta stocks saw a substantial rise in valuation in recent years, which could give them a one-time boost in recent years that might not be expected for the future.

However, Blitz and Vidojevic (2017) contradict the conclusion from Fama and French (2016) as they maintain that the FF5 does not explain the low volatility effect. They justify their argument, among other things, with the finding of market beta remaining unpriced in the cross-section of expected returns. Additionally, they find a flat relation between market beta and returns.[33] Besides, the absence of a low volatility factor in asset pricing models can have practical reasons as all models in Table 3 are based on a market factor, which is difficult to combine with a low volatility factor. The reason for this is that the market factor connects higher risk (or higher exposure to market beta) with higher returns. The low volatility effect suggests the exact opposite as low risk stocks tend to outperform high risk stocks on a risk-adjusted basis, as documented by Frazzini and Pedersen (2014). Therefore, asset pricing models that are based on the positive linear relation between market beta and returns dictated by the CAPM would be inconsistent if they added a factor that offsets this relationship.

Moreover, science and practice do not agree on the use of an investment factor. HQ5, FF6 and BS6 all use investment factors for describing stock returns. In contrast, neither the heuristic of Beck et al. (2016) nor any of the prominent providers of factor-based investment products considers an explicit investment factor. For example, the FF5, HQ4, and HQ5 are able to explain a broad set of anomalies through a combination of their investment and profitability factors. Arnott et al. (2019) maintain that practitioners often group together profitability and investment factors as a quality factor. However, Hsu et al. (2019) demonstrate that quality factors (defined as, for example, leverage or earnings growth) which practitioners apply lack evidence for generating a robust factor premium. They find stronger evidence for robust factor premiums for quality factors that are defined as, for example, investment or asset growth. There-

[33] However, findings in Blitz and Vidojevic (2017) are contradictory to the results of factor zoo shrinkage methods by, for example, Harvey and Liu (2018) and Pukthuanthong et al. (2019) as they find the beta or market factor of Sharpe (1964) as a significant predictor for stock returns.

fore, an exclusion of an investment factor, which is widely accepted in the academic literature, from practitioners seems a little surprising.

Finally, dividend yield or high net payout and illiquidity factors suggested by some well-known providers of factor investing products and Blitz and Van Vliet (2018) are not included in any of the discussed asset pricing model. This difference and differences between science and practice in general may be due to the fact that scholars aim to explain the cross-section of expected returns and mostly ignore transaction costs whereas practitioners seek to identify and implement factors that have the potential of generating prospective abnormal returns after controlling for transaction costs. A profitable factor-based investment strategy that can be explained by other already known factors does not contribute to the existing asset pricing literature (Blitz and Van Vliet, 2018). Hence, practical considerations have sometimes only little additional scientific insights and vice versa, as academic studies often ignore transaction costs.

TABLE 3
Choosing factors: Science vs. practice

	Factor models					Three step heuristic	Investment practice
Factor	BS6	HQ5	FF6	SY4	D3	Beck et al. (2016)	
Market	✓	✓	✓	✓	✓		
Size	✓	✓	✓	✓			✓
Profitability	✓	✓	✓				✓
Value	✓		✓			✓	✓
Low volatility						✓	✓
Illiquidity						✓	✓
Momentum	✓		✓			✓	✓
Investment	✓	✓	✓				
Long-horizon mispricing					✓		
Short-horizon mispricing					✓		
Other mispricing factors				✓			
Expected growth		✓					
Dividend yield							✓
Income							✓

Source: Author's own illustration. Factors in the column "factor models" represent factors proposed by science. The three-step heuristic including a trading cost perspective of Beck et al. (2016) represents factors suggested by practitioners and the column "investment practice" represents factors actually offered in practice. Factors in the column "investment practice" include those from Table 2.

4.3. Factor returns

Different factors suggested by both science and practice have been comprehensively discussed in the previous chapters. Before evaluating the implementation of these factors into workable investment strategies, this chapter discusses studies that demonstrate the long-term profitability of factor investing and why factor premiums potentially appear. The latter is an important topic to evaluate the prospective existence of factor returns.

In order to discover factors, scholars usually build long-short zero-cost portfolios. These portfolios buy/short-sell stocks which have a positive/negative exposure to the selected factor. For example, to exploit the value factor, a portfolio holds a long position in value stocks (high *B/M*) and takes a short position in growth stocks (low *B/M*). The rationale behind this strategy is comprehensive U.S. evidence that demonstrates the long-term outperformance of value stocks toward growth stocks (e.g., Basu, 1977, 1983; Fama and French, 1992). In practice, where transaction costs start to play an important role, buying stocks is much easier and less expensive than short-selling stocks. Therefore, practitioners typically use long-only factor strategies to gain exposure to certain factors (Blitz, 2016).[34]

Studies that document profitable long-term factor returns include, for example, the ones from Van Gelderen and Huij (2014), Koedijk et al. (2016), and Dimson et al. (2017). Van Gelderen and Huij (2014) find that approximately 20–30% of their examined U.S. equity mutual funds pursue a size or value factor-based long-short investment strategy. Moreover, they identify 1–6% of their sample as momentum, low beta and short-term and long-term reversal funds. In sum, they find that low beta, value and size funds outperform the market index over the period from 1990 to 2010. In addition, they argue that the abnormal factor returns do not seem to vanish after the publication of the underlying factors. A comparison of the single-factor performance against the market index reveals that size and value funds generate excess returns of 0.56–1.19% p.a., after costs. Funds adopting low beta strategies earn higher risk-adjusted returns compared to the market index as these funds exhibit similar returns, but lower risk. The findings from Van Gelderen and Huij (2014) do not provide consistent evidence for excess returns of momentum and reversal funds. The reason why

[34] See Chapter 4.4.1, for an in-depth discussion of long-only and long-short factor strategies.

funds that adopt momentum and short-term reversal strategies might not earn excess returns could be due to the high transaction costs that are necessary to implement these strategies (e.g., Korajczyk and Sadka, 2004; Lesmond et al., 2004; Avramov et al., 2006). In addition, Avramov et al. (2007) emphasize the higher risk associated with momentum strategies, which impedes the implementation of such strategies. However, other parts of the literature argue that anomalies like momentum and short-term reversal might be implementable if strategies to mitigate transaction costs are applied (e.g., De Groot et al., 2012; Novy-Marx and Velikov, 2016; Frazzini et al., 2018).[35]

Van Gelderen and Huij (2014) solely concentrate on equity factor returns in the U.S. In contrast, Koedijk et al. (2016) and Dimson et al. (2017) conduct an out-of-sample test by examining factor returns across different stock markets. Moreover, Koedijk et al. (2016) explore factor returns[36] within different asset classes. They construct long-only portfolios and find robust factor returns within equity and bond portfolios and also within a portfolio consisting of different asset classes (equities, real estate, commodities, and bonds). Additionally, they test their findings in markets beyond the U.S and Europe and find consistent results. They conclude that factor investing seems to be profitable over the long-term.

Dimson et al. (2017) also attempt to find out whether the patterns of well-known anomalies actually exist and are able to improve returns. More specifically, they estimate past factor returns for size, value, dividend yield, momentum and low volatility over an extremely long period (up to 117 years for certain factors and stock markets) and across up to 23 different stock markets. Before costs, they find the largest factor premium for the momentum factor and the smallest premium for the size factor. Furthermore, they argue that all of their investigated factors have substantial impact on stock returns. Hence, all investors should at least monitor their examined factors. However, besides the possible outperformance of factor investing, it is also probable to experience long periods of bad performance with this investment approach, as emphasized by Koedijk et al. (2016) and Arnott et al. (2019). Therefore, it seems reason-

[35] Chapter 4.5.1.3, more accurately examines these mitigation strategies.

[36] They distinguish between the following factors for equities: value, momentum, size, and low volatility; and the following factors for bonds: term spread (e.g., Fama and French, 1989, 1993), credit spread (e.g., Fama and French, 1989, 1993; Elton et al., 2001), and short treasury and credit (e.g., Bieri and Chincarini, 2005; Asness et al., 2012).

able to consider a longer investment horizon when investing in factor-based strategies.

4.3.1. Drivers of factor returns

A frequent question that investors ask is "[…] how likely are the factors' excess returns to persist in the future?" (Bender et al., 2013, p. 9). To answer this question, you have to find out what potentially drove factor returns in the past and whether those drivers will pursue in the future. This issue is intensively discussed by both academics and practitioners. Basically, there are three debated explanatory approaches for factor premiums in the literature: Risk-based, behavioral-based explanations and data mining. However, although researchers have explored this issue for decades, there is no agreement in the literature on what drives distinct anomalies. In the following subsections, only the risk-based and behavioral explanations are discussed for the most prominent factors as these factors are most likely not the product of data mining. It seems appropriate to scrutinize only the drivers of the well-known factors (size, value, momentum, quality, and low volatility) that at least generated robust factor premiums in the past and also within out-of-sample tests. Besides, it seems vitally important whether the driver of a factor is systematic risk or mispricing. This is substantial because mispricing could vanish as more investors pile into factor-based strategies, whereas true risk factors seem to possess more potential for robust prospective returns.

4.3.1.1. Risk-based explanations

Risk-based explanations are consistent with the view of the EMH that most investors act rational and stock prices accurately reflect all available and essential information. Systematic risk is unpredictable and cannot be eliminated by diversification. Consequently, factor premiums offer a compensation for taken higher systematic risk. Based on the findings from Banz (1981), Fama and French (1992) initially found that small caps earned less in the 1980's and could be more fragile in certain market cycles and therefore compensated with higher returns. Further research concluded that small firms might be rewarded for other risk factors such as financial distress (Chan and Chen, 1991), default risk (Vassalou and Xing, 2004), liquidity risk (Amihud, 2002; Acharya and Pedersen, 2005; Liu, 2006), lack of transparency (Zhang, 2006) or growth options which are by definition more risky (Carlson et al., 2004; Garleanu et al., 2012). More recent evidence from Asness et al. (2018) and Hwang and Rubesam

(2018) shows a high correlation between liquidity risk and the size premium. Additionally, Asness et al. (2018) demonstrate that the size effect is significantly reduced after controlling for liquidity risk measures. These results are consistent with the explanation that small caps generate higher returns because of their underlying liquidity risk. However, Asness et al. (2018) still find a remaining low but significant size effect after adjusting for liquidity risk. This indicates that their liquidity factors may be flawed or just a part of the size premium can be explained by liquidity risk.

Researchers (e.g., Cochrane, 1991b, 1996; Zhang, 2005) hypothesize that value firms face problems especially in economically difficult times. In these times, value firms lack flexibility in comparison to growth firms which makes them riskier (Zhang, 2005). Chen and Zhang (1998) highlighted that financial distress, high financial leverage effects and uncertain future revenues make value firms dicier. In addition, Fama and French (1993, 1995, 1996) also argue that the driver of the value anomaly is risk which is not captured by the CAPM. They support their argument with the finding that market returns cannot explain all of the common variation in the returns of value firms, and that market earnings are not able to explain all of the common variation in the earnings of value firms. Other risk-based approaches that attempt to explain the value (or size) premium link the value (or size) factor to macroeconomic factors such as the GDP. For instance, Liew and Vassalou (2000, p. 244) note that value (and size) "[…] contain significant information about future GDP growth." Admittedly, the correlation between value and the GDP is insignificant in some countries and cannot be observed from 1957–1998 in the U.S. When they repeat the regressions for the 1979–1996 period the value factor becomes statistically significant.

Compared to value and size, there is scarce evidence for an explanation of momentum that is risk-based and consistent with the EMH. Nevertheless, Conrad and Kaul (1998) found that the cross-sectional variation might explain the profitability of some momentum strategies. Cross-sectional variation can be interpreted as undiversifiable risk which means that higher cross-sectional variance could be compensated with higher returns. However, the evidence of Jegadeesh and Titman (2001) is clearly inconsistent with the Conrad and Kaul (1998) hypothesis. Furthermore, Berk et al. (1999) provided a model that outlines how individual firms take their investment decisions and how these decisions affect the firms' systematic risk. Their model might be able to describe several cross-sectional patterns such as the momentum premium which indicates that investment decisions and their related systematic risk could explain the momentum effect. Moreover, Johnson (2002) and Sagi and Seasholes (2007) offered further rational models of momentum. These models have admittedly been criticized for assuming implausible risk aversion (Dimson et al., 2017).

Theoretically, it is pretty difficult to explain a quality premium. There are a huge number of possible quality measures. For example, Novy-Marx (2013) defines quality as a ratio of the gross-profits-to-assets and Fama and French (2015) used operating profit as a measure of quality. Quality firms could be seen as companies that have sustainable profits, are competitive or have a high return on equity. This actually proposes that such companies are less risky and should consequently not be compensated for higher systematic risk. Therefore, it seems difficult to find a theory that suggests risk as the driver of the quality factor because a portfolio that buys high quality stocks and shorts low quality stocks earns excess returns on a risk-adjusted basis (Asness et al., 2019). Nonetheless, there are some risk-based explanations for the quality effect. Some propose that a quality premium is driven by the costly reverse of investments and differences in operating leverage (Zhang, 2005; Kisser, 2014). Campbell et al. (2010) point out that the systematic risk of value and growth stocks is primarily driven by their cash flow fundamentals. Admittedly, their results challenge both rational and behavioral explanations. An alternate story could be that profitable companies operate in dicier fields, which makes them riskier (Bouchaud et al., 2016). In general, risk premiums are actually rewarded for a significant negative skewness (Harvey and Siddique, 2000, Lemperiere et al., 2017). Bouchaud et al. (2016) illustrate that strategies, based on three different quality measures,[37] have in fact a positive skewness. These results contradict risk-based theories. Asness et al. (2019) also discover more evidence for a behavioral explanation of the quality premium, although they cannot rule out a risk-based explanation. Especially their findings that quality stocks tend to deliver a good performance during distressed market periods challenge risk-based theories.

The low volatility premium is obviously contradictory to the EMH. For this reason, most explanations for this anomaly are behavioral. Theory of asset pricing suggests that higher risk must be compensated with higher returns. However, the low volatility

[37] They use following quality variations: return on assets (EBIT/total assets), return on equity (net income/common equity) and cash flows (net operating cash flows/total assets).

anomaly is the exact opposite as low volatility stocks generate higher risk-adjusted returns than high volatility stocks in the long-term, as documented by Black et al. (1972), Haugen and Heins (1975) and Black (1993). Ang et al. (2009) argued that the drivers of the low volatility could be latent systematic risks. They found a comovement between low volatility in the U.S. and in international markets, which indicates that a common risk factor could possibly drive this effect. Baker et al. (2011) also propose a rationale for the low volatility effect that is not behavioral-based.

4.3.1.2. Behavioral explanations

The behavioral view of the drivers of factor premiums is based on mispricing because of investors' psychology and limits to arbitrage (Barberis and Thaler, 2003). Those explanations clearly contradict the EMH as they assume that some, but not all investors behave irrationally. Due to behavioral biases from investors, stock prices do not reflect their fundamental value. However, limits to arbitrage prevent sophisticated investors from taking advantage of these biases, which leads to the persistence of these anomalies.

Behavioral explanations for the occurrence of the size premium build on incorrect extrapolating of the past, overestimating returns of glamour (growth) stocks or overreaction to different growth measures (Lakonishok et al., 1994; Barberis et al., 1998; Daniel et al., 1998; Hong and Stein, 1999). Additionally, small companies have greater limits to arbitrage, which amplifies mispricing (Shleifer and Vishny, 1997). Asness et al. (2018) examine these behavioral theories and conclude that their findings cannot be explained by limits to arbitrage and mispricing.

Lakonishok et al. (1994) argue that growth stocks are often overvalued, while value stocks seem to be less attractive to many investors. The problem here is that investors often erroneously expect growth stocks to maintain their past growth. Technology stocks during the dot-com bubble perfectly exemplify the bias of overvaluing growth stocks. Further reasons for the value effect include loss aversion and mental accounting. Loss aversion fundamentally means that people prefer avoiding losses instead of gaining the equivalent amount (Kahneman and Tversky, 1979). Following Barberis and Huang (2001), prior losses and gains impact the degree of loss aversion. A loss after a previous gain normally aches less because it can be compensated with prior gains. Conversely, a loss after a previous loss aches more. This means that stocks that have performed well in the past could be seen less risky, and as a result investors lower their discount rates for future cash flows of these companies. Stocks that have had a bad performance in the past might seem riskier and are therefore associated with higher discount rates and higher expected returns.

As mentioned above, most explanations for momentum are behavioral. The momentum effect likely arises either due to overreaction (Barberis et al., 1998; Daniel et al., 1998) or underreaction (Hong and Stein, 1999; Hong et al., 2000). These models propose negative post-holding returns of momentum. Jegadeesh and Titman (2001) evaluated different explanations for the momentum premium. They found evidence for the behavioral models in Barberis et al. (1998), Daniel et al. (1998), and Hong and Stein (1999), which detect delayed overreaction as the explanation for the momentum premium. Nevertheless, these results are ambiguous. There is no evidence for a reversal effect for the first four years after the portfolio formation and only a scarce but significant reversal effect in the fifth year (Jegadeesh and Titman, 2001). In contrast, Grinblatt and Han (2005)[38] criticize the studies from Barberis et al (1998), Daniel et al. (1998), and Hong and Stein (1999) as these studies do not even suggest unrealized capital gains (or losses) to be the key variable that explains and predicts the momentum effect. Furthermore, they argue that only their model is consistent with those of George and Hwang (2004) and Frazzini (2006), which also offer a momentum explanation. The study by Chui et al. (2010) globally investigated momentum and determined that momentum premiums differ among countries, which challenges both risk-based and behavioral theories. Proponents of risk-based explanations are challenged by the fact that the momentum strategies are compensated for higher risk in some but not all markets. The behavioral camp must state why investors behave irrationally in the U.S. and Europe, but not in most East Asian countries and Japan, as noted by Chui et al. (2010).

A behavioral interpretation of the quality premium draws upon biased analysts that pay too little attention to profitability information from financial statements such as the cash flow statement. Instead, analysts overweight other indicators such as earnings per share (Bouchaud et al., 2016). However, the cash flow statement contains relevant information that is needed to efficiently value a company (Sloan, 1996). An issue

[38] The model from Grinblatt and Han (2005) is inspired by the prospect theory of Kahneman and Tversky (1979) and the mental accounting framework of Thaler (1985).

that complements this behavioral argument is the conservatism bias. In this view, especially unexperienced analysts have sticky beliefs and underreact to good or bad news about prospective earnings (Bouchaud et al., 2019). Subsequently, quality stocks could be underpriced and junk stocks overpriced. The evidence from Asness et al. (2019) also indicates that biased expectations of analysts might cause mispricing, which is consistent with the theory of quality stocks being underpriced.

As mentioned above, the low volatility effect is clearly inconsistent with the EMH. Baker et al. (2011) suggest a behavioral model that explains the connection between low volatility and prospective stock returns. They believe that some investors prefer risky stocks because of cognitive biases such as the lottery-effect,[39] representativeness,[40] and overconfidence.[41] This preference for highly volatile stocks could cause overpricing which consequently results in lower returns. According to Baker et al. (2011), fixed-benchmark mandates prevent rational investors from arbitraging the low volatility anomaly away. Li et al. (2014) found that also high transaction costs massively limit the possibility to extract the low volatility effect. A study conducted by Li et al. (2016) investigates whether the low volatility effect can be explained due to systematic risk or mispricing. They find evidence that supports behavioral arguments (e.g., those from Baker et al., 2011) and limits to arbitrage (e.g., high transaction costs) that inhibit investors from fully exploiting the mispricing (Li et al., 2014). For a more comprehensive discussion of the triggers of the low volatility effect, I refer to Blitz et al. (2014).

Moreover, there are papers (e.g. Jacobs, 2015; McLean and Pontiff, 2016; Yan and Zheng, 2017; Engelberg et al., 2018; Jacobs and Müller, 2020) that suggest mispricing and limits to arbitrage as an explanation for a broad set of anomalies. For example, Engelberg et al. (2018) posit that investors have overly pessimistic and overly optimistic expectations about firms' cash flows. New information leads investors to adapt their biased expectations, which could cause price changes. Additionally, McLean and Pontiff (2016) find lower post-publication returns for anomalies which

[39] The lottery-effect is understood as buying volatile stocks for a low price which has a similar risk–return payoff as a lottery ticket.

[40] Investors extrapolate the success from a few high volatility stocks to all risky stocks and ignore the fundamental risks of those stocks.

[41] Investors overestimate their ability to predict future returns. Basically, it is simple for investors to buy stocks instead of short-selling them. Consequently, biased investors might drive up the prices of high volatility stocks which results in lower returns for these stocks.

have lower limits to arbitrage (anomaly portfolios with liquid stocks and low idiosyncratic risk stocks). These findings consequently support the idea of mispricing as an explanation for some or all of the abnormal returns of anomalies, and that academic studies draw attention to anomalies which are exploited by investors and consequently reduce mispricing (McLean and Pontiff, 2016).

4.4. Factor implementation

In the previous chapter, this paper discussed studies that support added value from factor investing strategies. Furthermore, I examined the potential drivers of these factors. In this section, the implementation and allocation of these factors is analyzed with the purpose of helping investors to better understand the process of factor investing and thus to define more realistic forecasts about factor investing.

4.4.1. Long-only vs. long-short

Reported factor premiums in academic papers are generated through long-short trading strategies. However, the academic literature routinely assumes that transaction costs are zero. For example, the costs of short-selling a microcap are enormous. Therefore, the question as to whether factor premiums can be better captured through a long-only or a long-short approach arises. This is a contentious issue in the literature, as some advocate a long-only (Blitz, 2012; Bambaci et al., 2013; Huij et al., 2014; Fitzgibbons et al., 2017) and others a long-short approach (Bender et al., 2010; Ilmanen and Kizer, 2012; Briere and Szafarz, 2017). In that respect, Israel and Moskowitz (2013) and Asness et al. (2014)[42] argue that the factors' long and short sides offer exposure to risk premiums. Fundamentally, considering only the long leg of a long-short portfolio causes an efficiency loss of the performance (Jacobs and Levy, 1993; Miller, 2001). Israel and Moskowitz (2013) outline that, in the case of factor-based strategies, the loss of efficiency is less pronounced as the long legs of factor portfolios produce usually more than 50% of the returns.

The empirical study from Huij et al. (2014) compares both long-only and long-short factor investing strategies consisting of market, value, size, momentum, and low volatility factors under different scenarios. Their results without considering benchmark restrictions, implementation costs, and factor decay show that the annual Sharpe ratio

[42] Note that these two papers consider investments in individual factors.

of the long-short portfolio is at 0.73 significantly higher than the long-only portfolio with 0.47. In addition, Briere and Szafarz (2017) maintain that a long-short approach enhances the mean-variance performance of factor-based strategies. However, concluding that a long-short attempt is favorable over long-only is hasty at this point. Investors whose performance is measured by a benchmark such as the market index[43] have to be careful when they contemplate a long-short factor approach. Huij et al. (2014) demonstrate that the relative risk from a long-short approach can be two to three times higher than the benchmark. In addition, a long-only factor portfolio can have a significantly higher information ratio[44] and a lower drawdown (peak-to-through decline) than a corresponding long-short one. A further advantage of applying a long-only strategy in the context of benchmarking is that, if factor premiums vanish in the future, the long-only portfolio will keep up with the benchmark anyway because it captures the market premium. In contrast, the performance of the long-short approach is completely dependent on the persistence of the factor premiums. A full disappearance of future factor premiums would mean that the long-short factors' return amounts to zero (Huij et al., 2014). Thus, in contrast to long-only investors, long-short factor investors need a much greater belief in the further existence of factor premiums.

Alternatively, investors could adjust their 100/100 long-short portfolio with a 100% long-only position in the market portfolio by buying for example index futures. Following Huij et al. (2014), I refer to this portfolio as "long-short beta 1." A comparison between the long-short beta 1 portfolio that invests 200/100% in long-short and the long-only portfolio illustrates that the long-short portfolio has now a slight advantage over the long-only strategy in terms of the performance measures (Huij et al., 2014). However, short-selling contains a number of hazards[45] that are not apparent in the information ratio and other common risk measures. Moreover, long-only factor strategies are more robust against a possible factor decay (see, e.g., Harvey et al., 2016; McLean and Pontiff, 2016) and implementation costs (e.g., transaction costs, borrowing costs, margin requirements and management fees) than long-short strategies in

[43] Huij et al. (2014) use the value-weighted returns of all CRSP stocks minus the return on the one-month T-bill as a proxy for the market.

[44] Information ratio = outperformance against the benchmark divided by tracking error (standard deviation of outperformance).

[45] For example, margin requirements, counterparty risk, short squeezes, unlimited losses, unavailability of the most desired short positions, and the force to close down positions at the worst times (Jones and Lamont, 2002; Huij et al., 2014).

probably both optimistic and pessimistic scenarios (Huij et al., 2014). Table 4 summarizes the abovementioned findings.

TABLE 4
Long-only vs. long-short approaches, U.S., July 1963–December 2010

	Long-only	Long-short	Long-short beta 1
Absolute performance characteristics (without costs and decay)			
Return p.a. (%)	7.7	4.3	9.0
Risk (volatility) p.a. (%)	16.5	5.9	16.9
Sharpe ratio	0.47	0.73	0.54
Other risk characteristics			
Leverage	100/0	100/100	200/100
Counterparty risk	No	Yes	Yes
Liquidity	Medium	Low	Low
Benchmark-relative performance characteristics			
Beta	1.01	-0.07	1.01
Outperformance p.a. (%)	3.6	0.1	4.8
Tracking error	4.9	17.7	5.9
Information ratio	0.72	0.01	0.82
Drawdown (%)	-3.0	-49.1	-2.5
Optimistic decay scenario			
Sharpe ratio	0.37	0.29	0.34
Pesimistic decay scenario			
Sharpe ratio	0.24	-0.01	0.17
Optimistic cost scenario			
Sharpe ratio	0.32	0.25	0.29
Pesimistic cost scenario			
Sharpe ratio	0.29	0.03	0.22

Source: Author's own illustration based on Huij et al. (2014, p. 15 f.)

The decision regarding applying a long-only or long-short strategy also depends on the investors' capabilities. Some institutional investors such as insurance companies are prohibited to engage short-selling (Molk and Partnoy, 2019). In that regard, Bris et al. (2007) found that short-selling was principally permitted in 35 out of 47 countries at least as of December 2001 (the last month of their examined sample). Nevertheless, it is prohibited during financial crises (Bernal et al., 2014). Besides legal barriers, 66.1% of mutual funds in the U.S. were not allowed to engage in short-selling because of their investment policies in 2000 (Almazan et al., 2004).

Another fundamental question concerning long-only multi-factor investing refers on how to combine different factors. Basically, there are two ways to gain exposure to multiple factors. First, the portfolio mix which is a combination of standalone single factor indices or ETFs. In doing so, the problem of selecting stocks that have a positive exposure to one factor—but simultaneously a negative exposure to another factor that offsets the positive expected returns—might occur (Blitz and Vidojevic, 2019). The portfolio mix approach is inspired by asset pricing models, like the CAPM or the FF3, for instance. Second, the integrated portfolio that initially aggregates information from both the short and long legs of the single factors and thereafter builds a portfolio with the expected returns for each stock based on the gathered information (Fitzgibbons et al., 2017).[46] Therefore, the integrated approach seeks to identify stocks that simultaneously have positive exposure to numerous factors. Previous papers (e.g., Bender and Wang, 2016; Clarke et al., 2016; Fitzgibbons et al., 2017; Blitz and Vidojevic, 2019) empirically favor integrated portfolio construction over the portfolio mix as the integrated approach improves returns and information ratios. Fitzgibbons et al. (2017) justify this with the fact that integrated portfolios better avoid stocks that are based on different factors that offset each other. Additionally, the integrating style is more efficient in selecting stocks that have positive exposure to various factors instead of stocks that generate only a positive premium from one factor and concurrently generate a negative premium from another factor.

However, the findings that integrated portfolio construction leads to higher returns by simultaneously lower risk contradict risk-based explanations for anomalies and one basic principle in finance: higher returns are linked with more risk. In the investment practice, there is obviously no clear consensus about which approach is best suited for implementing multi-factor investing. Table 5 provides an overview of prominent multi-factor ETFs and shows that four of these ETFs pursue a mixed and five an integrated approach. Yet, the overall tenor in the investment community is that the mixed approach is dominated by the integrated approach (Leippold and Ruegg, 2018).

[46] Bender and Wang (2016) and Blitz and Vidojevic (2019) refer to these two approaches as top-down (portfolio mix) and bottom-up (integrated portfolio).

Therefore, Leippold and Ruegg (2018) use a robust multiple testing framework to reexamine the returns and risk differences between the mixed and the integrated approach. They test 26 different factor combinations and do not find any significant evidence that supports the hypothesis of improved return–risk ratios for the integrated portfolio approach. Moreover, they show that the integrated portfolio is more sensitive to the low volatility factor. This finding is consistent with the idea that the integrated approach reduces risks by selecting stocks with positive exposure to more factors. However, Leippold and Ruegg (2018) show that the lower risk of the integrated portfolios entails lower returns. Another paper that challenges the integrated portfolio is conducted by Fraser-Jenkins et al. (2016). They highlight that both mixed and integrated portfolios contain similar return–risk ratios. Another disadvantage of the integrated approach could be that, if there are too many considered factors (means more criteria for stock selection), only a few stocks meet the requirements. As a result, the constructed portfolio could experience a lack of diversification.

TABLE 5
Long-only multi-factor ETFs: Mix vs. integrate

Name	Asset Manager	AuM (02/14/20)	Inception	Approach
Goldman Sachs ActiveBeta U.S. Large Cap Equity ETF	Goldman Sachs	U.S. $ 8.43B	01/28/15	mix
FlexShares Morningstar US Market Factor Tilt Index Fund	FlexShares	U.S. $ 1.56B	09/16/11	integrate
John Hancock Multifactor Large Cap ETF	John Hancock	U.S. $ 966.75M	09/28/15	integrate
State Street Multi-Factor Global Equity Fund	State Street	U.S. $ 589.71M	09/30/14	mix
iShares Edge MSCI Multifactor USA ETF	iShares	U.S. $ 979.76M	04/30/15	integrate
JPMorgan Diversified Return U.S. Equity ETF	JP Morgan	U.S. $ 81.47M	09/29/15	integrate
The Global X Scientific Beta US ETF	Global X	U.S. $ 98.79M	05/12/15	mix
Franklin LibertyQ US Equity UCITS ETF	Franklin	U.S. $ 27.97M	06/01/16	integrate
ETFS Diversified-Factor U.S. Large Cap Index Fund	ETF Securities	U.S. $ 7.82M (closed)	01/28/15	mix

Source: Author's own illustration based on Leippold and Ruegg (2018, p. 831). Data compiled by Bloomberg and ETF.com

Finally, the conclusions of the above discussed papers reveal that it might be possible to capture factor premiums through both long-only and long-short approaches. However, before implementing, investors should carefully consider the pros and cons of either approach and also potential restrictions of short-selling.

4.4.2. Active vs. passive

This section investigates whether factor investing contains more active or passive characteristics and which of these two approaches could be more profitable in harvesting factor premiums after costs. Furthermore, implications of passive and active factor investing for market efficiency are discussed.

It is not straightforward to assess whether factor investing is purely active or passive. Before analyzing this issue, it is necessary to identify what defines an active and a passive fund. Active funds hold a limited number of stocks (e.g., 30 or fewer) that are differently weighted in the way it is determined by the fund manager (Berk and Green, 2004). Furthermore, active funds are able to hedge their bets and typically have high turnover rates, resulting in higher transaction costs and management fees (Mondello, 2013).

In comparison with active funds, passive funds are mostly market-cap weighted buy-and-hold strategies with low turnover rates. Fundamental reallocation takes place if the underlying index changes (Ang et al., 2011). In addition, passive funds attempt to hold all (liquid and available) stocks of the underlying index (e.g., S&P 500) as only concentrating on a subset of the market would go more in the direction of active investing. Finally, due to lower rebalancing activities and basically no costs for identifying over- or undervalued stocks, passive funds contain lower management fees (Pace et al., 2016). In the context of passive investing, ETFs have strongly grown in popularity for the last 25 years and have become widely used investment vehicles (Ben-David et al., 2017).

In theory, frequently rebalanced long-short strategies are most appropriate in harvesting factor premiums and might additionally provide diversification benefits, as maintained by Ilmanen and Kizer (2012). Practitioners, however, predominantly implement factor strategies by using a long-only approach that replicates the performance of passive smart beta indices (Blitz, 2016). Whether an active or a passive approach is better in delivering factor returns after costs also depends on the required turnover of the underlying factor. Factors which are associated with higher transaction costs and which require more frequent trading, such as momentum and illiquidity, are potentially better captured through active management (Beck et al., 2016). However, passive indices deliver with low costs and fees to end investors most of the benefits of the (more liquid) factor premiums (e.g., market, value or low volatility) that do not require frequent trading (Hsu et al., 2015; Beck et al., 2016). Hence, factor investing or smart beta is a big deal in the ETF industry. Examples of such smart beta ETFs include the following (data compiled by Bloomberg and ETF.com as end of February 2020):

- **Size –** iShares USA Size Factor ETF: This fund attempts to replicate an index comprising large- and midcap U.S. stocks with comparatively small market capitalization; expense ratio (ER)[47] = 0.20% p.a.
- **Momentum** – iShares Edge MSCI USA Momentum Factor ETF: This fund replicates an index consisting of large- and midcap U.S. stocks with higher price momentum; ER = 0.15% p.a.
- **Low volatility** – Invesco S&P 500 Low Volatility ETF: This fund buys the 100 stocks of the S&P 500 with the lowest volatility over the last 12 months and is rebalanced quarterly; ER = 0.25% p.a.
- **Multi-factor** – iShares Edge MSCI World Multifactor UCITS ETF: This fund tracks an index consisting of international large- and midcap stocks with exposure to quality, momentum, size and value; ER = 0.50% p.a.

As the descriptions of these smart beta ETFs indicate, some of them combine both active and passive elements. For example, the Invesco S&P 500 Low Volatility ETF requires quarterly rebalancing. In addition, the ER of the Invesco S&P 500 low volatility ETF with 0.25% p.a. is pretty high compared to the 0.04% p.a. ER of a traditional index ETF, such as the Vanguard S&P 500 UCITS ETF. However, in comparison with the ER of 0.82% p.a. from the Fidelity Contrafund, which is one the largest active funds in the U.S., the ER of the Invesco S&P 500 low volatility ETF seems fairly low.

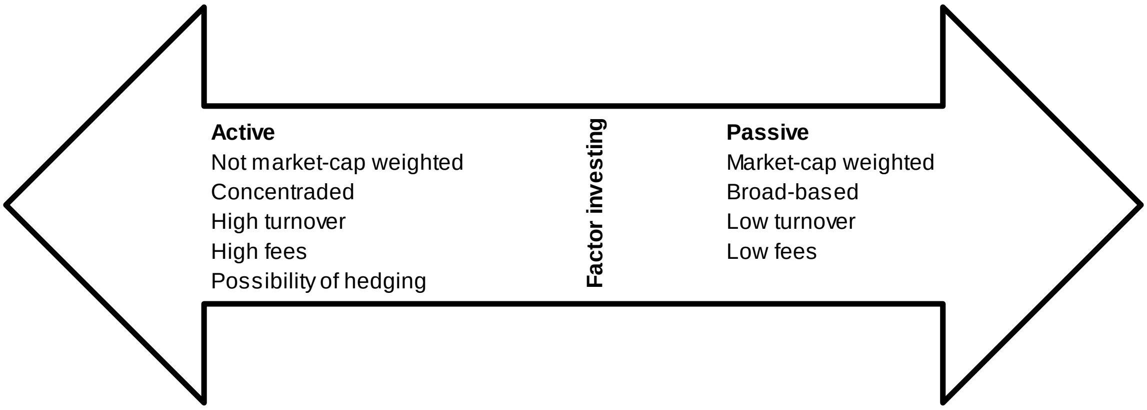
Regarding factor investing and the debate between active vs. passive, Altaf Kassam, the head of the MSCI's Index Applied Research, has made the following statement: "It is a third way of investing: between active and passive. It does not replace market-cap passive investing, nor does it fully replace active management. Factor investing has some of the features of passive investing, such as investing systematically at low cost. It also has some of the features of active management by aiming to generate returns above the market cap-weighted index." (Robeco, 2014, internet).

To sum up, in order to determine whether a factor-based strategy is active or passive, it seems appropriate to analyze every smart beta ETF or factor-based investment

[47] The ER measures the amount of the fund volume that is used to operate the fund. It includes, for example, management fees and advertising expenses.

strategy in isolation. Smart beta ETFs can probably be seen as a hybrid between active and passive as some smart beta ETFs contain more characteristics of an active fund and some do not. The framework in Figure 3 should help to identify the passive and active characteristics of the considered smart beta ETF or factor-based investment strategy.

FIGURE 3
Active vs. passive spectrum



Source: Author's own illustration

Another topic that could reintroduce a more skill-based and active management approach to factor investing is the timing and tilting of factors. Actually, this topic has been investigated by academics and practitioners 30 years ago. An early study in this area has been conducted by Arnott et al. (1989). Currently, there is a controversial debate in the literature about the added value of factor timing strategies. Opponents (e.g., Asness, 2016; Asness et al., 2017; Lee, 2017) argue that some factor timing strategies are highly correlated with the underlying factor strategies. This is especially the case with valuation-based factor timing strategies, which add further (suboptimal) value exposure to the portfolio, as emphasized by Asness et al. (2017). Furthermore, factor timing strategies produce high turnover rates and transaction costs that consequently erode most or all of its outperformance. Proponents (e.g., Arnott et al., 2016b; Hodges et al., 2017; Bender et al., 2018; Haddad et al., 2019) acknowledge the difficulties of factor timing by means of market, macroeconomic and sentiment predictors. Nonetheless, they think that, if investors consider long investment horizons and understand the drivers behind the factors, factor timing strategies might be able to improve factor returns.

Asness et al. (2017) examine whether a valuation-based contrarian factor timing strategy is able to generate outperformance against a passive factor allocation. In general, valuation-based factor timing relies on the changing valuation levels of factors (see, e.g., Arnott et al., 2016). If valuation levels of factors are high, and if mean-reversion is predictable and results from price changes, an investment strategy that exploits mean-reversion might generate excess returns. However, Asness et al. (2017) do not find a contrarian factor timing strategy to outperform a passive well-diversified factor portfolio. Findings in Asness (2016) are consistent with Asness et al. (2017) as he finds very weak performance and inapplicability of factor timing strategies based upon valuation of factors. Dichtl et al. (2019) attempt factor timing based on fundamental or factor-specific technical time-series predictors and factor tilting based on valuation and momentum characteristics. In a scenario where transaction costs are ignored, they find statistical and economic significant results for their factor timing and factor tilting strategies. By taking transaction costs into account, they come to the conclusion that most of the factor predictability's benefits vanish. However, using transaction cost penalties and the Black and Litterman (1992) shrinkage approach helps to maintain some of the added value from their factor timing, but not from their factor tilting strategies. Thus, the results in Dichtl et al. (2019) continue to ask whether factor timing and tilting strategies are able to obtain superior returns to some extent open.

The other stream of the literature that believes in the potential of improving returns through factor timing and tilting finds the combination of various predictors (e.g., valuation, macroeconomic, relative strength and dispersion metrics) superior to individual predictors (Hodges et al., 2017). Moreover, Bender et al. (2018) empirically identify valuation, past performance (momentum), and sentiment as predictors for future factor returns. However, these predictors do not work for every single factor and time horizon. For example, there is only a positive relation between sentiment and size or value. For sentiment and profitability or investment, they find a negative relation. Additionally, sentiment metrics only become strong predictors at time horizons for one year and beyond. Haddad et al. (2019) find that without consideration of transaction costs, factor-timing strategies substantially improve portfolio returns compared to static factor investing. However, they draw theoretical conclusions from their findings for the estimation of the SDF instead of deriving practical suggestions. To sum up, the possibility whether factor timing and tilting strategies improve outcomes compared to passive allocated factor portfolios when transaction costs are taken into account remains unclear and is a fecund area for future research that should be of interest to both practitioners and academics.

Besides the discussion whether an active or passive factor investing approach is more advantageous than the other, this topic has also broader implications for the stock market and its efficiency. Fundamentally, there is no agreement in the literature about the implications of ETFs on financial markets. Some researchers (e.g., Madhavan, 2016; Lettau and Madhavan, 2016; Madhavan and Sobczyk, 2016) argue that ETFs improve market efficiency as price determination of ETFs also results in price determination of the underlying stocks. Empirical studies that confirm this argument are conducted by Richie et al. (2008), Marshall et al. (2013), and Glosten et al. (2016). More precisely, for example, Glosten et al. (2016) show that prices of stocks in ETFs more readily incorporate information.

Other parts of the literature (e.g., Bradley and Litan, 2011a, 2011b; Broman, 2016; Brown et al., 2016; Da and Shive, 2018) find contradictory evidence as they show that ETFs diminish the information efficiency of their underlying stocks. For example, Da and Shive (2018) show that individual underlying stocks of ETFs probably react more slowly and less accurately to news, which could contribute to anomalies, such as the post-earnings announcement drift. Broman (2016) and Brown et al. (2016) demonstrate that ETFs attract noise traders, which could cause stock prices to diverge from their fundamental values.

ETFs also affect the liquidity of their underlying stocks. Marshall et al. (2015) argue that arbitrage trades between the ETF and the underlying stocks increase the liquidity of these stocks. Conversely, due to their inexpensive and uncomplicated investability, ETFs can lead to crowding out investors from the underlying stocks which decreases liquidity (Ben-David et al., 2017). Dannhauser (2017) finds such a crowding out effect for bond ETFs as the liquidity of the examined underlying bonds decreases after the introduction of the ETF. Furthermore, investors might accept to pay a premium for stocks that are more liquid. Piccotti (2014) and Petajisto (2017) find evidence for this hypothesis as they show that the value of some ETFs permanently deviates from the value of their underlying stocks. Moreover, Ben-David et al. (2017) argue that especially during times of market turmoil, the liquidity of ETFs drastically decreases.

Although passive investment approaches and particularly ETFs mostly outperform active funds after costs, active managers contribute to increasing market efficiency as they identify mispricing. If investors pursued only passive investment strategies, this would probably have substantial negative implications for both investment performance and aggregate capital allocation (Arnott et al., 2016b). In this context, the costs for active management of 0.67% p.a. estimated by French (2008) seem to be a fair price for society for achieving price discovery and efficient capital allocation.

4.5. The ignored risks of factor investing

After a comprehensive discussion of factors suggested by academics and practitioners, and of the different ways of translating factor premiums into feasible investment strategies, the following chapter examines the risks involved with factor investing.

4.5.1. Exaggerated expectations

4.5.1.1. Data mining and backtest overfitting

The reasons that factor investing could generate disappointing returns are manifold. Data mining (Lo and MacKinlay, 1990; Harvey et al., 2016; Harvey, 2017; Hou et al., 2018b; Chordia et al., 2019) and backtest overfitting (Bailey et al., 2014; Harvey and Liu, 2015; Suhonen et al., 2017) are two main concerns in this context. Backtesting means simulating an investment strategy with historical data. Data mining or data snooping refers to finding significant results, which are actually spurious findings, by reusing (or backtesting) the same data set (in our case the cross-section of returns). If thousands or millions of factor strategies are backtested, some of them will have outstanding returns and significant *t*-statistics only due to chance (Harvey and Liu, 2018). Bailey et al. (2014) demonstrate that only few configurations on a backtest can substantially improve the performance of the tested investment strategy. They refer to this practice as backtest overfitting. Additionally, they argue that the probability of overfitting drastically increases with the number of tried configurations on a backtest. Alas, the number of attempted configurations for a presented backtest is rarely published, which makes it difficult to determine the likelihood of overfitting. Moreover, the most recently uncovered factors lack an economic foundation (Harvey, 2017). In order to address the issue of data mining, Harvey and Liu (2015) propose a method for the real-time evaluation of trading strategies that also accounts for multiple testing. Moreover, Bailey et al. (2014) propose a minimum backtest length in order to determine the reliability of a backtested investment strategy.

Post-publication decline of factor returns is a further issue. McLean and Pontiff (2016) showed that 12 out of 97 anomalies failed to replicate their in-sample performance within out-of-sample observations. Consequently, superior past returns could vanish after the anomaly is published. Besides, Hou et al. (2018b) find that after excluding the bottom 2% in terms of market capitalization from the backtest, only 64% of their examined anomalies still generate significant abnormal returns. This finding demonstrates that, if the backtest contains illiquid and non-investable stocks, it is very likely that the live trading results will differ from the expected outcome. In sum, the previously discussed issues can lead investors to develop exaggerated expectations about prospective performance of factor-based strategies that have fantastic backtested returns but might fail to deliver in reality.

4.5.1.2. Crowding

Crowding is another hazard of factor investing that is associated with the post-publication decline.[48] Since information about the factors' excess returns is publicly available, the possibility of the factors becoming crowded rises. Once more investors trade the same factors; they might become rapidly unprofitable as prices rise and prospective returns plunge, as a result of disappearing mispricing. Subsequently, expected arbitrage, which could already have gone away, leads to frustrating real-time results (Arnott et al., 2019). The crowding hypothesis is consistent with the findings from McLean and Pontiff (2016), Arnott et al. (2019), and Jacobs and Müller (2020),[49] who point out that factor returns in the U.S. experience a significant decline after dissemination. Arnott et al. (2019) examine the average performance of 46 factors from the 10 years before and after the in-sample period. Figure 4 illustrates that the cumulative performance of the 46 factors used by Arnott et al. (2019) significantly decreased in the following 10 years after the in-sample observation. In detail, the average factor return in the 10 in-sample years becomes more than twice the return from

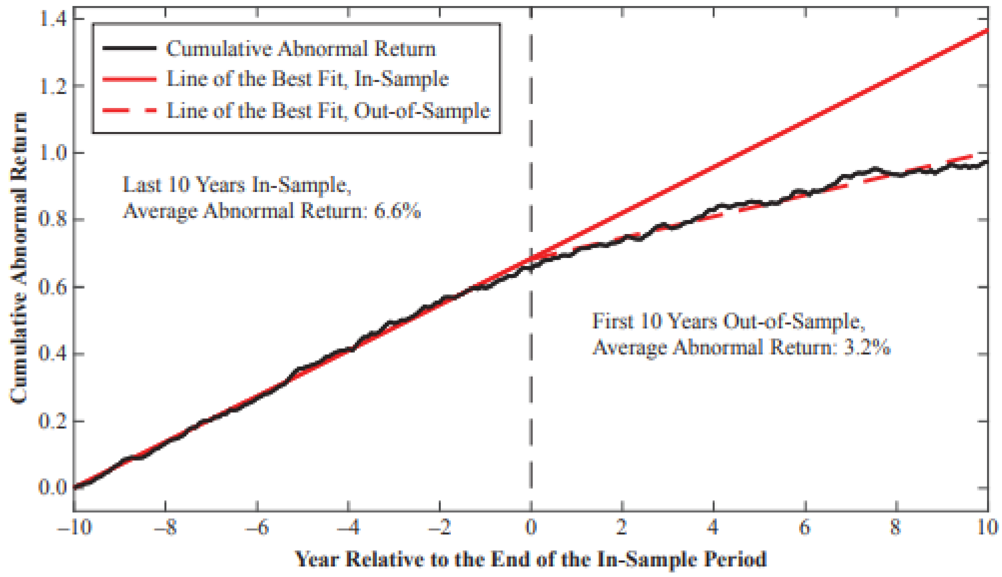
[48] For further papers that discuss crowding in financial markets, see: Wermers (1999), Pedersen (2009), Khandani and Lo (2011), Hong et al. (2016), and Bhansali and Harris (2018).

[49] Note that Jacobs and Müller (2020) only find a reliable post-publication decline in the U.S., but not in other stock markets across the globe. Probably, these anomalies in international markets do not become crowded after publication.

the subsequent 10 years after the end of the initial sample. Arnott et al. (2019) argue that there are at least three reasons for this: (1) trying various definitions of one factor can lead to find a high in-sample performance only by luck that turns out to be relatively weak out-of-sample; (2) factors might get crowded after their publication, which could reduce mispricing and hence mitigate previous abnormal returns; and (3) rising valuations of factors (e.g., high valuation levels at the end of the backtest compared to low valuation levels at the beginning could cause lower returns toward the end of the observation).[50]

FIGURE 4

Cumulative factor performance before and after publication, U.S.



Source: Arnott et al. (2019, p. 24).

4.5.1.3. Transaction costs

As soon as factor-based strategies are implemented, transaction costs will come into play. Novy-Marx and Velikov (2016) point out that most long-short factor portfolios with monthly more than 50% turnover lose all their returns when transaction costs are taken into account. In addition, the profits of the momentum premium seem to completely disappear after accounting for trading costs, as pointed out by Korajczyk and Sadka (2004) and Patton and Weller (2019). Moreover, Patton and Weller (2019) ar-

50 See Arnott et al. (2016b) for an investigation of how rising valuations affect the excess returns of factor-based strategies.

gue that implementation costs prevent mutual funds also from earning significant returns to the value premium but have less impact on size and market strategies.

Based on other papers on the decay of factor premiums, Chen and Velikov (2019) suggest that the real-world excess returns of factor strategies within U.S. stock markets amount to nearly zero. Preceding studies illustrate that publication bias constitutes around 12% of a factors' in-sample return (Chen and Zimmermann, 2018)[51] and approximately 32% is mispricing which disappears in consequence of arbitrage activities (McLean and Pontiff, 2016). The leftover 56% of the return are almost completely devoured by transaction costs (Chen and Velikov, 2019). In contrast to this deduction, Van Gelderen and Huij (2014) demonstrate that the excess factor returns from U.S. mutual funds do not disappear after the publication of the factors. However, considering U.S. stock markets, evidence from McLean and Pontiff (2016), Chen and Zimmermann (2018), Arnott et al. (2019), and Jacobs and Müller (2020) better supports the argument from Chen and Velikov (2019) as they find a significant post-publication decline for anomalies in the U.S. Jacobs and Müller (2020) do not detect a reliable post-publication effect in the other 38 examined stock markets across the globe.

In order to reduce transaction costs and hence improve returns, researchers (e.g., De Groot et al., 2012; Novy-Marx and Velikov, 2016; Frazzini et al., 2018) analyze different cost mitigation strategies. Novy-Marx and Velikov (2016) examine three different strategies, whereby the buy/hold strategy dominates the other two techniques. The hold range of this buy/hold spread strategy is larger than the buy range, meaning that it buys a stock when it reaches a certain and predefined signal (e.g., highest 10% on a certain signal), but does not sell the stock if it falls under the buy signal. Regarding the short side of the strategies, the threshold for shorting a stock is higher than the range for holding a previously opened short position. Basically, the buy/hold strategy certainly experiences marginally lower returns but simultaneously reduces average turnover by 41% and transaction costs by 42%. Frazzini et al. (2018) examine a dataset of real trading orders from an institutional investor across 21 international stock markets over a time period of 19 years. In general, they point out that transaction costs are at least for patient investors substantially lower than suggested by Korajczyk and Sadka

[51] Note that McLean and Pontiff (2016) initially estimate a higher publication bias than Chen and Zimmermann (2018). However, the evidence in Jacobs and Müller (2020) is more consistent with Chen and Zimmermann (2018).

(2004), Novy-Marx and Velikov (2016), Chen and Velikov (2019) and Patton and Weller (2019).

4.5.2. Tail behavior of factor strategies

Investors often misjudge the tail behavior of factors. Many investors believe that factor returns are essentially normally distributed. But in fact, factor returns are not even close to being normally distributed as they generally tend to experience large drawdowns (Arnott et al., 2019). In order to illustrate this issue, Table 6 shows the worst monthly returns of the examined factors over Arnott et al.'s (2019) 55-year sample and the estimated frequency in which such one-month drawdowns would occur under the assumption of normal distribution. As illustrated in Table 6, the probability of such monthly drawdowns reaches from once in 106 years (for long-term reversals) to once in 4.7 quadrillion (10^{15}) years (for operating profitability). Moreover, if factor returns were normally distributed, 10 of the 15 examined factors would experience such drawdowns in less than once in 1.6 million years. Nevertheless, most of these drawdowns actually occurred over the last 15 years. In addition, it is widely believed that extreme tail behavior can be eliminated by combining various factors in one portfolio (Arnott et al., 2019). However, as shown in Table 6, also the different factor portfolios actually suffer bad one-month drawdowns that would rarely arise under the assumption of normal distribution. Kalesnik and Linnainmaa (2018) examine six factors and one factor portfolio strategy and find similar results to those found by Arnott et al. (2019).

Table 6 illustrates the annual skewness and excess kurtosis for each factor, which is another way to examine the factors' deviation from normal distribution.[52] Unsurprisingly, momentum and illiquidity have the most negative skewness (the negative tail is much longer). Momentum is highly susceptible to crashes as demonstrated by Jegadeesh and Titman (1993) and Daniel and Moskowitz (2016), and illiquid stocks also tend to crash. Furthermore, the excess kurtosis is positive for every factor and substantially positive for operating profitability, momentum and net share issues. These findings propose that extreme outcomes are more common than some inves-

[52] Skewness is a measure for the symmetry properties of returns. If a distribution is negatively skewed, this means that it has a fatter tail on the negative side (more large negative outliers than positive ones). Excess kurtosis (actual kurtosis minus kurtosis in case of the normal distribution) measures the probability of extreme outcomes (or returns) in both positive and negative directions. Under the assumption of normal distribution, skewness and excess kurtosis equal zero.

tors would have believed. To sum up, factor returns have negative fat tails and an asymmetrical distribution to the downside. Hence, investors must take big drawdowns into account and must be prepared for such events. The assumption that returns are primarily normally distributed is simply false and leads to a totally wrong estimation of the actual risks of factor investing. In order to deal with extreme drawdowns, an appropriate long investment horizon should be there to catch up those negative outliers.

TABLE 6

Tail behavior of monthly factor returns, U.S., July 1963–June 2018

Factor	Average Annualized Return	Skewness	Excess Kurtosis	Worst Monthly Return	Frequency (in years) of Expected Worst Realized Monthly Return, Assuming Normal Return Distribution
Market	4.20%	–0.54	2.03	–15.3%	1 in 1.6 million
Value	4.15%	0.11	2.05	–11.5%	1 in 2,522
Size	2.53%	0.46	5.54	–16.2%	1 in 8.9 million
Operating Profitability	3.70%	–0.25	12.64	–24.3%	1 in 4.7 quadrillion (10^{15})
Investment	4.32%	0.20	0.79	–9.2%	1 in 109
Momentum	5.48%	–1.41	11.38	–24.3%	1 in 4.1 quadrillion (10^{15})
Low Beta	0.16%	–0.39	3.33	–17.1%	1 in 49.8 million
Idiosyncratic Volatility	1.62%	–0.32	4.39	–16.4%	1 in 12.3 million
Short-Term Reversals	5.34%	0.37	5.82	–13.5%	1 in 58,097
Illiquidity	3.01%	–0.58	5.11	–17.1%	1 in 55.8 million
Accruals	4.31%	–0.11	0.97	–10.0%	1 in 320
Cash Flow to Price	4.82%	–0.35	5.81	–18.9%	1 in 2.5 billion (10^{9})
Earnings to Price	3.76%	–0.32	5.94	–19.5%	1 in 10.5 billion (10^{9})
Long-Term Reversals	3.43%	0.65	2.62	–9.1%	1 in 106
Net Share Issues	5.28%	–0.40	9.85	–23.4%	1 in 357.5 trillion (10^{12})
Average of Other Factors	3.58%	0.03	4.19	–14.4%	
Portfolio of Factors 1–6	7.38%	0.46	8.12	–16.1%	1 in 6.5 million
Portfolio of Factors 7–14	7.28%	0.12	5.04	–17.6%	1 in 147.6 million
Portfolio of Other Factors	10.28%	0.63	4.93	–13.3%	1 in 38,989

Source: Arnott et al. (2019, p. 28).

Particularly momentum is the factor that experienced the worst drawdowns in the history of factor returns. In 2009, a long-short momentum portfolio crashed with a drawdown of 44% (Arnott et al., 2019). However, Barroso and Santa-Clara (2015) maintain that the risk of momentum varies over time and is predictable. More precisely, carefully controlling the risk of a long-short momentum strategy can drastically decrease the excess kurtosis, and hence the maximum drawdown, by half. Conversely, careful risk management nearly increases the Sharpe ratio of a momentum strategy by half (Barroso and Santa-Clara, 2015).

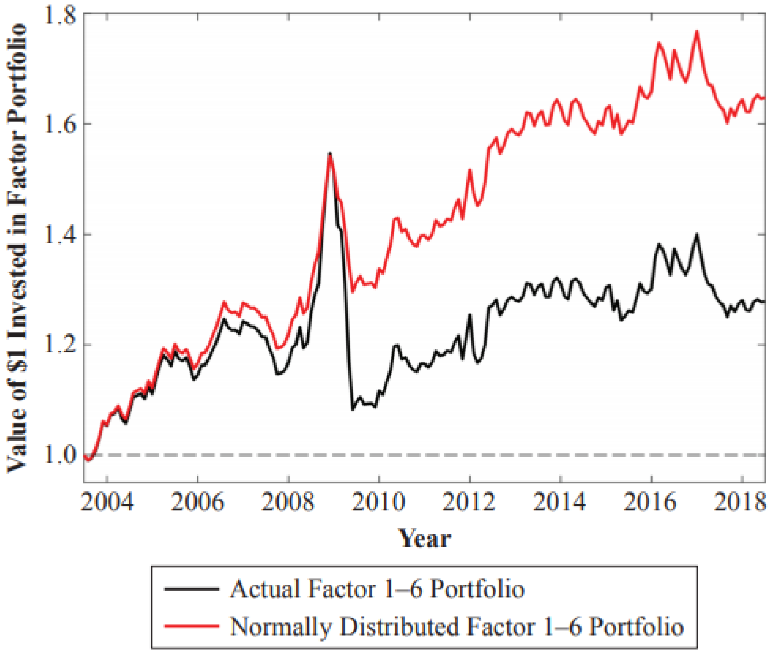
4.5.3. Mistaken diversification

Factors individually tend to have extreme tails. In contrast, a portfolio that combines different factors could be more diversified and closer to a normal distribution. The central limit theorem (Polya, 1920) might support this hypothesis as it basically states that an increase in the sample size leads to an approximation to a normal distribution. However, if factors are cross-correlated among each other, the approximation to a normal distribution will not happen. As shown in Table 6, the portfolio comprising the factors 1–6 has a similar excess kurtosis and experiences with minus 16% a similar worst monthly drawdown as the individual factors, which have an average worst month of minus 17%. This indicates that a portfolio of different factors is actually not beneficial to individual factors in terms of avoiding drawdowns. The reason for this is that factor returns become highly correlated in periods of market stress (Arnott et al., 2019).

Is the conclusion that a portfolio of different factors does not involve any diversification benefits justified? The answer to this is probably no. In this context, it is essential to highlight that the correlations of factor returns vary through time. Figure 5 shows that the returns of a factor portfolio comprising the market, value, size, operating profitability, investment, and momentum factor behaves similarly to the returns of the same portfolio that is assumed to be normally distributed, during the time period from mid-2003 till November 2008 (before the factor portfolio crash in 2009). Therefore, in normal times, returns of a factor portfolio are well approximated to be normally distributed and not as highly correlated as in distressed times. However, Figure 5 demonstrates the disappearance of diversification benefits regarding a portfolio of factors in times where actual factor returns experience a larger drawdown.

FIGURE 5

Factor portfolio crash: Actual factor portfolio returns vs. normally distributed factor portfolio returns, U.S., July 2003–July 2018



Source: Arnott et al. (2019, p. 30).

5. Summary and conclusions

This thesis aimed to provide an overview of the academic discussion of factors that attempt to explain the cross-section of expected returns and the practical implications of these factors. In order to answer the research questions in Chapter 1.2., a critical evaluation of the existing literature has been done.

In recent years, the production of factors for predicting stock returns escalated as there are over 400 anomalies or factors proposed in the academic literature (Harvey et al., 2016; Hou et al., 2018b; Harvey and Liu, 2019). One strand of the literature (e.g., Harvey et al., 2016; Harvey, 2017; Hou et al., 2018b; Linnainmaa and Roberts, 2018; Chordia et al., 2019) sees data mining or multiple testing and different forms of *p*-hacking as the main reasons for the proliferation of the factor zoo. After recognizing the problem of data mining or multiple testing, Harvey et al. (2016, p. 37) come to the damning verdict that "[…] many of the factors discovered in the field of finance are likely false discoveries […]." To address the problem of data mining or multiple testing, Harvey et al. (2016) suggest out-of-sample validation (Schwert, 2003; Asness et al., 2013; Frazzini and Pedersen, 2014; McLean and Pontiff, 2016; Linnainmaa and Roberts, 2018; Jacobs and Müller, 2020) and statistical frameworks that allow for multiple tests (Harvey et al., 2016; Green et al., 2017) as possible ways. However, another strand of the literature (e.g., Yan and Zheng, 2017; Engelberg et al., 2018; Wahal, 2019; Jacobs and Müller, 2020) argues that at least a large proportion of factors are real discoveries and not found by chance. These studies suggest that anomalies are better explained through mispricing caused by irrational behavior of investors.

Against this background, determining whether an already known or newly discovered factor provides reliable and independent information about stock returns has become a vibrant research area in the asset pricing field, as noted by Feng et al. (2020). The literature provides different statistical methods to approach this issue. Approaches range from a bootstrap method (Harvey and Liu, 2018), examining the maximum squared Sharpe ratio of factors (Barillas and Shanken, 2018; Fama and French, 2018), various regression analysis such as different LASSO procedures (Freyberger et al., 2018; Feng et al., 2020; Kozak et al., 2020) or Bayesian variable selection (Hwang and Rubesam, 2018), to other machine learning methods (Gu et al., 2019; Kelly et al., 2019). Surprisingly, Harvey and Liu (2018) find the original market factor of Sharpe (1964) as the dominant predictor in the cross-section of expected returns.

Asset pricing models combine a small number of some of the various factors from the factor zoo to explain the cross-section of stock returns. The CAPM with the beta or market factor was the first popular and parsimonious factor model that attempted to explain the relation between risk and returns with only one factor. Currently, prominent asset pricing models such as the SY4, D3, H*Q*5, FF6, and BS6 use up to six factors for relating expected returns to the sensitivity of factors. In particular, these models include factors such as market, size, value, investment, profitability, expected growth, momentum, mispricing factors that correspond to multiple anomalies, long- or short-horizon mispricing. Although the current asset pricing models contain different factors and to some extent similar factors which are, however, differently constructed, these models are closely related on empirical grounds (Hou et al., 2019). However, recent evidence in Kozak et al. (2020) suggests that the era of summarizing the cross-section of returns with only a few factors (e.g., four, five or six) is finally over. The number of return predictors found in the literature seems to be too large and the redundancy between these factors is too small for explaining stock returns with characteristic-sparse factor models (Kozak et al., 2020).

Investment companies provide investment products based on factors that are included in prominent asset pricing models on the one hand, but on the other hand, there are divergences between science and practice in terms of factor selection. These divergences are most conspicuous regarding the low volatility factor. This factor is offered by all prominent providers of factor-based investment products and suggested by many practitioners (e.g., Blitz and Van Vliet, 2007; Hsu et al., 2015; Beck et al., 2016; Blitz and Van Vliet, 2018; Blitz et al., 2020). In contrast, none of the well-known asset pricing models relates stock returns to a low volatility factor. Blitz et al. (2020), all of whom work for Robeco, back up their recommendation for investing in low volatility with empirical evidence that documents abnormal risk-adjusted returns of this factor across international stock markets (e.g., Blitz and Van Vliet, 2007; Blitz et al., 2013; Frazzini and Pedersen, 2014), different asset classes such as corporate bonds (e.g., Carvalho et al., 2014; Houweling and Van Zundert, 2017) and the options market (Falkenstein, 2009).

However, the profitability factor of Hou et al. (2015) captures the idiosyncratic volatility effect of Ang et al. (2006), which is associated with low volatility. Additionally, Fama and French (2016) demonstrate that their investment and profitability factors also largely explain the idiosyncratic volatility and the low volatility anomaly. Furthermore, all of the asset pricing models discussed in this thesis share a market factor that predicts a positive and linear relation between risk and returns. It would be contradictory if these factor models added a low volatility factor that offsets this positive risk-return relation.

Before translating different factors into feasible investment strategies, the central question that investors face is: How likely are the factors' abnormal returns to exist in the future? To address this issue, the potential drivers of the factor returns have to be analyzed along with the question of the longevity of those drivers in the future. Fundamentally, two main drivers, namely risk-based and behavioral explanations, are discussed in the literature. Risk-based explanations are consistent with the EMH and view factor returns as a compensation for systematic risk (Fama and French, 1992, 1993; Hou et al., 2015). Behavioral explanations argue that markets are to some extent inefficient and biased expectations generate mispricing which remains due to limits to arbitrage (Barberis and Thaler, 2003; Jacobs, 2015; McLean and Pontiff, 2016; Yan and Zheng, 2017; Engelberg et al., 2018; Jacobs and Müller, 2020). However, there is no consensus in the literature whether factor returns are primarily driven by systematic risk or mispricing. Besides, it is essential to know whether the abnormal returns of a factor are due to systematic risk or mispricing, because the latter, and hence abnormal returns, could vanish as more investors pile into these factor-based strategies. For this reason, the examination of whether factor returns are better explained by systematic risk or mispricing is an essential topic for prospective research.

When it comes to the practical implementation of factors, investors can choose between long-only or long-short strategies. In this respect, there is no agreement in the literature as one part of the literature suggests long-only (Blitz, 2012; Bambaci et al., 2013; Huij et al., 2014; Fitzgibbons et al., 2017) and the other part recommends long-short factor strategies (Bender et al., 2010; Ilmanen and Kizer, 2012; Briere and Szafarz, 2017). Nevertheless, factor investing is predominantly implemented through long-only smart beta ETFs in practice, as stated by Blitz (2016). Moreover, a distinction can be drawn between single-factor and multi-factor strategies. Within long-only multi-factor strategies there are two popular approaches: The portfolio mix and the integrated portfolio. Mixed portfolios combine different standalone single-factor ETFs. This can create the problem of selecting stocks that have positive exposure to one factor but simultaneously negative exposure to another factor (Blitz and Vidojevic, 2019). Integrated portfolios select stocks that contain positive exposure to multiple factors (Fitzgibbons et al., 2017). However, if too many factors are considered, the risk that only a few stocks meet all the criteria arises. The constructed factor portfolio could consequently not be well diversified. Whether the mixed or the integrated approach is more advantageous in terms of return risk ratios has not yet been fully elucidated as some researchers (e.g., Bender and Wang, 2016; Clarke et al., 2016; Fitzgibbons et al., 2017; Blitz and Vidojevic, 2019) favor the integrated style and others (e.g., Fraser-Jenkins et al., 2016; Leippold and Ruegg, 2018) argue that none of the two approaches dominates the other. Besides the question as to whether the mixed or the integrated portfolio is superior in terms of return risk ratios, both frameworks could mostly be implemented through passive investment approaches such as automated factor selection based on indicators that is conducted by computers and artificial intelligence. Hence, the question arises as to which role asset managers play in the implementation process of factor-based strategies. Factor timing strategies could reintroduce a more active and skill-based approach to factor investing. However, the question as to whether factor timing is able to improve returns remains open and is therefore a subject for future research that will be of interest to both academics and practitioners.[53]

Although researchers (e.g., Ang et al., 2009; Van Gelderen and Huij, 2014; Blitz, 2012, 2015; Koedijk et al., 2016; Dimson et al., 2017) advocate factor investing, this investment framework entails substantial risks that are often underestimated. First, data mining (Lo and MacKinlay, 1990; Harvey et al., 2016; Harvey, 2017; Hou et al., 2018b; Chordia et al., 2019), backtest overfitting (Bailey et al., 2014; Harvey and Liu, 2015; Suhonen et al., 2017), crowding (Arnott et al., 2019), and unrealistic estimations of transaction costs (Korajczyk and Sadka, 2004; Novy-Marx and Velikov, 2016; Chen and Velikov, 2019; Patton and Weller, 2019) lead investors to develop exaggerated expectations about future factor returns. Second, investors tend to misjudge the tail behavior of factor strategies. Especially in distressed times, factor returns are far away from a normal distribution that manifests in large drawdowns. Finally, investors often believe that they can eliminate the risk of factor investing by combining different factors into one portfolio. This assumption is also dangerous as the correlation of factors varies over time. In periods of market stress, when diversification is essential, factors become highly correlated and diversification benefits can largely vanish (Kalesnik and Linnainmaa, 2018; Arnott et al., 2019). Furthermore, factor investing is no free lunch as factors returns are cyclical. As pointed out by Bender et al. (2013), there are time periods where factors exhibit significant underperformance. Factor investing can be a useful investment framework that has the potential of improving long-term returns. However, before adopting this investment approach, investors need to understand the implementation process and the risks involved.

[53] Opponents of factor timing strategies include Asness (2016), Asness et al. (2017) and Lee (2017), and proponents include Arnott et al. (2016), Hodges et al. (2017), Bender et al. (2018) and Haddad et al. (2019).

References

Acharya, V., Pedersen, L.H., (2005). Asset pricing with liquidity risk, *Journal of Financial Economics* 77(2), 375-410.

Adrian, T., Rosenberg, J. (2008). Stock returns and volatility: Pricing the short-run and long-run components of market risk, *Journal of Finance* 63(6), 2997-3030.

Aharoni, G., Grundy, B., Zeng, Q. (2013). Stock returns and the Miller Modigliani valuation formula: Revisiting the Fama French analysis, *Journal of Financial Economics* 110(2), 347-357.

Ahmed, S., Bu, Z., Tsvetanov, D. (2019). Best of the best: A comparison of factor models, *Journal of Financial and Quantitative Analysis* 54(4), 1713-1758.

Aked, M., Moroz, M.(2015). The market impact of passive trading, *Journal of Trading* 10(3), 5-12.

Almazan, A., Brown, K.C., Carlson, M., Chapman, D.A. (2004). Why constraint your mutual fund manager?, *Journal of Financial Economics* 73(2), 289-321.

Amenc, N., Goltz, F. (2013). Smart beta 2.0, *Journal of Index Investing* 4(3), 15-23.

American Finance Association, (2016). Code of professional conduct and ethics, retrieved from: http://www.afajof.org, 10th December 2019.

American Statistical Association, (2016). Statement on statistical significance and p-values, retrieved from: https://www.amstat.org/asa/files/pdfs/P-ValueStatement.pdf, 10th December 2019.

Amihud, Y. (2002). Illiquidity and stock returns: Cross-section and time-series effects, *Journal of Financial Markets* 5(1), 31-56.

Ang, A., Goetzmann, W.N., Schaefer S.M. (2009). Evaluation of active management of the Norwegian government pension fund – global, report to the Norwegian parliament.

Ang, A., Goetzmann, W.N., Schaefer, S.M. (2011). The efficient market theory and evidence: Implications for active investment management, *Foundations and Trends in Finance* 5(3), 157-242.

Ang, A., Hodrick, R., Xing, Y., Zhang, X. (2006). The cross-section of volatility and expected returns, *Journal of Finance* 61(1), 259-299.

Ang, A., Hodrick, R., Xing, Y., Zhang, X. (2009). High Idiosyncratic volatility and low returns: International and further U.S. evidence, *Journal of Financial Economics* 91(1), 1-23.

Arnott, R.D., Beck, N., Kalesnik, V. (2017a). Forecasting factor and smart beta returns (hint: History is worse than useless), available at SSRN: https://papers.ssrn.com/abstract=3040953.

Arnott, R.D., Beck, N., Kalesnik, V. (2016a). To win with "smart beta" ask if the price is right, available at SSRN: https://papers.ssrn.com/abstract=3040976.

Arnott, R.D., Beck, N., Kalesnik, V., West, J. (2016b). How can "smart beta" go horribly wrong?, available at SSRN: https://papers.ssrn.com/abstract=3040949.

Arnott, R.D., Harvey, C.R., Kalesnik, V., Linnainmaa, J. (2019). Alice's adventures in factorland: Three blunders that plague factor investing, *Journal of Portfolio Management* 45(4), 18-36.

Arnott, R.D., Kalesnik, V., Wu, L. (2017b). The incredible shrinking factor return, available at SSRN: https://papers.ssrn.com/abstract=3040964.

Arnott, R.D., Kelso, C.M., Kiscadden, S., Macedo, R. (1989). Forecasting factor returns: An intriguing possibility, *Journal of Portfolio Management* 16(1), 28-35.

Asness, C.S. (2016). The siren song of factor timing aka "smart beta timing" aka "style timing", *Journal of Portfolio Management* 42(5), 1-6.

Asness, C.S., Chandra, S., Ilmanen, A., Israel, R. (2017). Contrarian factor timing is deceptively difficult, *Journal of Portfolio Management* 43(5), 72-87.

Asness, C.S., Frazzini, A., Israel, R., Moskowitz, T.J. (2014). Fact, fiction, and momentum investing. *Journal of Portfolio Management* 40(5), 75-92.

Asness, C.S., Frazzini, A., Israel, R., Moskowitz, T.J., Pedersen, L.H. (2018). Size matters, if you control your junk, *Journal of Financial Economics* 129, 479-509.

Asness, C.S., Frazzini, A., Pedersen, L.H. (2012). Leverage aversion and risk parity, *Financial Analysts Journal* 68(1), 47-59.

Asness, C.S., Frazzini, A., Pedersen, L.H. (2019). Quality minus junk, *Review of Accounting Studies* 24(1), 34-112.

Asness, C.S., Moskowitz, T.J., Pedersen, L.H. (2013). Value and momentum everywhere, *Journal of Finance* 68(3), 929-985.

Avramov, D., Chordia, T., Goyal, A. (2006). Liquidity and autocorrelations in individual stock returns, *Journal of Finance* 61(5), 2365-2394.

Avramov, D., Chordia, T., Gergana, J., Philipov, A. (2007). Momentum and credit rating, *Journal of Finance* 62(5), 2503-2520.

Bailey, D.H., Borwein, J.M., López de Prado, M., Zhu, Q.J. (2014). Pseudo-mathematics and financial charlatanism: The effects of backtest overfitting on out-of-sample performance, *Notices of the American Mathematical Society* 61(5), 458-471.

Baker, M., Bradley, B., Wurgler, J. (2011). Benchmarks as limits to arbitrage: Understanding the low-volatility anomaly, *Financial Analyst Journal* 67(1), 40-54.

Baker, N. L., Haugen, R. A. (2012). Low risk stocks outperform within all observable markets of the world, available at SSRN: https://papers.ssrn.com/abstract=2055431.

Ball, R. (1978). Anomalies in relationships between securities' yields and yield-surrogates, *Journal of Financial Economics* 6(2-3), 103-126.

Bambaci, J., Bender, J., Briand, R., Gupta, A., Hammond, B., Subramanian, M. (2013). Harvesting risk premia for large scale portfolios: Analysis of risk premia indices for the ministry of finance, Norway, retrieved from: https://www.regjeringen.no/contentassets/f453b48778c342d9a7ed8d6810d6cea4/harvesting_risk.pdf, 14th February 2020.

Banz, R.W. (1981). The relationship between return and market value of common stocks, *Journal of Financial Economics* 9(1), 3-18.

Barberis, N., Huang, M. (2001). Mental accounting, loss aversion and individual stock returns, *Journal of Finance* 56(4), 1247-1292.

Barberis, N., Shleifer, A., Vishny, R. (1998). A model of investor sentiment, *Journal of Financial Economics* 49(3), 307-343.

Barberis, N., Thaler, R. (2003). A survey of behavioural finance, in: Constantinides, G., Harris, M., Stulz, R.M. (Eds.), *Handbook of the Economics of Finance*, Amsterdam: Elsevier, 1053-1128.

Barillas, F., Shanken, J. (2017). Which alpha?, *Review of Financial Studies* 30(4), 1316-1338.

Barillas, F., Shanken, J. (2018). Comparing asset pricing models, *Journal of Finance* 73(2), 715-754.

Barroso, P., Santa-Clara, P. (2015). Momentum has its moments, *Journal of Financial Economics* 116(1), 111-120.

Basu, S. (1977). Investment performance of common stocks in relation to their price earnings ratios: A test of the efficient market hypothesis, *Journal of Finance* 32(3), 663-682.

Basu, J. (1983). The relationship between earnings' yield, market value and return for NYSE common stocks: Further evidence, *Journal of Financial Economics* 12(1), 129-156.

Beck, N., Hsu, J.C., Kalesnik, V., Kostka, H. (2016). Will your factor deliver? An examination of factor robustness and implementation costs, *Financial Analysts Journal* 72(5), 58-82.

Belloni, A., Chernozhukov, V., Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls, *Review of Economic Studies* 81(2), 608-650.

Ben-David, I., Franzoni, F.A., Moussawi, R. (2017). Exchange-traded funds, *Annual Review of Financial Economics* 9, 169-189.

Bender, J., Briand, R., Melas, D., Subramanian, R.A. (2013). Foundations of factor investing, available at SSRN: https://papers.ssrn.com/abstract=2543990.

Bender, J., Briand, R., Nielsen, F., Stefek, D. (2010). Portfolio of risk premia: A new approach to diversification, *Journal of Portfolio Management* 36(2), 17-25.

Bender, J., Sun, X., Thomas, R., Zdorovtsov, V. (2018). The promises and pitfalls of factor timing, *Journal of Portfolio Management* 44(4), 79-92.

Bender, J., Wang, T. (2016). Can the whole be more than the sum of the parts? Bottom-up versus top-down multifactor portfolio construction, *Journal of Portfolio Management* 42(5), 39-50.

Berk, J.B., Green, R.C. (2004). Mutual fund flows and performance in rational markets, *Journal of Political Economy* 112(6), 1269-1295.

Berk, J.B., Green, R.C., Naik, V. (1999). Optimal investment, growth options, and security returns, *Journal of Finance* 54(5), 1553-1607.

Bernal, O., Herinckx, A., Szafarz, A. (2014). Which short-selling regulation is the least damaging to market efficiency? Evidence from Europe, *International Review of Law and Economics* 37, 244-256

Bhansali, V., Harris, L. (2018). Everybody’s doing it: Short volatility strategies and shadow financial insurers, *Financial Analysts Journal* 74(2), 12-23.

Bieri, D.S., Chincarini, L.B. (2005). Riding the yield curve: A variety of strategies, *Journal of Fixed Income* 15(2), 6-35.

Black, F. (1993). Beta and return, *Journal of Portfolio Management* 20(1), 8-18.

Black, F., Jensen, M.C., Scholes, M.S. (1972). The capital asset pricing model: Some empirical tests, in: Jensen, M.C. (Ed.), *Studies in the Theory of Capital Markets*, New York: Praeger, 79-121.

Black, F., Litterman, R. (1992). Global portfolio optimization, *Financial Analysts Journal* 48(5), 28-43.

BlackRock (2018). What is factor investing?, retrieved from: https://www.blackrock.com/us/individual/investment-ideas/what-is-factor-investing/factor-commentary/andrews-angle/factor-growth, 24th October 2019.

Blitz, D. (2012). Strategic allocations to premiums in the equity market, *Journal of Index Investing* 2(4), 42-49.

Blitz, D. (2015). Factor investing revisited, *Journal of Index Investing* 6(2), 7-17.

Blitz, D. (2016). Factor investing with smart beta indices, *Journal of Index Investing* 7(3), 1-6.

Blitz, D., Falkenstein, E., Van Vliet, P. (2014). Explanations for the volatility effect: An overview based on the CAPM assumptions, *Journal of Portfolio Management* 40(3), 61-76.

Blitz, D., Pang, J., Van Vliet, P. (2013). The volatility effect in emerging markets, *Emerging Markets Review* 16, 31-45.

Blitz, D., Van Vliet, P. (2007). The volatility effect, *Journal of Portfolio Management* 34(1), 102-113.

Blitz, D., Van Vliet, P. (2018). The conservative formula: Quantitative investing made easy, *Journal of Portfolio Management* 44(7), 24-38.

Blitz, D., Van Vliet, P., Baltussen, G. (2020). The volatility effect revisited, *Journal of Portfolio Management Quantitative Special Issue* 46(2), 45-63.

Blitz, D., Vidojevic, M. (2017). The profitability of low-volatility, Journal of Empirical Finance 43, 33-42.

Blitz, D., Vidojevic, M. (2019). The characteristics of factor investing, *Journal of Portfolio Management* 45(3), 69-86.

Boguth, O., Kuehn, L.A. (2013). Consumption volatility risk, *Journal of Finance* 68(6), 2589-2615.

Bouchaud, J.P., Ciliberti, S., Landier, A., Simon, G., Thesmar, D. (2016). The excess returns of “quality” stocks: A behavioral anomaly, *Journal of Investment Strategies* 5(3), 51-61.

Bouchaud, J.P., Krüger, P., Landier, A., Thesmar, D. (2019). Sticky expectations and the profitability anomaly, *Journal of Finance* 74(2), 639-674.

Boudoukh, J., Michaely, R., Richardson, M., Roberts, M.R. (2007). On the importance of measuring payout yield: Implications for empirical asset pricing, *Journal of Finance* 62(2), 877-915.

Bradley, H., Litan, R.E. (2011a). Choking the recovery: Why new growth companies aren’t going public and unrecognized risks of future market disruptions, available at SSRN: https://papers.ssrn.com/abstract=1706174.

Bradley, H., Litan, R.E. (2011b). ETFs and the present danger to capital formation, available at SSRN: https://papers.ssrn.com/abstract=1947346.

Briere, M., Szafarz, A. (2017). Factor investing: The rocky road from long only to long short, available at SSRN: https://papers.ssrn.com/abstract=2908491.

Bris, A., Goetzmann, W.N., Zhu, N. (2007). Efficiency and the bear: Short sales and markets around the world, *Journal of Finance* 62(3), 1029-1079.

Broman, M.S. (2016). Liquidity, style investing and excess comovement of exchange-traded fund returns, *Journal of Financial Markets* 30, 27-53.

Brown, D.C., Davies, S.W., Ringgenberg, M.C. (2016). ETF arbitrage and return predictability, Working Paper, University of Arizona.

Bryzgalova, S. (2016). Spurious factors in linear asset-pricing models, Working Paper, Stanford University.

Campbell, J.Y., Polk, C., Vuolteenaho, T. (2010). Growth or glamour? Fundamentals and systematic risk in stock returns, *Review of Financial Studies* 23(1), 305-344.

Carhart, M.M. (1997). On persistence in mutual fund performance, *Journal of Finance* 52(1), 57-82.

Carlson, M., Fisher, A., Giammarino, R. (2004). Corporate investment and asset price dynamics: Implications for the cross-section of returns, *Journal of Finance* 59(6), 2577-2603.

Carvalho, R.L., Dugnolle, P., Lu, X., Moulin, P. (2014). Low-risk anomalies in global fixed income: Evidence from major broad markets, *Journal of Fixed Income* 23(4), 51-70.

Cazalet, Z., Roncalli, T. (2014). Facts and fantasies about factor investing, available at SSRN: https://papers.ssrn.com/abstract=2524547.

Chan, K.C., Chen, N.F. (1991). Structural and return characteristics of small and large firms, *Journal of Finance* 46(4), 1467-1484.

Chan, K.C., Chen, N.F., Hsieh, D.A. (1985). An exploratory investigation of the firm size effect, *Journal of Financial Economics* 14(3), 451-471.

Chan, L.K.C., Hamao, Y., Lakonishok, J. (1991). Fundamentals and stock returns in Japan, *Journal of Finance* 46(5), 1739-1764.

Chen, A., Pong, E., Wang, Y. (2018). Accessing the China a-shares market via minimum-variance investing, *Journal of Portfolio Management* 45(1), 106-117.

Chen, N.F., Roll, R., Ross, S.A. (1986). Economic forces and the stock market, *Journal of Business* 59(3), 383-403.

Chen, A.Y., Velikov, M. (2019). Accounting for the anomaly zoo: A trading cost perspective, available at SSRN: https://papers.ssrn.com/abstract=3073681.

Chen, N.F., Zhang, F. (1998). Risk and return of value stocks, *Journal of Business* 71(4), 501-535.

Chen, A.Y., Zimmermann, T. (2018). Publication bias and the cross-section of stock returns, available at SSRN: https://ssrn.com/abstract=2802357.

Chordia, T., Goyal, A., Saretto, A. (2019). Anomalies and false rejections, available at SSRN: https://papers.ssrn.com/abstract=3017677.

Chui, A.C.W., Titman, S., Wei, K.C.J. (2010). Individualism and momentum around the world, *Journal of Finance* 65(1), 361-392.

Clarke, R.G., De Silva, H., Thorley, S. (2016). Fundamentals of efficient factor investing, *Financial Analysts Journal* 72(6), 9-26.

Cochrane, J.H. (1991a). Production-based asset pricing and the link between stock returns and economic fluctuations, *Journal of Finance* 46(1), 209-237.

Cochrane, J.H. (1991b). Volatility tests and efficient markets: A review Essay, *Journal of Monetary Economics* 27(3), 463-485.

Cochrane, J.H. (1996). A cross-sectional test of an investment-based asset pricing model, *Journal of Political Economy* 104(3), 572-621.

Cochrane, J.H. (2011). Presidential address: Discount rates, *Journal of Finance* 66(4), 1047-1108.

Connor, G. (1995). The three types of factor models: A comparison of their explanatory power, *Financial Analyst Journal* 51(3), 42-46.

Conrad, J., Kaul, G. (1998). An anatomy of trading strategies, *Review of Financial Studies* 11(3), 489-519.

Cooper, M.J., Gulen, H., Schill, M.J. (2008). Asset growth and the cross-section of stock returns, *Journal of Finance* 63(4), 1609-1651.

Da, Z., Shive, S. (2018). Exchange traded funds and asset return correlations, *European Financial Management* 24(1), 136-168.

Daniel, K., Hirshleifer, D., Subrahmanyam, A. (1998). Investor psychology and security market under- and overreactions, *Journal of Finance* 53(6), 1839-1885.

Daniel, K., Hirshleifer, D., Sun, L. (2020). Short- and long-horizon behavioral factors, *Review of Financial Studies* 33(4), 1673-1736.

Daniel, K., Moskowitz, T. (2016). Momentum crashes, *Journal of Financial Economics* 122(2), 221-247.

Dannhauser, C.D. (2017). The impact of innovation: Evidence from corporate bond exchange-traded funds (ETFs), *Journal of Financial Economics* 125(3), 537-560.

De Bondt, W.F.M, Thaler, R. (1985). Does the stock market overreact?, *Journal of Finance* 40(3), 793-805.

De Groot, W., Huij, J., Zhou, W. (2012). Another look at trading costs and reversal profits, *Journal of Banking and Finance* 36(2), 371-382.

Dichtl, H., Drobetz, W., Lohre, H., Rother, C., Vosskamp, P. (2019). Optimal timing and tilting of equity factors, *Financial Analyst Journal* 75(4), 84-102.

Dimson, E., Marsh, P., Staunton, M. (2017). Factor-based investing: The long-term evidence, *Journal of Portfolio Management* 43(5), 15-37.

Elton, E.J., Gruber, M.J., Agrawal, D., Mann, C. (2001). Explaining the rate spread on corporate bonds, *Journal of Finance* 56(1), 247-277.

Engelberg, J., McLean, R.D., Pontiff, J. (2018). Anomalies and news, *Journal of Finance* 73(5), 1971-2001.

Eun, C.S., Lai, S., de Roon, F.A., Zhang, Z. (2010). International diversification with factor funds, *Management Science* 56(9), 1500-1518.

Falkenstein, E. G. (2009). Risk and return in general: Theory and evidence, available at SSRN: https://papers.ssrn.com/abstract=1420356.

Fama, E.F. (1970). Efficient capital markets: A review of theory and empirical work, *Journal of Finance* 25(2), 383-417.

Fama, E.F., Fisher, L., Jensen, M.C., Roll, R. (1969). The adjustment of stock prices to new information, *International Economic Review* 10(1), 1-21.

Fama, E.F., French, K.R. (1989). Business conditions and expected returns on stocks and bonds, *Journal of Financial Economics* 25(1), 23-49.

Fama, E.F., French, K.R. (1992). The cross-section of expected stock returns, *Journal of Finance* 47(2), 427-465.

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

Fama, E.F., French, K.R. (1995). Size and book-to-market factors in earnings and returns, *Journal of Finance* 50(1), 131-155.

Fama, E.F., French, K.R. (1996). Multifactor explanations of asset pricing anomalies, *Journal of Finance* 51(1), 55-84.

Fama, E.F., French, K.R. (1998). Value versus growth: The international evidence, *Journal of Finance* 53(6), 1975-1999.

Fama, E.F., French, K.R. (2006). Profitability, investment and average returns, *Journal of Financial Economics* 82(3), 491-518.

Fama, E.F., French, K.R. (2008). Dissecting Anomalies, *Journal of Finance* 63(4), 1653-1678.

Fama, E.F., French, K.R. (2012). Size, value, and momentum in international stock returns, *Journal of Financial Economics* 105(3), 457-472.

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

Fama, E.F., French, K.R. (2016). Dissecting anomalies with a five-factor model, *Review of Financial Studies* 29(1), 69-103.

Fama, E.F., French, K.R. (2017). International tests of a five-factor asset pricing model, *Journal of Financial Economics* 123(3), 441-463.

Fama, E.F., French, K.R. (2018). Choosing factors, *Journal of Financial Economics* 128(2), 234-252.

Fama, E.F., MacBeth, J.D. (1973). Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy* 81(3), 607-636.

Fanelli, D. (2013). Positive results receive more citations, but only in some disciplines, *Scientometrics* 94(2), 701-709.

Farcomeni, A. (2007). A review of modern multiple hypothesis testing, with particular attention to the false discovery proportion, *Statistical Methods in Medical Research* 17(4), 347-388.

Feng, G., Giglio, S., Xiu, D. (2020). Taming the factor zoo: A test of new factors, *Journal of Finance*, forthcoming.

Ferson, W.E., Harvey C.R. (1991). The variation of economic risk premiums, *Journal of Political Economy* 99(2), 385-415.

Fitzgibbons, S., Friedman, J., Pomorski, L., Serban, L. (2017). Long-only style investing: Don’t just mix, integrate, *Journal of Investing* 26(4), 153-164.

Fraser-Jenkins, I., Guerrini, A., Harmsworth, A., Diver, M., McCarthy, S., Stancikas, R., Hughes, M. (2016). Global quantitative strategy: How to combine factors? It depends why you are doing it, London: Sanford Bernstein.

Frazzini, A. (2006). The disposition effect and underreaction to news, *Journal of Finance* 61(4), 2017-2046.

Frazzini, A., Israel, R., Moskowitz, T.J. (2018). Trading costs, available at SSRN: https://papers.ssrn.com/abstract= 3229719.

Frazzini, A., Pedersen, L.H. (2014). Betting against beta, *Journal of Financial Economics* 111(1), 1-25.

French, K.R. (2008). Presidential address: The cost of active investing, *Journal of Finance* 63(4), 1537-1573.

Freyberger, J., Neuhierl, A., Weber, M. (2018). Dissecting characteristics nonparametrically, available at SSRN: https://papers.ssrn.com/abstract=2820700.

Garleanu, N., Panageas, S., Yu, J. (2012). Technological growth and asset pricing, *Journal of Finance* 67(4), 1265-1292.

George, T.J., Hwang, C.Y. (2004). The 52-Week high and momentum investing, *Journal of Finance* 59(5), 2145-2176.

Gibbons, M.R., Ross, S.A., Shanken, J. (1989). A test of the efficiency of a given portfolio, *Econometrica* 57(5), 1121-1152.

Glosten, L., Nallareddy, S., Zou, Y. (2016). ETF trading and informational efficiency of underlying securities, available at SSRN: https://papers.ssrn.com/abstract=2846157.

Glushkov, D. (2016). How smart are smart beta exchange-traded funds? Analysis of relative performance and factor exposure, *Journal of Investment Consulting* 17(1), 50-74.

Gordon, M.J., Shapiro, E. (1956). Capital equipment analysis: The required rate of profit, *Management Science* 3(1), 102-110.

Green, J., Hand, J.R.M, Zhang, X.F. (2013). The supraview of return predictive signals, *Review of Accounting Studies* 18(3), 692-730.

Green, J., Hand, J.R.M, Zhang, X.F. (2017). The Characteristics that provide independent information about average U.S. monthly stock returns, *Review of Financial Studies* 30(12), 4389-4436.

Griffin, J.M. (2002). Are the Fama and French factors global or country specific?, *Review of Financial Studies* 15(3), 783-803.

Grinblatt, M., Han, B. (2005). Prospect theory, mental accounting, and momentum, *Journal of Financial Economics* 78(2), 311-339.

Grossman, S.J., Stiglitz, J.E. (1980). On the impossibility of informationally efficient markets, *American Economic Review* 70(3), 393-408.

Gu, S., Kelly, B., Xiu, D. (2019). Empirical asset pricing via machine learning, available at SSRN: https://papers.ssrn.com/abstract=3159577.

Haddad, V., Kozak, S., Santosh, S. (2019). Factor timing, available at SSRN: https://papers.ssrn.com/abstract=2945667.

Han, X., Li, K., Li, Y. (2018). Investor overconfidence and the security market line: New evidence from China, available at SSRN: https://papers.ssrn.com/abstract=3284886.

Harvey, C.R. (2017). Presidential address: The scientific outlook in financial economics, *Journal of Finance* 72(4), 1399-1440.

Harvey, C.R., Liu, Y. (2015). Backtesting, *Journal of Portfolio Management* 42(1), 13-28.

Harvey, C.R., Liu, Y. (2018). Lucky factors, available at SSRN: https://papers.ssrn.com/abstract=2528780.

Harvey, C.R., Liu, Y. (2019). A census of the factor zoo, available at SSRN: https://papers.ssrn.com/abstract=3341728.

Harvey, C.R., Liu, Y., Zhu, H. (2016). … and the cross-section of expected returns, *Review of Financial Studies* 29(1), 5-68.

Harvey, C.R., Siddique, A. (2000). Conditional skewness in asset pricing tests, *Journal of Finance* 55 (3), 1263-1295.

Haugen, R.A., Heins, J.A. (1975). Risk and the rate of return on financial assets: Some old wine in new bottles, *Journal of Financial and Quantitative Analysis* 10(5), 775-784.

He, Z., Kelly, B., Manela, A. (2017). Intermediary asset pricing: New evidence from many asset classes, *Journal of Financial Economics* 126(1), 1-35.

Hodges, P., Hogan, K., Peterson, J.R., Ang, A. (2017). Factor timing with cross-sectional and time-series predictors, *Journal of Portfolio Management* 44(1), 30-43.

Hong, H., Li, W., Ni, S.X., Scheinkman, J.A., Yan, P. (2016). Days to cover and stock returns, available at SSRN: https://papers.ssrn.com/abstract=2568768.

Hong, H., Lim, T., Stein, J.C. (2000). Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies, *Journal of Finance* 55(1), 265-295.

Hong, H., Stein, J.C. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets, *Journal of Finance* 54(6), 2143-2184.

Hou, K., Mo, H., Xue, C., Zhang, L. (2018a). q^5, available at SSRN: https://papers.ssrn.com/abstract=3191167.

Hou, K., Mo, H., Xue, C., Zhang, L. (2019). Which factors?, *Review of Finance* 23(1), 1-35.

Hou, K., Xue, C., Zhang, L. (2015). Digesting anomalies: An investment approach, *Review of Financial Studies* 28(3), 650-705.

Hou, K., Xue, C., Zhang, L. (2018b). Replicating anomalies, *Review of Financial Studies*, forthcoming.

Houweling, P., Van Zundert, J. (2017). Factor investing in the corporate bond market, *Financial Analysts Journal* 73(2), 100-115.

Hsu, J.C., Kalesnik, V., Kose, E. (2019). What is quality?, *Financial Analyst Journal* 75(2), 44-61.

Hsu, J.C., Kalesnik, V., Viswanathan, V. (2015). A framework for assessing factors and implementing smart beta strategies, *Journal of Index Investing* 6(1), 89-97.

Huij, J., Lansdorp, S., Blitz, D., Van Vliet, P. (2014). Factor investing: Long-only versus long-short, available at SSRN: https://papers.ssrn.com/abstract=2417221.

Hwang, S., Rubesam, A. (2018). Searching the factor zoo, available at SSRN: https://papers.ssrn.com/abstract=3100811.

Ilmanen, A., Kizer, J. (2012). The death of diversification has been greatly exaggerated, *Journal of Portfolio Management* 38(3), 15-27.

Ioannidis, J.P.A. (2005). Why most published research findings are false?, *PLoS Medicine* 2(8): e124, 696-701.

Israel R., Moskowitz, T. (2013). The role of shorting, firm size, and time on market anomalies, *Journal of Financial Economics* 108(2), 275-301.

Jacobs, H. (2015). What explains the dynamics of 100 anomalies?, *Journal of Banking and Finance* 57, 65-85.

Jacobs, B.I., Levy, K.N. (1993). Long short equity investing, *Journal of Portfolio Management* 20(1), 52-63.

Jacobs, H., Müller, S. (2020). Anomalies across the globe: Once public, no longer existent?, *Journal of Financial Economics* 135(1), 213-230.

Jaffe, J., Keim, D.B., Westerfield, R. (1989). Earnings yields, market values, and stock returns, *Journal of Finance* 44(1), 135-148.

Jegadeesh, N., Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency, *Journal of Finance* 48(1), 65-91.

Jegadeesh, N., Titman, S. (2001). Profitability of momentum strategies: An evaluation of alternative explanations, *Journal of Finance* 56(2), 699-720.

Johnson, T. (2002). Rational momentum effects, *Journal of Finance* 57(2), 585-608.

Jones, C.M., Lamont, O.A. (2002). Short-sale constraints and stock returns, *Journal of Financial Economics* 66(2-3), 207-239.

Jordan, B.D., Riley, T.B. (2015). Volatility and mutual fund manager skill, *Journal of Financial Economics* 118(2), 289-298.

Joshipura, M., Joshipura, N. (2016). The volatility effect: Evidence from India, *Applied Finance Letters* 5(1), 12-27.

Kahneman, D., Tversky, A. (1979). Prospect theory: An analysis of decision under risk, *Econometrica* 47(2), 263-291.

Kalesnik, V., Linnainmaa, J. (2018). Ignored Risks of Factor Investing, retrieved from: https://www.researchaffiliates.com/documents/Ignored%20Risks%20of%20Factor%20Investing.pdf, 17th February 2020.

Karolyi, A.G. (2016). Home bias, an academic puzzle, *Review of Finance* 20(6), 2049-2078.

Kelly, B.T., Pruitt, S., Su, Y. (2019). Characteristics are covariances: A unified model of risk and return, *Journal of Financial Economics* 134(3), 501-524.

Khan, M. (2011). Conceptual foundations of capital market anomalies, in: L. Zacks (Ed.), *The handbook of equity market anomalies: Translating market inefficiencies into effective investment strategies*, Hoboken: John Wiley & Sons, 1-22.

Khandani, A. E., Lo, A.W. (2011). What happened to the quants in August 2007? Evidence from factors and transaction data, *Journal of Financial Markets* 14(1), 1-46.

Kisser, M. (2014). What explains the gross profitability premium?, available at SSRN: https://papers.ssrn.com/abstract=2526686.

Koedijk, K.G., Slager, A.M.H., Stork, P.A. (2016). Investing in systematic factor premiums, *European Financial Management* 2(2), 193-234.

Koijen, R.S.J., Moskowitz, T.J., Pedersen, L.H., Vrugt, E.B. (2018). Carry, *Journal of Financial Economics* 127(2), 197-225.

Korajczyk, R.A., Sadka, R. (2004). Are momentum profits robust to trading costs?, *Journal of Finance* 59(3), 1039-1082.

Kozak, S., Nagel, S., Santosh, S. (2018). Interpreting factor models, *Journal of Finance* 73(3), 1183-1223.

Kozak, S., Nagel, S., Santosh, S. (2020). Shrinking the cross-section, *Journal of Financial Economics* 135(2), 271-292.

Lakonishok, J., Shleifer, A., Vishny, R. (1994). Contrarian investment, extrapolation, and risk, *Journal of Finance* 49(5), 1541-1578.

Lee, W. (2017). Factors timing factors, *Journal of Portfolio Management* 43(5), 66-71.

Leippold, M., Rueegg, R. (2018). The mixed vs the integrated approach to style investing: Much ado about nothing?, *European Financial Management* 24(5), 829-855.

Lemperier, Y., Deremble, C., Nguyen, T.T., Seager, P., Potters, M., Bouchaud, J.P. (2017). Risk premia: Asymmetric tail risks and excess returns, *Quantitative Finance* 17(1), 1-14.

Lesmond, D.A., Schill, M.J., Zhou, C. (2004). The illusory nature of momentum profits, *Journal of Financial Economics* 71(2), 349-380.

Lettau, M., Madhavan, A.N. (2016). ETF 101 for economists, *Journal of Economic Perspectives* 32(1), 135-154.

Li, X., Sullivan, R.N., Garcia-Feijoo, L. (2014). The limits to arbitrage and the low-volatility anomaly, *Financial Analysts Journal* 70(1), 52-63.

Li, X., Sullivan, R.N., Garcia-Feijoo, L. (2016). The low-volatility anomaly: Market evidence on systematic risks versus mispricing, *Financial Analyst Journal* 72(1), 36-47.

Liew, J., Vassalou, M. (2000). Can book-to-market, size and momentum be risk factors that predict economic growth?, *Journal of Financial Economics* 57(2), 221-245.

Linnainmaa, J.T., Roberts, M.R. (2018). The history of the cross-section of stock returns, *Review of Financial Studies* 31(7), 2606-2649.

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets, *Review of Economics and Statistics* 47(1), 13-37.

Liu, W. (2006). A liquidity-augmented capital asset pricing model, *Journal of Financial Economics* 82(3), 631-671.

Lo, A.W., MacKinlay, A.C. (1990). Data snooping biases in tests of financial asset pricing models, *Review of Financial Studies* 3(3), 431-467.

Madhavan, A.N., (2016). Exchange-traded funds and the new dynamics of investing, New York: Oxford University Press.

Madhavan, A.N., Sobczyk, A. (2016) Price dynamics and liquidity of exchange-traded funds, *Journal of Investment Management* 14(2), 1-17.

Malkiel, B.G. (2014). Is smart beta really smart?, *Journal of Portfolio Management* 40(5), 127-134.

Marshall, B.R., Nguyen, N.H., Visaltanachoti, N. (2013). ETF arbitrage: Intraday evidence, *Journal of Banking and Finance* 37(9), 3486-3498.

Marshall, B.R., Nguyen, N.H., Visaltanachoti, N. (2015). ETF liquidity, Working Paper, Massey University.

McLean, R.D., Pontiff, J. (2016). Does academic research destroy stock return predictability?, *Journal of Finance* 71(1), 5-31.

Miller, E.M. (2001). Why the low returns to beta and other forms of risk?, *Journal of Portfolio Management* 27(2), 40-55.

Miller, M.H., Modigliani, F. (1961). Dividend policy, growth, and the valuation of shares, *Journal of Business* 34(4), 411-433.

Molk, P., Partnoy, F. (2019). Institutional investors as short sellers?, available at SSRN: https://papers.ssrn.com/abstract=3309731.

Mondello, E. (2013). Portfoliomanagement, Wiesbaden: Springer Gabler.

Moskowitz, T.J., Ooi, Y.H., Pedersen, L.H. (2012). Time series momentum, *Journal of Financial Economics* 104(2), 228-250.

Mossin, J. (1966). Equilibrium in a capital asset market, *Econometrica* 34(4), 768-783.

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

Novy-Marx, R., Velikov, M. (2016). A taxonomy of anomalies and their trading costs, *Review of Financial Studies* 29(1), 104-147.

Odean, T. (1998). Are investors reluctant to realize their losses?, *Journal of Finance* 53(5), 1775-1798.

Ohlson, J.A. (1995). Earnings, book values, and dividends in security valuation, *Contemporary Accounting Research* 11(2), 661-687.

Ouysse, R., Kohn, R. (2010). Bayesian variable selection and model averaging in the arbitrage pricing theory model, *Computational Statistics and Data Analysis* 54(12), 3249-3268.

Pace, D., Hili, J., Grima, S. (2016). Active versus passive investing: An empirical study on the US and European mutual funds and ETFs, in: Emerald Group Publishing Limited (Ed.), *Contemporary Issues in Bank Financial Management* 97, S. 1-35.

Pastor, L., Stambaugh, R.F. (2003). Liquidity risk and expected stock returns, *Journal of Political Economy* 111(3), 642-685.

Patton, A.J., Weller, B.M. (2019). What you see is not what you get: The costs of trading market anomalies, *Journal of Financial Economics,* forthcoming.

Pedersen, L.H. (2009). When everyone runs for the exit, available at SSRN: https://papers.ssrn.com/abstract=1463890.

Perrins, A. (2019). An introduction to factor investing, retrieved from: https://www.savvyinvestor.net/blog/factor-investing-academic-research-and-papers, 26th January 2020.

Petajisto, A. (2017). Inefficiencies in the pricing of exchange-traded funds, *Financial Analysts Journal* 73(1), 24-54.

Piccotti, L.R., (2014). An ETF premium puzzle and a market segmentation explanation, Working Paper, Rutgers University.

Polya, G. (1920). Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem, *Mathematische Zeitschrift* 8(3-4), 171-181.

Poteshman, A.M., Serbin, V. (2003). Clearly irrational financial market behavior: Evidence from the early exercise of exchange traded stock options, *Journal of Finance* 58(1), 37-70.

Preinreich, G.A.D. (1938). Annual survey of economic theory: The theory of depreciation, *Econometrica* 6(3), 219-241.

Pukthuanthong, K., Roll, R., Subrahmanyam, A. (2019). A protocol for factor identification, *Review of Financial Studies* 32(4), 1573-1607.

Richie, N., Daigler, R.T., Gleason, K.C. (2008) the limits to stock index arbitrage: Examining S&P 500 futures and SPDRs, *Journal of Futures Markets* 28(12), 1182-1205.

Robeco (2014). How factor investing fits into active vs passive?, retrieved from: https://www.robeco.com/en/insights/2014/10/how-factor-investing-fits-into-active-vs-passive.html, 09th February 2020.

Ross, S.A. (1976). The arbitrage theory of capital asset pricing, *Journal of Economic Theory* 13(3), 341-360.

Sagi, J., Seasholes, M. (2007). Firm-specific attributes and the cross-section of momentum, *Journal of Financial Economics* 84(2), 389-434.

Scheffé, H. (1959). The analysis of variance, New York: John Wiley & Sons.

Schwert, G.W. (2003). Anomalies and market efficiency, in: Constantinides, G., Harris, M., Stulz, R.M. (Eds.), *Handbook of the Economics of Finance*, Amsterdam: Elsevier, 939-974.

Shanken, J. (1990). Intertemporal asset pricing: An empirical investigation, *Journal of Econometrics* 45(1-2), 99-120.

Shanken, J. Weinstein, M.I. (2006). Economic forces and the stock market revisited, *Journal of Empirical Finance* 13(2), 129-144.

Sharpe, W.F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk, *Journal of Finance* 19(3), 425-442.

Shefrin, H., Statman, M. (1985). The disposition to sell winners too early and ride losers too long: Theory and evidence, *Journal of Finance* 40(3), 777-790.

Shiller, R.J. (2003). From efficient markets theory to behavioral finance, *Journal of Economic Perspectives* 17(1), 83-104.

Shleifer, A., Vishny, R.W. (1997). The limits of arbitrage, *Journal of Finance* 52(1), 35-55.

Siddiqi, H. (2018). Anchoring-adjusted capital asset pricing model, *Journal of Behavioral Finance* 19(3), 249-270.

Sloan, R. (1996). Do stock prices fully reflect information in accruals and cash flows about future earnings?, *Accounting Review* 71(3), 289-315.

Stambaugh, R.F., Yuan, Y. (2017). Mispricing factors, *Review of Financial Studies* 30(4), 1270-1315.

Suhonen, A., Lennkh, M., Perez, F. (2017). Quantifying backtest overfitting in alternative beta strategies, *Journal of Portfolio Management* 43(2), 90-104.

Sullivan, R., Timmermann, A., White, H. (1999). Data-snooping, technical trading rule performance, and the bootstrap, *Journal of Finance* 54(5), 1647-1691.

Sullivan, R., Timmermann, A., White, H. (2001). Dangers of data mining: The case of calendar effects in stock returns, *Journal of Econometrics* 105(1), 249-286.

Thaler, R. (1985). Mental accounting and consumer choice, *Marketing Science* 4(3), 199-214.

Titman, S., Wei, K.C.J., Xie, F. (2004). Capital investments and stock returns, *Journal of Financial and Quantitative Analysis* 39(4), S. 677-700.

Titman, S., Wei, K.C.J., Xie, F. (2013). Market development and the asset growth effect: International evidence, *Journal of Financial and Quantitative Analysis* 48(5), 1405-1432.

Tobin, J. (1969). A general equilibrium approach to monetary theory, *Journal of Money, Credit and Banking* 1(1), 15-29.

Tukey, J.W. (1951). Reminder sheets for “Discussion of paper on multiple comparisons by Henry Scheffé.”, in: Braun, H.I. (Ed), *The collected works of John W. Tukey VIII. Multiple Comparisons: 1948-1983*, New York: Chapman and Hall, 469-475.

Tversky, A., Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases, *Science* 185(4157), 1124-1131.

Van Gelderen, E., Huij, J. (2014). Academic knowledge dissemination in the mutual fund industry: Can mutual funds successfully adopt factor investing strategies?, *Journal of Portfolio Management* 40(4), 157-167.

Van Vliet, P., De Koning, J. (2017). High returns from low risk: A remarkable stock market paradox, New York: John Wiley & Sons.

Vassalou, M., Xing, X. (2004). Default risk in equity returns, *Journal of Finance* 59(2), 831-868.

Wahal, S. (2019). The profitability and investment premium: Pre-1963 evidence, *Journal of Financial Economics* 131(2), 362-377.

Walkshäusl, C. (2014). International low-risk investing, *Journal of Portfolio Management* 41(1), 45-56.

Watanabe, A., Xu, Y., Yao, T., Yu, T. (2013). The asset growth effect: Insights from international equity markets, *Journal of Financial Economics* 108(2), 529-563.

Wermers, R. (1999). Mutual fund herding and the impact on stock prices, *Journal of Finance* 54(2), 581-622.

White, H. (2000). A reality check for data snooping, *Econometrica* 68(5), 1097-1126.

Yan, X.S., Zheng, L. (2017). Fundamental analysis and the cross-section of stock returns: A data-mining approach, *Review of Financial Studies* 30(4), 1382-1423.

Yogo, M. (2006). A consumption-based explanation of expected stock returns, *Journal of Finance* 61(2), 539-580.

Zhang, L. (2005). The value premium, *Journal of Finance* 60(1), 67-103.

Zhang, X.F. (2006). Information uncertainty and stock returns, *Journal of Finance* 61(1), 105-136.

Zivot, E., Wang, J. (2003). Modeling financial time series with S-Plus®, New York: Springer.