WISEflow
Europe/Oslo(CEST)
30 Jun 2022

Handelshøyskolen BI

GRA 19703 Master Thesis

Thesis Master of Science 100% - W

Predefinert informasjon

Startdato:	16-01-2022 09:00	**Termin:**	202210
Sluttdato:	01-07-2022 12:00	**Vurderingsform:**	Norsk 6-trinns skala (A-F)
Eksamensform:	T		
Flowkode:	202210||10936||IN00||W||T		
Intern sensor:	(Anonymisert)		

Deltaker

Jørgen Tetlie, Henrik Armand Hamnnes

Informasjon fra deltaker

Tittel *:	Are Factor ETFs a Valuable Investment?
Navn på veileder *:	Chunyu Yang

Inneholder besvarelsen konfidensielt materiale?:	Nei	**Kan besvarelsen offentliggjøres?:**	Ja

Gruppe

Gruppenavn:	(Anonymisert)
Gruppenummer:	296
Andre medlemmer i gruppen:	

BI Norwegian Business School

Master Thesis

Are Factor ETFs a Valuable Investment?

Name of Students:

Henrik Armand Hamnnes

Jørgen Tetlie

Name of Supervisor:

Chunyu Yang

Hand-in Date:

01.07.2022

Campus:

BI Oslo

Examination Code and Name:

GRA 19703 Master Thesis

Study Programme:

Master of Science in Business – Major in Finance

Abstract

In this thesis, we study the performance of factor ETFs. We estimate risk-adjusted returns (alpha) using renowned asset pricing models to capture different sources of systematic risk. Moreover, we study whether factor ETFs add value to different types of investors using different performance metrics. In addition, we study how style characteristics of factor ETFs can be exploited to diversify investors' portfolios and improve their risk-return tradeoff.

We find that, on average, factor ETFs do not generate significant risk-adjusted returns. Furthermore, we find that most funds do not add value to different types of investors, apart from growth and momentum factor ETFs. We conclude that the average investor obtains a better risk-return tradeoff by holding the market portfolio. However, certain factors like growth and momentum can be exploited to improve the overall risk-adjusted performance.

Acknowledgements

We want to express our most profound appreciation to our supervisor *Chunyu Yang* for his helpful comments and support. In addition, we want to extend our sincere thanks to *Bruno Gerard* for providing insightful content on portfolio performance evaluation in his *Topics in Portfolio Management* course. Finally, thanks should also go to *Sven Klingler*, who impacted and inspired us to research factor ETFs by arranging *The Trading Game* in his *Asset Management* course.

Table of Contents

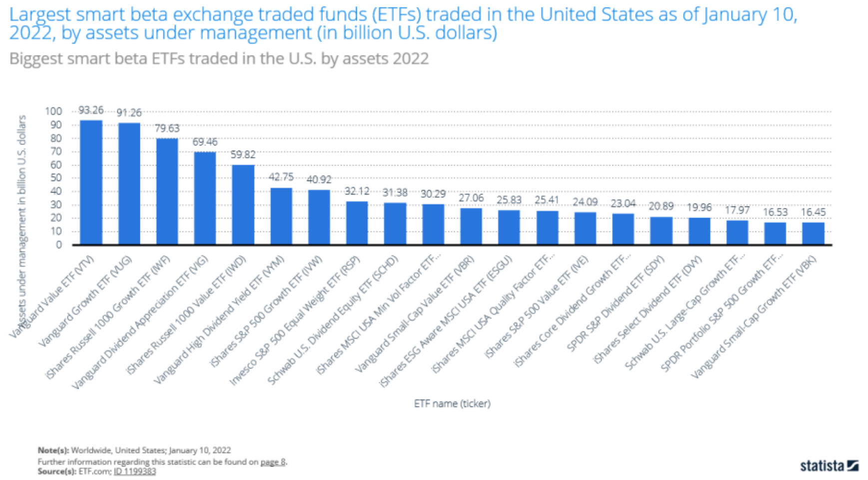
1. Introduction

Equity risk factors have generated abnormal risk-adjusted returns in empirical asset pricing and outperformed the value-weighted market portfolio. In this context, a relatively new financial innovation is the exchange-traded funds (ETFs), which overvalue investments in assets with specific characteristics instead of replicating a market index. For example, common equity risk factors may include size, growth, value, momentum, idiosyncratic volatility, quality, et cetera.

In this thesis, we use asset pricing models to estimate the expected risk-adjusted returns of factor ETFs to examine if investors can exploit them to diversify their portfolios. Because ETFs cannot short-sell assets, the idea is to determine if factor investing with long-only factors proceed a good reward to risk tradeoff. We examine if factor ETFs generate significant alphas in this context. Moreover, we use different risk-adjusted performance measures to examine whether the funds add value to different types of investors. Accordingly, we sort the factor ETFs on style characteristics and compare them to their peers and respective benchmarks.

We are motivated to research the combined topics of factor investing and ETFs. The first ETF was launched in the 1990s, and its popularity has surged rapidly. *The EY Global ETF Report, (2017)* predicted that factor ETFs would achieve a total of $1.2 trillion in assets under management (*AUM*) by 2020. The expectations were met and exceeded. As of 2022, factor ETFs are estimated to make up $1.5 trillion of the financial markets, with the biggest US funds composing approximately $800 billion in AUM (*see Figure 1*).

Figure 1: Biggest Factor ETFs in the US by AUM (2022).



ETF.com. (January 10, 2022). Largest smart beta exchange traded funds (ETFs) traded in the United States as of January 10, 2022, by assets under management (in billion US dollars) [Graph]. In Statista. Retrieved June 2, 2022, from https://www-statista-com.ezproxy.library.bi.no/statistics/1199383/largest-smart-beta-etfs-traded-usa/

Jacobs & Levy, (2015) and Malkiel, (2014) argue that factors cannot be captured effectively in real-life portfolios. Ang, (2014) and Huij et al., (2014), on the other hand, show that factors work in real-life. Given the popularity of ETFs and conflicting research on the topic, we see this as a great opportunity to contribute to existing literature. Consequently, we intend to provide valuable insights for both institutional and individual investors by researching the following topic:

Are Factor ETFs a Valuable Investment?

To best answer the research question, we examine diversified factor ETFs across international markets over five years, from Jan 2017 to Dec 2021. Moreover, the study comprises several multifactor and pure factor ETFs. The factors we sort for as style characteristics of the ETFs are the ones most known from academia. Respectively, size, value, growth, quality, volatility, dividend, momentum, and blended multifactor funds. Hence, the thesis analyses and compares multiple factors implemented through ETFs.

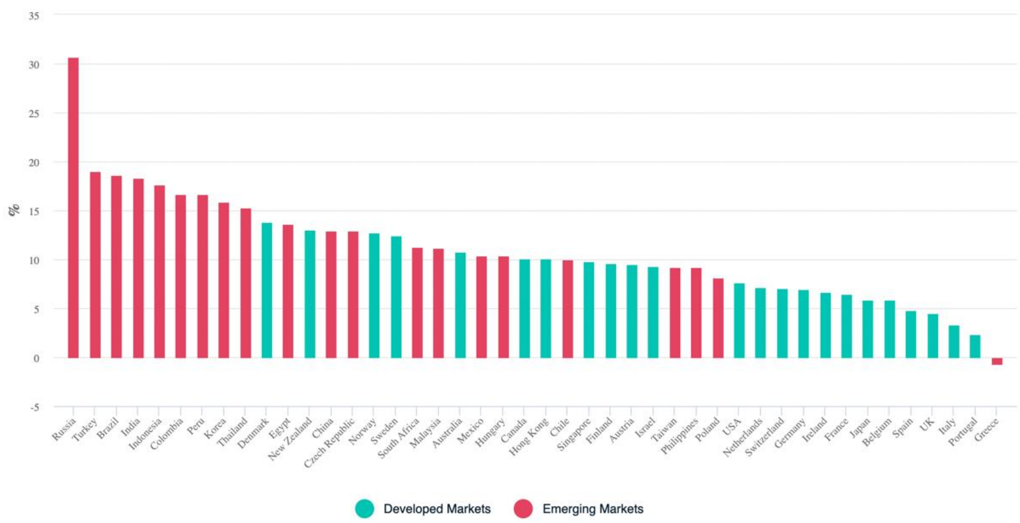
There are several reasons for our choice of topic. First and foremost, we wanted to research a topic we find truly exciting and where we could utilize our knowledge and skills within investments, portfolio management, and financial theory. Second, and equally important, we wanted to contribute new insights into which style characteristics of factor ETFs investors should consider valuable. Moreover, due to the experienced significant growth in the number of factor funds during the 21st century, we predict that factor ETFs will become even more relevant in the investment sphere in the future.

We compare the performance of factor ETFs to regression-based benchmarks that capture different sources of systematic risk. Moreover, we compute and compare risk-adjusted performance metrics between peers and the overall market. We find no conclusive empirical evidence from our analysis that factor ETFs outperform their respective benchmarks or the overall market. Hence, investors obtain a better risk-return tradeoff from holding the market portfolio rather than the average factor ETF. Correspondingly, we do not consider the average ETF a valuable investment.

However, we find that growth factor ETFs outperform their respective benchmarks, overall market, and their peers prior to and during Covid-19. The pandemic of Covid-19 is a unique period in our analysis with high financial distress. During this period, factor ETFs such as momentum enhanced their risk-adjusted performance and outperformed the overall market.

There are, however, a few implications for our results. First, we have examined factor ETFs from international markets over several years. Thus, the comparison might not seem fair if a specific factor is overweighted in a good or bad-performing market compared to its peers. E.g., not all economies have performed equally, and equity premiums vary across economic regions (*see figure 2*). Thus, one style characteristic of factor ETFs can have outperformed another due to differences in the overall economies. This might be inflicted for the results, as we have performed regressions only differing on factors, not on regions/markets.

Figure 2: Historical Average Annual Equity Premium by Country.



Source: MSCI, Retrieved June 2, 2022, from (https://www.msci.com/research-and-insights/global-investing-trends/emerging-and-developed-market-equity-performance).

Another point to consider is that the world economy passes different stages over five years where different factors will perform better than others. So it could be that different factor ETFs are performing well for one year, while the following year it is performing poorly compared to their peers. For example, we find that this happened for idiosyncratic volatility factor ETFs during covid-19.

Another implication is that the data might be biased as we have included more data from the US than from other markets. The reason for this is that the US is the largest capital market in the world. Thus, it is natural that most ETFs are based there or invested there. Nevertheless, another reason is that when acquiring the data, we manually collected data for non-US ETFs from Refinitiv, as CRSP only provides data for securities listed on US exchanges. Consequently, this led to fewer data for non-US factor ETFs.

2. Literature Review and Theory

In this section, we provide and review relevant literature, theory, and necessary background information for our analysis.

2.1 Factor Investing

Factor investing is a relatively new approach to investment management that focuses on capturing the risk premiums that arise from exposure to systematic risk factors (Elton et al., 2017). In market equilibrium, there is a positive expected rate of return over the risk-free rate associated with identifiable factors. Therefore, every asset exposed to underlying factor risks earns a premium. As a result, investors can increase their exposure to such factors to obtain a higher expected average return. Correspondingly, factor investing is about improving the portfolio's beta and enhancing the alpha. Thus, we examine whether factor ETFs generate risk-adjusted expected returns above benchmark portfolios.

2.1.1 Factors vs. Anomalies

Factors are investment styles that deliver high returns over the long run (Ang, 2014). There are more than 400 documented factors in academic research, and some say that the rate of factor production is out of control (Harvey & Liu, 2019). What characterizes a factor is that it is persistent beyond the initial sample when documented and across different exchanges. For example, momentum and value are potent factors, as they are consistent across diverse markets and asset classes (Asness et al., 2013).

In comparison, anomalies derive from inefficiency in the market, such as lack of information causing mispricing. Anomalies do not persist and are corrected when made aware of them. Anomalies are found to generate significant returns when corporate news and announcements are released to the public (Engelberg et al., 2018). One of the main drivers of such anomaly returns is biased expectations, which are partly corrected upon the release of corporate news and announcements.

Factors are a proxy and premium for risk, even if risk premiums change over time, and many factors are approximately capturing the same risks. Accordingly, the most potent factors can be delimited to the most significant and persistent.

A recent literature review documented alpha-generating strategies by analyzing specific factors in European markets and found significant abnormal returns in factors such as value, momentum, and profitability (Bermejo et al., 2021). Moreover, their research found mixed, conditional, and combined strategies of factor investing to outperform pure factor strategies significantly.

Pure factor strategies are single-factor portfolios that invest in assets with high return metrics on one particular factor. However, these portfolios are proven to be suboptimal because they ignore the possibility that these assets may be unattractive from the perspective of other factors (Blitz & Vidojevic, 2019).

2.1.2 Long-Only vs. Long-Short Strategies

Several academic studies recommend investing in factor premiums beyond the classic market risk premium. However, an interesting research topic is whether factor investing can be best implemented using a long-only or long-short approach. Although a long-short approach is superior theoretically, a long-only strategy is a preferred alternative in most scenarios (Huij et al., 2014). This result is derived from accounting for practical issues such as benchmark restrictions, implementation costs, and factor decay. Academic research has also discovered the long-only approach to generate significant alphas in factor timing strategies (Leippold & Rueegg, 2021). These findings arise because factor timing strategies exploit the momentum and mean reversion in factor returns. In addition, a long-only approach usually has lower implementation costs than a long-short approach. However, in other financial markets, e.g., commodity futures, the advantage of long-short exposures has been documented to be superior to long-only (Miffre, 2016).

The superior approach has also been researched in academia by comparing hedge funds with actively managed long-only portfolios. The differences in alpha distributions and return-driven risk factors make a good comparison as hedge funds do not have any short-selling restrictions. Thereupon, hedge funds have been documented to generate more consistent alphas in both equity and bond asset classes than long-only portfolios, even in extreme market conditions (Kao, 2002). However, potential explanations for these findings are differences in portfolio structures, investment constraints, management fees, and lack of data reliability.

2.2 ETFs vs. Mutual Funds

The Economist once said that ETFs are one of the more successful financial innovations in recent decades and that their success is driven by cheapness and convenience (The Economist, 2013). ETFs are beneficial for investors as they can be traded instantaneously and as there is a large variety of funds. Thus, factor ETFs are a considerably convenient investment product for many investors.

In addition, the factor ETFs are relatively cheap as they are semi-actively managed. Factor ETFs are passively managed as there is no input from a manager or others, and it follows a prespecified rule of style. However, it is also active in the extent of regular portfolio rebalancing as the composition changes over through horizon. The expenses related to factor ETFs include an average management fee of approximately 50 basis points and transaction costs. The transaction costs are subject to trading commissions, bid-ask spreads, and market impact costs.

Factor investing through ETFs has a few advantages over mutual funds. First, ETFs have the advantage of immediate liquidity as they can be traded instantaneously. In comparison, mutual funds are traded at the same price by the end of the day. Second, ETFs have the advantage of tax efficiency because mutual funds must sell assets to pay out redeemed shares, and then the capital gains are passed through to remaining investors. Third, ETFs have the advantage of more transparency as their holdings are published daily and not quarterly like mutual funds. Last but not least, ETFs have the advantage that investors can short sell them, while mutual funds are a long-only product (Ang, 2014). However, only long positions can be added to ETFs and mutual funds because it is rule-based.

The prices quoted on an exchange for ETFs should correspond closely with the funds' net asset values (NAV) through no-arbitrage pricing. Although investors might risk paying a slightly different price from the NAV, ETFs are documented to be efficiently priced most of the time. The market price deviations from the

NAV are minor, last for a short time, and are not persistent (DeFusco et al., 2011; Engle & Sarkar, 2006).

However, Ang, (2014) imposes two main disadvantages of ETFs for individual investors. First, ETFs have the traditional disadvantage of being easily traded and therefore excessively traded by individual investors. It is well documented that individuals lose money when they excessively trade because they are too quick to realize their winners and reluctant to sell their losers (Barber & Odean, 2000; Odean, 1999). Hence, Factor ETFs are subject to be traded pro-cyclically. Second, the large variety of ETFs can easily lead to individuals holding overly narrow portfolios that do not give adequate diversification (Ang, 2014). However, many ETFs provide significant diversification, and there has been documented that combining a few ETFs can even replicate active fund performance (Filho et al., 2021).

The market risk factor is documented to explain a substantial part of the expected returns of factor ETFs (Apergis et al., 2022). Their findings document that the market factor explains a substantial part of the expected returns, with the remaining factors, except momentum, posting smaller or no contribution. In addition, style ETFs exhibit mixed results in capturing their referenced style, with almost all exhibiting non-neutral momentum. These findings interest investment managers, investors, risk managers, and stock exchanges.

2.3 Fear and Risk Premia: Covid-19

Durand et al., (2011) documented how market volatility captured by the VIX volatility index affects expected returns in the US. Their research found that the market risk premium and the value premium were highly sensitive to changes in the VIX volatility index. Unfortunately, their research is limited as their sample period ends before August 2007, prior to the global financial crisis. Thus, we do not know if their results are valid in extreme market conditions. However, they found that an increase in expected volatility is associated with flights to quality firms and increases in estimated required returns (Durand et al., 2011).

Our sample period includes the period of Covid-19, during which the VIX volatility index reached its all-time highest closing price. The global financial crisis is the only time the index has reached a higher price intraday.

Hence, such increases in expected market volatility might result in increased risk premiums during Covid-19. Accordingly, factor ETFs might have an increased estimated required return during our sample period. In addition, the exceedingly higher expectations in market volatility during Covid-19 might increase the performance of factor ETFs with high-quality factor exposures.

2.4 Similar Studies

Factor funds are sometimes referred to as smart beta funds as they seek to increase exposure to systematic risks. Frazzini & Pedersen, (2014) compared the performance of portfolios with high- and low-beta assets. As for performance measures, they used estimated alphas and Sharpe ratios. Their research discovered that portfolios with high-beta assets have significantly lower alphas and Sharpe ratios than portfolios of low-beta assets. Consequently, an even new factor was introduced, called betting against beta.

Mateus et al., (2020) analyzed smart beta ETFs' performance persistence from June 2000 to May 2017. As for performance measures, they used estimated alphas, Sharpe, Treynor, and Information ratios. Their analysis found that about 40% of the ETFs generated positive alpha. Moreover, their research documented that the performance of winners and losers persisted a year ahead in 7 out of 9 peer categories. The peer categories were growth, value, and blended funds sorted on small-, mid-, and large-cap sizes.

Glushkov, (2015) found that 60% of smart beta ETFs in the US outperformed their raw passive benchmarks but did not outperform their risk-adjusted benchmarks during 2003-2014. Moreover, this research examines style characteristics beyond Mateus et al., (2020), which interest us. However, due to the relatively new innovation of factor ETFs, there are many style categories of Glushkov, (2015) with a considerably low number of funds. Hence, the means and variances of style characteristics will likely deviate from our results because our analysis includes many more funds for each peer category of style characteristics in more recent years.

Kothari & Warner, (2001) studied the empirical properties of performance measures for mutual funds. Their methodology includes regression-based performance measures by utilizing multiple asset pricing models. Their study found it hard to detect substantial abnormal performance for factor funds due to the style differences of factor mutual funds compared to the value-weighted market portfolio.

Simons, (1998) summarizes risk-adjusted performance measures and states that two major issues must be addressed in any performance ranking. The first is the choice of an appropriate benchmark for comparison. The second is how to adjust expected returns for risk.

Arugaslan et al., (2008) investigated the risk-adjusted performance of international mutual funds. First, they use the S&P 500 and MSCI EAFE indices as appropriate benchmarks. Second, they adjust expected returns for risk by utilizing the M-squared measure. Their analysis discovered that losers outperformed several winners on a risk-adjusted basis. However, their analysis leaves out essential effects such as management fees.

Many studies utilize risk-adjusted measures in the performance ranking of funds with different style characteristics. However, there are very few similar studies of factor ETFs to our knowledge. Glushkov, (2015) and Mateus et al., (2020) are significant breakthroughs in this research topic. However, the paper of Mateus et al., (2020) is not yet published. Nevertheless, their paper is forthcoming in the *Journal of Asset Management.* Their study claim to provide the first evidence on factor ETF's performance persistence. Moreover, comparing some of our results to Glushkov, (2015) and Mateus et al., (2020) will be interesting, as we have different sample periods and include different style characteristics of funds.

2.5 Asset Pricing Theory and Models

2.5.1 Capital Asset Pricing Model (CAPM)

The capital asset pricing model (CAPM) is the first theory of factor risk, introduced by Sharpe, (1964) and Lintner, (1965). There are two essential definitions in factor theory: bad times and risk premium. CAPM defines bad times as times of low market returns. Risk premiums are defined as the excess return required from an investment in a risky asset over that required from a riskless asset. However, CAPM only includes the market risk premium. Thus, CAPM is a single factor model that explains the relationship between systematic risk and expected return.

Moreover, CAPM states that assets with poor performance during bad times are risky and, therefore, must reward their investors with a higher risk premium. In contrast, assets that do well during bad times are attractive and have a lower risk premium.

Numerous assumptions underpin the CAPM. Most importantly, the model assumes that investors have homogenous expectations and hold mean-variance efficient portfolios. Furthermore, the model assumes there is a frictionless market. Accordingly, the mean-variance efficient portfolio for the average investor will be the market portfolio. The market portfolio is a value-weighted portfolio of all assets available in the financial markets. However, the actual market portfolio is unobservable. Consequently, the model's validity is widely discussed in academia due to its poor empirical record. For this reason, when applying the model, we must use a proxy for the market portfolio.

Sharpe and Lintner assume investors can borrow and lend capital at risk-free rates. For any asset *i*, the CAPM can thus be written as:

$$E[R_i] = R_f + \beta_i\left(E[R_m] - R_f\right) \tag{1}$$

where

$E[R_i]$ is the expected return of the asset *i*.

R_f is the risk-free rate.

β_i is the beta of asset *i*, defined as the sensitivity of the expected excess asset returns to the expected excess market returns.

$E[R_m]$ is the expected return of the market portfolio.

$E[R_m] - R_f$ is the market risk premium.

Beta becomes a measure of risk by using the definitions of bad times and risk premium. During bad times with low market returns, an asset with a low beta will be expected to do better than an asset with a higher beta. Thus, investors require compensation for holding riskier assets and are compensated in the form of a risk premium.

The beta is calculated as:

$$\beta_i = \frac{Cov(R_i, R_m)}{Var(R_m)} = \rho_{i,m} \frac{\sigma_i}{\sigma_m} \tag{2}$$

where

$\rho_{i,m}$ is the correlation coefficient between the asset i and the market.

σ_i is the standard deviation for the asset i.

σ_m is the standard deviation for the market.

As the beta is obtained from the correlation between an asset and the market, it only captures the systematic risk. Correspondingly, investors are only compensated for exposure to systematic risk, not idiosyncratic risk. Systematic risk is inherent in the market and cannot be diversified. In contrast, idiosyncratic risk or unsystematic risk is asset-specific and can be diversified.

There are six key takeaways from the CAPM theory (Ang, 2014). 1) The market diversifies away idiosyncratic risk, which implies diversification works. 2) Every investor has their own optimal exposure to the market portfolio. 3) The average investor holds the market portfolio. 4) The market factor is priced in equilibrium under the CAPM assumptions. 5) The CAPM beta measures the risk of an asset. 6) Assets paying off in bad times when the market return is low are attractive, and these assets have lower risk premiums.

2.5.2 Arbitrage Pricing Theory (APT)

Arbitrage Pricing Theory (APT) is the inspiration for factor investing, introduced by Ross, (1976). Compared to CAPM, APT allows for multiple systematic risk factors in estimating expected returns. Thus, APT capture multiple definitions of bad times across many factors and states of nature (Ang, 2014). Accordingly, APT replaces the mean-variance market portfolio with a set of indexes consisting of observable portfolios that aims to capture multiple systematic risk factors and return attributes.

The APT model is derived from the following multifactor return-generating process:

$$R_i = \alpha_i + \sum_{j=1}^{J} \beta_{ij} I_j + e_i \tag{3}$$

where

α_i is constant and equals the expected level of return for asset *i* if all indexes have a value of zero.

I_j is the systematic risk factor captured by the j^{th} index that impacts the return on asset *i*.

β_{ij} is the sensitivity of stock *i*'s return to the j^{th} index.

e_i is the random error term with a mean equal to zero.

For an asset with no idiosyncratic risk, the APT model can be written as:

$$R_i = R_f + \sum_{j=1}^{J} \beta_{ij} \lambda_j \tag{4}$$

where

λ_j is the risk premium and extra expected return required because of an asset's sensitivity to the j^{th} attribute of the asset.

The law of one price underpins the APT model. Hence, two identical items cannot be sold at different prices. Moreover, the assumptions concerning utility theory in CAPM are not necessary. As a result, the APT model's description of equilibrium is significantly broader than CAPM in that asset pricing can be influenced by factors other than means and variances of asset returns. It is, however, necessary to assume homogenous expectations (Elton et al., 2017).

There are six key takeaways from the APT model apply to the multifactor models (Ang, 2014). 1) The tradeable version of a factor diversifies away idiosyncratic risk, which implies diversification works. 2) every investor has their own optimal exposure to different factor risks. 3) The average investor holds the market portfolio. 4) Risk premiums exist for each factor, assuming no arbitrage or equilibrium. 5) The risk of an asset is measured in terms of the factor exposures of that asset. Finally, 6) Assets paying off in bad times are attractive, and these assets have lower risk premiums.

Contrary to CAPM, the APT model does not define which factors attribute to an asset's expected return. Accordingly, the challenge is to separate factors from anomalies and choose which factors best explain the cross-section of returns. The currently dominant approach to capturing risk premiums and specific factors is empirical research of firm characteristics that identify sources of systematic risk exposure (Bodie et al., 2020). The following multifactor models approach this issue by including different factors that predict average stock returns well in empirical evidence.

2.5.3 Asset Pricing Models

2.5.3.1 CAPM

The expected return on an ETF at time *t* is derived from equation (1) and can be written as:

$$R_t^{ETF} = R_{f,t} + \beta_{Mkt}^{ETF}\left[R_{Mkt,t} - R_{f,t}\right] \tag{5}$$

The market factor beta from equation (2) can also be estimated by running the following time-series regression:

$$R_t^{ETF} - R_{f,t} = \hat{\alpha}_{CAPM}^{ETF} + \hat{\beta}_{Mkt}^{ETF}\left[R_{Mkt,t} - R_{f,t}\right] + \epsilon_t^{ETF} \tag{6}$$

The ETFs' CAPM single factor regression-based benchmark is then:

$$\hat{R}_{CAPM\ Benchmark,t}^{ETF} = R_{f,t} + \hat{\beta}_{Mkt}^{ETF}\left[R_{Mkt,t} - R_{f,t}\right] \tag{7}$$

The ETFs' CAPM alpha is then:

$$\hat{\alpha}_{CAPM}^{ETF} = \bar{R}^{ETF} - \bar{R}_{CAPM\ Benchmark}^{ETF} \tag{8}$$

Or equivalently

$$\hat{\alpha}_{CAPM}^{ETF} = \bar{R}^{ETF} - \left[\bar{R}_f + \hat{\beta}_{Mkt}^{ETF}\left[\bar{R}_{Mkt} - \bar{R}_f\right]\right] \tag{9}$$

2.5.3.2 Fama-French 3 Factor Model (FF3)

Fama and French introduced a three-factor model to account for return attributing characteristics such as a firm's size and value (book to market equity) in extension to the market risk factor in CAPM (Fama & French, 1993). The size and value risk factors were included as they were empirically documented in academia to influence stock returns (Banz, 1981; Rosenberg et al., 2021). Fama and French found that higher returns were on average consistent with low book-to-market equity and small capitalization. In addition, lower returns were on average consistent with high book-to-market equity and large capitalization. Fama and French further documented that their three-factor model persistently outperformed CAPM in explaining the cross-section of stock returns (Fama & French, 1998).

In addition to the market portfolio, this approach includes two additional hedged size and value factor portfolios. The size factor portfolio is long small firms and short large firms, denoted (SMB). The value factor portfolio is long value and short growth, that is long high book-to-market equity firms and short low book-to-market equity firms, denoted (HML).

The expected return on an ETF is derived from the following equation:

$$R_t^{ETF} = R_{f,t} + \beta_{Mkt}^{ETF}\left[R_{Mkt,t} - R_{f,t}\right] + \beta_{SMB}^{ETF} R_{SMB,t} + \beta_{HML}^{ETF} R_{HML,t} \tag{10}$$

The factor betas are estimated by running the following time series regression:

$$\begin{aligned} R_t^{ETF} - R_{f,t} = \hat{\alpha}_{FF3}^{ETF} + \hat{\beta}_{Mkt}^{ETF}\left[R_{Mkt,t} - R_{f,t}\right] + \hat{\beta}_{SMB}^{ETF} R_{SMB,t} \\ + \hat{\beta}_{HML}^{ETF} R_{HML,t} + \epsilon_t^{ETF} \end{aligned} \tag{11}$$

The ETFs' FF 3 factor regression-based benchmark is then:

$$\begin{aligned} \hat{R}_{FF3\,Benchmark,t}^{ETF} = R_{f,t} + \hat{\beta}_{Mkt}^{ETF}\left[R_{Mkt,t} - R_{f,t}\right] \\ + \hat{\beta}_{SMB}^{ETF} R_{SMB,t} + \hat{\beta}_{HML}^{ETF} R_{HML,t} \end{aligned} \tag{12}$$

The ETFs' generalized FF 3 factor alpha is then:

$$\hat{\alpha}_{FF3}^{ETF} = \bar{R}^{ETF} - \bar{R}_{FF3\,Benchmark}^{ETF} \tag{13}$$

Or equivalently

$$\hat{\alpha}_{FF3}^{ETF} = \bar{R}^{ETF} - \left[\bar{R}_f + \hat{\beta}_{Mkt}^{ETF}\left[\bar{R}_{Mkt} - \bar{R}_f\right] + \hat{\beta}_{SMB}^{ETF} \bar{R}_{SMB} + \hat{\beta}_{HML}^{ETF} \bar{R}_{HML}\right] \tag{14}$$

2.5.3.3 Fama-French-Carhart 4 Factor Model (FFC4)

Carhart introduced momentum as a persistent risk factor in addition to the FF 3 factor model, from investigating persistence in mutual fund performance (Carhart, 1997). Utilizing the persistence in stocks' performance was documented as a successful strategy to generate significant abnormal returns (Jegadeesh & Titman, 1993).

In addition to the FF3 portfolios, this approach includes an additional hedged momentum factor portfolio. The momentum factor portfolio is long the winners and short the losers of the last six months, denoted (WML).

The expected return on an ETF is derived from the following equation:

$$R_t^{ETF} = R_{f,t} + \beta_{Mkt}^{ETF}\left[R_{Mkt,t} - R_{f,t}\right] + \beta_{SMB}^{ETF} R_{SMB,t} + \beta_{HML}^{ETF} R_{HML,t} + \beta_{WML}^{ETF} R_{WML,t} \tag{15}$$

The factor betas are estimated by running the following time series regression:

$$R_t^{ETF} - R_{f,t} = \hat{\alpha}_{FFC4}^{ETF} + \hat{\beta}_{Mkt}^{ETF}\left[R_{Mkt,t} - R_{f,t}\right] + \hat{\beta}_{SMB}^{ETF} R_{SMB,t} + \hat{\beta}_{HML}^{ETF} R_{HML,t} + \hat{\beta}_{WML}^{ETF} R_{WML,t} + \epsilon_t^{ETF} \tag{16}$$

The ETFs' FFC 4 factor regression-based benchmark is then:

$$\hat{R}_{FFC4\,Benchmark,t}^{ETF} = R_{f,t} + \hat{\beta}_{Mkt}^{ETF}\left[R_{Mkt,t} - R_{f,t}\right] + \hat{\beta}_{SMB}^{ETF} R_{SMB,t} + \hat{\beta}_{HML}^{ETF} R_{HML,t} + \hat{\beta}_{WML}^{ETF} R_{WML,t} \tag{17}$$

The ETFs' generalized FFC 4 factor alpha is then:

$$\hat{\alpha}_{FFC4}^{ETF} = \bar{R}^{ETF} - \bar{R}_{FFC4\,Benchmark}^{ETF} \tag{18}$$

Or equivalently

$$\hat{\alpha}_{FFC4}^{ETF} = \bar{R}^{ETF} - \left[\bar{R}_f + \hat{\beta}_{Mkt}^{ETF}\left[\bar{R}_{Mkt} - \bar{R}_f\right] + \hat{\beta}_{SMB}^{ETF} \bar{R}_{SMB} + \hat{\beta}_{HML}^{ETF} \bar{R}_{HML} + \hat{\beta}_{WML}^{ETF} \bar{R}_{WML}\right] \tag{19}$$

2.5.3.4 Fama-French 5 Factor Model (FFC5)

Fama and French suggested modifying their three-factor model to account for systematic links between returns, firms' profitability, and investment behavior (Fama & French, 2015). Their research documented that this modified model better explains the cross-section of stock returns than the three-factor model.

In addition to the FF 3 portfolios, this approach includes two additional hedged profitability and investment factor portfolios. The profitability factor portfolio is long robust profitability firms and short weak profitability firms, denoted (RMW). The investment factor portfolio is long firms that invest conservatively and short firms that invest aggressively, denoted (CMA).

The expected return on an ETF is derived from the following equation:

$$R_t^{ETF} = R_{f,t} + \beta_{Mkt}^{ETF}\left[R_{Mkt,t} - R_{f,t}\right] + \beta_{SMB}^{ETF} R_{SMB,t} + \beta_{HML}^{ETF} R_{HML,t} + \beta_{RMW}^{ETF} R_{RMW,t} + \beta_{CMA}^{ETF} R_{CMA,t} \quad (20)$$

The factor betas are estimated by running the following time series regression:

$$R_t^{ETF} - R_{f,t} = \hat{\alpha}_{FF5}^{ETF} + \hat{\beta}_{Mkt}^{ETF}\left[R_{Mkt,t} - R_{f,t}\right] + \hat{\beta}_{SMB}^{ETF} R_{SMB,t} + \hat{\beta}_{HML}^{ETF} R_{HML,t} + \hat{\beta}_{RMW}^{ETF} R_{RMW,t} + \hat{\beta}_{CMA}^{ETF} R_{CMA,t} + \epsilon_t^{ETF} \quad (21)$$

The ETFs' FF 5 factor regression-based benchmark is then:

$$\hat{R}_{FF5\ Benchmark,t}^{ETF} = R_{f,t} + \hat{\beta}_{Mkt}^{ETF}\left[R_{Mkt,t} - R_{f,t}\right] + \hat{\beta}_{SMB}^{ETF} R_{SMB,t} + \hat{\beta}_{HML}^{ETF} R_{HML,t} + \hat{\beta}_{RMW}^{ETF} R_{RMW,t} + \hat{\beta}_{CMA}^{ETF} R_{CMA,t} \quad (22)$$

The ETFs' generalized FF 5 factor alpha is then:

$$\hat{\alpha}_{FF5}^{ETF} = \bar{R}^{ETF} - \bar{R}_{FF5\ Benchmark}^{ETF} \quad (23)$$

Or equivalently

$$\hat{\alpha}_{FF5}^{ETF} = \bar{R}^{ETF} - \left[\bar{R}_f + \hat{\beta}_{Mkt}^{ETF}\left[\bar{R}_{Mkt} - \bar{R}_f\right] + \hat{\beta}_{SMB}^{ETF} \bar{R}_{SMB} + \hat{\beta}_{HML}^{ETF} \bar{R}_{HML} + \hat{\beta}_{RMW}^{ETF} R_{RMW} + \hat{\beta}_{CMA}^{ETF} \bar{R}_{CMA}\right] \quad (24)$$

2.5.3.5 Hou-Xue-Zhang 4 Factor Model (HXZ4)

Hou et al., (2015) suggest accounting for investment behavior and profitability with an investment and a return on equity factor, denoted (IA) and (ROE). Moreover, they propose to replace the value and momentum factor of the FFC four-factor model with these factors.

The expected return on an ETF is derived from the following equation:

$$R_t^{ETF} = R_{f,t} + \beta_{Mkt}^{ETF}\left[R_{Mkt,t} - R_{f,t}\right] + \beta_{SMB}^{ETF} R_{SMB,t} + \beta_{IA}^{ETF} R_{INV,t} + \beta_{ROE}^{ETF} R_{ROE,t} \tag{25}$$

The factor betas are estimated by running the following time series regression:

$$R_t^{ETF} - R_{f,t} = \hat{\alpha}_{HXZ4}^{ETF} + \hat{\beta}_{Mkt}^{ETF}\left[R_{Mkt,t} - R_{f,t}\right] + \hat{\beta}_{SMB}^{ETF} R_{SMB,t} + \hat{\beta}_{INV}^{ETF} R_{IA,t} + \hat{\beta}_{ROE}^{ETF} R_{ROE,t} + \epsilon_t^{ETF} \tag{26}$$

The ETFs' HXZ 4 factor regression-based benchmark is then:

$$\hat{R}_{HXZ4\,Benchmark,t}^{ETF} = R_{f,t} + \hat{\beta}_{Mkt}^{ETF}\left[R_{Mkt,t} - R_{f,t}\right] + \hat{\beta}_{SMB}^{ETF} R_{SMB,t} + \hat{\beta}_{INV}^{ETF} R_{IA,t} + \hat{\beta}_{ROE}^{ETF} R_{ROE,t} \tag{27}$$

The ETFs' generalized HXZ 4 factor alpha is then:

$$\hat{\alpha}_{HXZ4}^{ETF} = \bar{R}^{ETF} - \bar{R}_{HXZ4\,Benchmark}^{ETF} \tag{28}$$

Or equivalently

$$\hat{\alpha}_{HXZ4}^{ETF} = \bar{R}^{ETF} - \left[\bar{R}_f + \hat{\beta}_{Mkt}^{ETF}\left[\bar{R}_{Mkt} - \bar{R}_f\right] + \hat{\beta}_{SMB}^{ETF} \bar{R}_{SMB} + \hat{\beta}_{IA}^{ETF} \bar{R}_{IA} + \hat{\beta}_{ROE}^{ETF} \bar{R}_{ROE}\right] \tag{29}$$

2.5.3.6 Cremers-Petajisto-Zitzewitz 4 Factor Model (CPZ4)

Cremers et al., (2013) documented that passive benchmark portfolios generally generated negative alphas when evaluated using the other factor models. The reasoning behind their findings is that the portfolios in the other factor models are not truly traded portfolios, and being passive benchmark portfolios should theoretically generate zero alpha. Accordingly, their research suggests replacing the FFC four-factor model with traded passive investable benchmarks.

This approach includes the S&P 500 index as a market portfolio, the Russell 2000 index in excess of the S&P 500 as a hedged size portfolio, the Russell 3000 Value index in excess of the Russell 3000 Growth index as a hedged value portfolio, together with the hedged momentum factor portfolio.

The expected return on an ETF is derived from the following equation:

$$\begin{aligned} R_t^{ETF} = R_{f,t} &+ \beta_{BIG}^{ETF}\left[R_{S\&P\,500,t} - R_{f,t}\right] \\ &+ \beta_{SMALL}^{ETF}\left[R_{Russell\ 2000,t} - R_{S\&P\,500,t}\right] \\ &+ \beta_{VMG}^{ETF}\left[R_{Russell\ 3000\ Value,t} - R_{Russell\ 3000\ Growth,t}\right] \\ &+ \beta_{WML}^{ETF} R_{WML,t} \end{aligned} \tag{30}$$

The factor betas are estimated by running the following time series regression:

$$\begin{aligned} R_t^{ETF} - R_{f,t} = \hat{\alpha}_{CPZ4}^{ETF} &+ \hat{\beta}_{BIG}^{ETF}\left[R_{S\&P\,500,t} - R_{f,t}\right] \\ &+ \hat{\beta}_{SMALL}^{ETF}\left[R_{Russell\ 2000,t} - R_{S\&P\,500,t}\right] \\ &+ \hat{\beta}_{VMG}^{ETF}\left[R_{Russell\ 3000\ Value,t} - R_{Russell\ 3000\ Growth,t}\right] \\ &+ \hat{\beta}_{WML}^{ETF} R_{WML,t} + \epsilon_t^{ETF} \end{aligned} \tag{31}$$

The ETFs' CPZ 4 factor regression-based benchmark is then:

$$\begin{aligned} \hat{R}_{CPZ4\,Benchmark,t}^{ETF} & \\ = R_{f,t} &+ \hat{\beta}_{BIG}^{ETF}\left[R_{S\&P\,500,t} - R_{f,t}\right] \\ &+ \hat{\beta}_{SMALL}^{ETF}\left[R_{Russell\ 2000,t} - R_{S\&P\,500,t}\right] \\ &+ \hat{\beta}_{VMG}^{ETF}\left[R_{Russell\ 3000\ Value,t} - R_{Russell\ 3000\ Growth,t}\right] \\ &+ \hat{\beta}_{WML}^{ETF} R_{WML,t} \end{aligned} \tag{32}$$

The ETFs' generalized CPZ 4 factor alpha is then:

$$\hat{\alpha}_{CPZ4}^{ETF} = \bar{R}^{ETF} - \bar{R}_{CPZ4\,Benchmark}^{ETF} \tag{33}$$

Or equivalently

$$\begin{aligned} \hat{\alpha}_{CPZ4}^{ETF} = \bar{R}^{ETF} - \Big[\bar{R}_f &+ \hat{\beta}_{BIG}^{ETF}\left[\bar{R}_{S\&P\,500} - \bar{R}_f\right] \\ &+ \hat{\beta}_{SMALL}^{ETF}\left[\bar{R}_{Russell\ 2000} - \bar{R}_{S\&P\,500}\right] \\ &+ \hat{\beta}_{VMG}^{ETF}\left[\bar{R}_{Russell\ 3000\ Value} - \bar{R}_{Russell\ 3000\ Growth}\right] \\ &+ \hat{\beta}_{WML}^{ETF} \bar{R}_{WML}\Big] \end{aligned} \tag{34}$$

3. Hypotheses and Methodology

In this thesis, we focus on the performance of factor ETFs relative to given benchmark portfolios and examine if they add value to different types of investors through different metrics. We follow a similar strategy to Mateus et al., (2020) but differentiate ourselves in three fundamental ways.

The first difference is the sample period and sample size. We use a shorter sample period and larger sample size. For example, Mateus et al., (2020) evaluated 152 US-based funds on the 3x3 Morningstar Box Style classification. Thus, they evaluated the value and growth factors sorted on capitalization size. We increase our sample size to include more factor strategies implemented with ETFs across developed countries. Accordingly, we examine and sort the factor ETFs into eight peer categories by their style characteristics. Including size, value, growth, quality, low volatility, dividend, momentum, and blended/multifactor funds. Because some of the factor strategies have been implemented recently with ETFs, there is a shorter time series. Therefore, the short time series is a weakness in the regression-based performance evaluation.

The second difference in methodology is the choice of performance metrics used to evaluate whether the funds add value to different investors. Mateus et al., (2020) use the Sharpe ratio, Information ratio, and Treynor ratio. We add the M2 and Appraisal ratio to include more performance evaluating metrics. The M^2 does not affect the rankings based on the Sharpe ratio. However, we include it because it is more intuitively interpreted than the Sharpe ratio (Simons, 1998). Moreover, we add the Appraisal ratio because it is often confused with the information ratio but allows us to rank funds differently on the reward to additional risk.

The third essential difference is the choice of model and benchmark portfolios in the regression-based performance analysis. Mateus et al., (2020) use the Fama-French-Carhart 4 factor model to evaluate the performances. Furthermore, they construct and apply active peer benchmarks (APB) to adjust their alphas with the approach of Hunter et al., (2014). As a result, their methodology is unbiased of non-zero benchmark alphas and shows performance persistence a year ahead. In comparison, we do not use active peer benchmarks but instead apply passive benchmarks by including more models to get more regression-based performance measures, such as Kothari & Warner, (2001). Accordingly, we apply all the models described in section (*2.5.3*). Moreover, to evaluate performance persistence, we compare the performance metrics before and during Covid-19.

In short, we have evaluated 334 factor ETFs in the sample period of Jan 2017 to Dec 2021. We have chosen a value-weighted market portfolio, multiple factor portfolios, and truly traded portfolios as benchmarks for the regression-based performance measures. In addition, we utilize the value-weighted market portfolio as a benchmark for the other performance measures.

The review of relevant theories and literature has led to the following hypotheses and methodologies regarding the research question.

3.1 Do Factor ETFs Generate Alpha?

The first hypothesis we will test is whether factor ETFs generate alpha. Alpha is a performance metric that reflects the abnormal rate of return on an asset in excess of the predicted return by an asset pricing model in equilibrium. Thus, factor ETFs will add significant value to the average investor if the alpha is positive. Correspondingly, we test the following hypotheses:

$$H_0: \alpha_{ETF} \leq 0 \quad (35)$$

$$H_A: \alpha_{ETF} > 0 \quad (36)$$

The alpha metric is also known as Jensen's measure, introduced by Jensen, (1968) to evaluate mutual fund managers. Jensen, (1968) used the CAPM to derive the alpha as we did in equation (8). However, we believe CAPM will likely overestimate the alphas of factor ETFs because it only adjusts the performance for one systematic risk factor. Certainly, factor funds are likely to have exposure to multiple systematic risk factors. Thus, we will utilize the same methodology as Jensen, (1968), but include multifactor models to correct for possible other sources of systematic risk or so-called omitted variable biases.

The alphas are estimated along with the betas through a set of ordinary least squares (OLS) time-series regressions, as in equations (6), (11), (16), (21), (26), and (31). OLS is a regression approach that explains the linear relationship between a set of dependent and independent variables. The dependent variable in our regressions is the excess returns of factor ETFs. The independent variables are the risk factor portfolios that explain average stock return well in the asset pricing model's market equilibrium. Accordingly, the OLS method estimates the relationship by minimizing the sums of squares in the difference between observed and predicted excess returns. As a result, we obtain the alphas from equations (8), (13), (18), (23), (28), and (33).

As a result, we have applied a methodology that examines the risk-adjusted returns for every factor ETF. However, the applied procedures might cause some implications and limitations to the results. First, the OLS approach assumes that the independent variables are independent and not correlated with each other. Accordingly, there might be a presence of multicollinearity that causes the standard errors of regression estimates to increase. Consequently, the test statistics might result in statistically insignificant estimates at conventional levels due to thinner tails in the normal distribution. Thus, the conclusions might become inappropriate. However, we use OLS because it is the most common and standard approach to estimating linear regression models (Brooks, 2019).

In addition, a lack of factors included in the regression models might cause an inappropriate presence of positively estimated alphas. Therefore, we include a magnitude of asset pricing models that accounts for several sources of systematic risk. Moreover, we examine asset returns in the time-series analyses, which are stationary without unit-roots (Brooks, 2019). Hence, there should not be a presence of spurious regressions. For additional considerations, we refer to section (*6*).

The multifactor models we use are developed to explain the cross-section of stock returns better than the CAPM. Accordingly, we also study the cross-section of factor ETFs using Brooks' methodology (Brooks, 2019, p. 577). This approach is similar to Fama-MacBeth, (1973). However, (Brooks, 2019) provides a methodology that examines the link between factor characteristics and abnormal returns. In comparison, the methodology of Fama-MacBeth, (1973) is a two-step procedure applied to test asset pricing models and explain the relationship between expected returns and risks. We will use cross-sectional regressions to research the link between factor risks and excess returns of factor ETFs.

In the first step, we obtain the estimated betas from the time-series regressions in equations (6), (11), (16), (21), (26), and (31). In the second step, we perform cross-sectional regressions that examine the explanatory power of the estimated betas from the first step. Moreover, we examine the cross-sectional intercepts to test whether the average excess return is significantly different from zero after allowing for the impacts of exposures to different sources of systematic risks.

Thus, we obtain the following cross-sectional regressions from time t = 1,…,T:

$$R_{i,t} - R_{f,t} = \hat{\lambda}_{0,t} + \sum_{j=1}^{J} \hat{\lambda}_{j,t}\, \hat{\beta}_{i,j} \tag{37}$$

where

$\hat{\lambda}_{0,t}$ is the estimated cross-sectional intercept at time *t*.

$\hat{\beta}_{i,j}$ is the estimated betas from the first step that represents the factor loadings of fund *i* to the j^{th} factor from the different asset pricing models.

$\hat{\lambda}_{j,t}$ is the estimated risk premium for the j^{th} factor from the different asset pricing models at time *t*.

Accordingly, we obtain T cross-sectional estimates for the intercept and every factor. Consequently, the average estimates are the following:

$$\overline{\hat{\lambda}_0} = \frac{1}{T}\sum_{t=1}^{T} \hat{\lambda}_{0,t} \quad and \quad \hat{\lambda}_j = \frac{1}{T}\sum_{t=1}^{T} \hat{\lambda}_{j,t} \quad for\ j = 1, \dots, J \tag{38}$$

3.2 Do Factor ETFs Add Value to Different Investors?

In this section, we introduce appropriate risk-adjusting measures for different types of investors. Investors care about different risks, depending on their current portfolio construction and risk tolerance. Hence, we suggest different measures to different investors. Overall, we consider three different types of investors with different risk-return trade-offs when evaluating a new investment opportunity. These investors require respectively different compensation and reward for exposure to total risk, additional risk, and systematic risk.

Type 1:

The first type of investor considers investing their entire risky portfolio into a single factor ETF instead of the market portfolio. Thus, we recommend the Sharpe ratio and M^2 measure because these investors care about the compensation for total risk exposure.

Type 2:

The second type of investor considers adding factor ETFs to a well-diversified market portfolio. Hence, we recommend the Information and Appraisal ratios because these investors care about extra compensation for the additional risk exposure.

Type 3:

The third type of investor considers diversifying across the factor ETFs to obtain a well-diversified portfolio. Thus, we recommend the Treynor's ratio because these investors care about the compensation for systematic risk exposure.

These metrics are helpful for our thesis as they can be used to determine the risk-adjusted return, e.g., help determine which investors should consider factor ETFs as a potentially valuable investment. We have chosen to include more than one metric because the different metrics provide different insights, accounting for different types of risk.

3.2.1 Sharpe Ratio (SR)

The Sharpe ratio is the most commonly used risk-adjusted return measure, introduced by Sharpe, (1966). The metric is an asset's excess return per unit of its total risk. Excess returns are calculated as the difference between the arithmetic average return of a portfolio and the risk-free rate, and the volatility of excess returns captures the absolute risk. Correspondingly, the metric is derived from dividing the excess returns by the standard deviation of the excess returns:

$$SR_{ETF} = \frac{\bar{R}_{ETF} - \bar{R}_f}{\sigma_{(R_{ETF} - R_f)}} \tag{39}$$

The beauty of the Sharpe ratio is that it reflects a trade-off between expected return and risk independently of a market benchmark. A low Sharpe ratio would imply that the ETF delivers a low return for its level of volatility and vice versa. Typically, a good Sharpe ratio is above one, where the ETF would at least pay off some excess returns relative to its risk. Moreover, the metric compares risk-adjusted performance for ETFs regardless of their volatilities and correlations with a market benchmark.

The Sharpe ratio might be limited due to its assumptions, such as the assumption that returns are normally distributed. However, skewed distributions are likely to occur in the real world. Moreover, the metric assumes that the probability of upside and downside risks are equally distributed. In addition, the metric could be manipulated by adjusting the interval of observations or excluding volatile periods with low excess returns.

Correspondingly, we test the following hypotheses:

$$H_0: SR_{ETF} \leq SR_{Market\ Benchmark} \tag{40}$$

$$H_A: SR_{ETF} > SR_{Market\ Benchmark} \tag{41}$$

3.2.2 M²

M^2 is the Modigliani-Modigliani measure, introduced by Modigliani-Modigliani, (1997). The metric is also known as the risk-adjusted performance (RAP) and the Graham-Harvey (GH) measures. M^2 measures the risk-adjusted performance relative to a benchmark portfolio in contrast to the Sharpe ratio. However, the M^2 is derived from the Sharpe ratio and is expressed as follows:

$$\begin{aligned} M_{ETF}^2 &= [SR_{ETF} - SR_{Mkt}] \times \sigma_{(R_{Mkt}-R_f)} \\ &= [\bar{R}_{ETF} - \bar{R}_f] \times \frac{\sigma_{(R_{Mkt}-R_f)}}{\sigma_{(R_{ETF}-R_f)}} - [\bar{R}_{Mkt} - \bar{R}_f] \end{aligned} \tag{42}$$

Thus, the metric provides the same ETF rankings as the Sharpe ratio but is stated in percentage terms. The intuition is that investors can construct portfolios with ETFs and riskless assets to obtain a greater return for the same volatility as the market benchmark. Correspondingly, the metric is the excess return of such a portfolio over the market portfolio. However, M^2 is an approach based on the assumption that investors can lend and borrow at a risk-free rate.

Correspondingly, we test the following hypotheses:

$$H_0: \mathrm{M}^2_{ETF} \leq 0 \tag{43}$$

$$H_A: \mathrm{M}^2_{ETF} > 0 \tag{44}$$

3.2.3 Information Ratio (IR)

The Information ratio, introduced by Treynor and Black, (1973) is one of the most prominent performance measures (Grinold, 1989). The ratio measures the excess return of a portfolio relative to a market benchmark. Moreover, it adjusts this excess return for the consistency of such performance. The correspondingly tracking error measures this consistency of excess returns.

$$IR_{ETF} = \frac{\bar{R}_{ETF} - \bar{R}_{Mkt}}{\sigma_{(R_{ETF} - R_{Mkt})}} \tag{45}$$

Accordingly, the Information ratio is the ETFs' average alpha divided by the standard deviation of diversifiable risk.

Correspondingly, we test the following hypotheses:

$$H_0: \mathrm{IR_{ETF}} \leq 0 \tag{46}$$

$$H_A: \mathrm{IR_{ETF}} > 0 \tag{47}$$

3.2.4 Appraisal Ratio (AR)

The Appraisal ratio is similar to the Information ratio, and the two are often confused. However, the Appraisal ratio measures the simulated CAPM alphas from equation (8) divided by the residual's standard deviation. Like the tracking error, the residual's standard deviation captures systematic risks.

$$AR_{ETF} = \frac{\alpha_{ETF}}{\sigma_{\epsilon_{ETF}}} \tag{48}$$

where

$$\sigma_{\epsilon_{ETF}} = \sqrt{\sigma_{ETF}^2 - \beta_{ETF}^2 \times \sigma_{Mkt}^2} \tag{49}$$

The appraisal ratio is a good measure of performance consistency and reflects the ETFs' active return per unit of risk from factor investing. The ratio can be generalized to other alphas estimated from multifactor regressions, but one must adjust for the obtained residual's standard deviations in each model. However, we will not generalize the ratio for all models. Accordingly, we utilize the CAPM to compute the appraisal ratio.

Correspondingly, we test the following hypotheses:

$$H_0\text{: }\mathrm{AR_{ETF}} \leq 0 \tag{50}$$

$$H_A\text{: }\mathrm{AR_{ETF}} > 0 \tag{51}$$

3.2.5 Treynor Ratio (TR)

The Treynor Ratio, introduced by Treynor, (1966) is similar to the Sharpe ratio. However, it adjusts performance for systematic risk measured by beta and not absolute risk measured by volatility. Hence, Treynor's ratio divides excess returns by the CAPM beta we derived in equation (2).

$$TR_{ETF} = \frac{\bar{R}_{ETF} - \bar{R}_f}{\beta_{ETF}} \tag{52}$$

Consequently, Treynor's ratio captures how much investors are compensated for taking on systematic risk. Thus, the choice of an appropriate market benchmark is essential to estimate fair betas for all ETFs. Moreover, Traynor's ratio is the slope of the security market line (SML), and indicate under- and overvalued funds.

Correspondingly, we test the following hypotheses:

$$H_0\text{: }TR_{ETF} \leq TR_{Market\ Benchmark} \tag{53}$$

$$H_A\text{: }TR_{ETF} > TR_{Market\ Benchmark} \tag{54}$$

3.3 Markowitz Mean-Variance Optimization

Markowitz, (1952) introduced the foundation of modern portfolio theory (MPT) in a study of portfolio selection. MPT proposes that investors should only care about means and variances of returns. These underlying principles are applied to choose optimal portfolios based on risk-return trade-offs and efficient diversification (Bodie et al., 2020). Markowitz found that diversification generated higher expected returns with less risk than single assets. As a result, the diversified portfolios became known as mean-variance efficient portfolios.

In this context, we examine how the average investor can exploit the diversification benefits across style characteristics of factor ETFs together with the market benchmark. Hence, we construct equally weighted portfolios sorted on style characteristics and study what portfolio allocation that maximizes the risk-return trade-off measured by the Sharpe ratio.

As a result, we identify style characteristics of factor ETFs that are valuable on average in terms of diversification to the average investor. Moreover, we identify shorting candidates to leverage the positions that maximize the risk-return trade-off. Accordingly, we implicitly assume that the average investor holds the market portfolio and does not have any borrowing or short-selling constraints. Moreover, we analyze whether the inclusion of factor ETFs in the average investor's portfolio is considered to improve the risk-adjusted expected returns.

The mean-variance optimization indicates the optimal capital allocation line (CAL). Thus, we identify the best allocation of capital between different factor strategies of factor ETFs together with the market benchmark. The factor strategies and style characteristics are sorted on size, value, growth, quality, volatility, dividend, momentum, and blended multifactor funds. Correspondingly, we utilize modern portfolio theory to examine what type of factor ETFs provide diversification benefits and can be considered valuable to the average investor. Moreover, this analysis provide a framework to how investors can exploit the style characteristics of factor ETFs.

4. Data

This section describes the data used in this thesis to research whether factor ETFs are a valuable investment. First, we present the screening process and data collection of factor ETFs. Second, we present the selection of benchmark portfolios for performance evaluation. Lastly, we present how the benchmark portfolios are constructed.

4.1 Selection of Factor ETFs – The Screening Process

We first conducted a positive and negative screening process in Bloomberg to collect data from factor ETFs of interest. Then we conducted the data collection of monthly holding period returns in The Center for Research in Security Prices (CRSP) and from Refinitiv. Due to the retrospective data collection, there is a presence of survivorship bias. For a more detailed discussion of survivorship bias, we refer to section (*6.1*).

The positive screening process of our data sample first took place in the Bloomberg Terminal and included ETFs that were tracking a basket of assets with return attributing characteristics. Specifically, we screened for volatility, beta, momentum, quality, growth, dividend, size, and value characteristics. In addition, we screened for ETFs with "factor" and "multifactor" stated in their names and funds with descriptions that claimed to be factor ETFs. Thus, we collected data from ETFs with portfolios concentrated on pure factors and portfolios diversified across multiple factors.

The second part of the positive screening process took place in CRSP and Refinitiv. We got the necessary ETF tickers from Bloomberg, which was the input in CRSP and Refinitiv. We used both CRSP and Refinitiv to collect time series data because CRSP only provides data for US-listed data. Thus, for non-US data, we had to use Refinitiv.

The negative screening process of our data sample excludes ETFs that do not provide data in our sample period. Hence, we have excluded factor ETFs with a start-up date later than Jan 1st, 2017. We have also excluded factor ETFs that closed down for any reason before Dec 31st, 2021. In addition, we excluded levered or inverse ETFs.

To evaluate factor ETFs' performances, we have collected historical monthly returns of ETFs that provide at least five years of historical returns. The sample period is five years to obtain at least 12 observations for each independent variable in the factor models. Thus, our dataset has 60 observations of monthly returns for 334 factor ETFs. Accordingly, we follow the rule of thumb of at least 30 observations for each fund, a minimum of 10 observations per variable in the factor models, and a minimum of 100 funds in the sample. Thus, we do not consider it necessary to change from monthly to weekly or daily observations.

Table 1: Criteria we have restricted the dataset of factor ETFs to meet.

Criteria No.:	Criteria Description	Source
1	The ETFs follow a prespecified rule of factor style/strategy when investing.	Bloomberg
2	The ETFs have a start-up date before Jan 1st, 2017 and are still running as of Dec 31st, 2021.	Bloomberg
3a	The ETFs' historical monthly returns are available in CRSP every month in the sample period (US data).	CRSP
3b	The ETFs' historical monthly returns are available in Refinitiv every month in the sample period (non-US data).	Refinitiv

4.2 Selection of Benchmark Portfolios

An appropriate benchmark is essential to evaluate the performance of any asset. Before launching a fund, the management team and prominent investors typically agree on a benchmark. This benchmark is known as the self-reported benchmark. We have identified the self-reported benchmarks in the prospectus of the ETFs.

However, the main problem is whether the self-reported benchmarks appropriately reflect the risk profile of the funds. There are typically two ways to identify appropriate benchmarks. First, one could use factor models to generate regression-based benchmarks for every factor ETF concerning their systematic exposure to risk factors. Second, one could use style analysis as Sharpe, (1992) to identify a closely correlated benchmark with the ETFs. We have done the former and not the latter because (Fama & French, 1993) characterize this application of multifactor benchmarks as straightforward and simple. Moreover, they claim these regression-based passive benchmarks are well specified in the cross-section.

We use several benchmarks to avoid whether a specific benchmark reflects the risk profile of the funds appropriately. Including, a proxy for the value-weighted market portfolio, multiple different factor portfolios, and some truly traded portfolios in our regression-based performance evaluation. We assume that the average investor holds the value-weighted market portfolio. Correspondingly, we rank and compare the other performance measures of the factor ETFs to the market portfolio proxy. Accordingly, we utilize a broad range of benchmark portfolios.

From the positive and negative screening process, we have collected 334 ETFs in our data sample. Accordingly, we examined the portfolios of the ETFs and found that many funds diversified across assets in the US, European, and Asian markets. Thus, we have selected proxy portfolios from developed countries provided by Kenneth French's library.

The market portfolio consists of value-weighted portfolios from Australia, Austria, Belgium, Canada, Switzerland, Germany, Denmark, Spain, Finland, France, Great Britain, Greece, Hong Kong, Ireland, Italy, Japan, Netherlands, Norway, New Zealand, Portugal, Sweden, Singapore, and United States.

In addition, we also collected factor portfolios from Hou-Xue-Zhang's q-factors data library. However, the benchmark portfolios are indeed not truly traded portfolios. Therefore, we substitute them in parts of our regression-based analysis with easily traded US portfolios following (Cremers et al., (2013).

Table 2: Sources for dataset of market benchmark and factor portfolios used to construct regression-based benchmarks.

No.:	Description	Source
1	The Risk-Free Rate	Kenneth French
2	Value-weighted Market Portfolio	Kenneth French
3	FF 3-Factor Portfolios	Kenneth French
4	FFC 4-Factor Portfolios	Kenneth French
5	FF 5-Factor Portfolios	Kenneth French
6	HXZ 4-Factor Portfolios	Kenneth French & Hou-Xue-Zhang
7	CPZ 4-Factor Portfolios	Kenneth French & Compustat

The data for the asset pricing model's portfolios are provided to us by Kenneth French, Hou-Xue-Zhang, and Compustat.

First, we have collected the market (MKT), size (SMB), value (HML), momentum (WML), operating profitability (RMW), and Investment (CMA) factor portfolios from the Kenneth French Data Library (*Kenneth R. French - Data Library*, 2022). These data are available and provided for developed markets.

Second, we have collected the investment to assets (IA) and return on equity (ROE) factor portfolios from the Hou-Xue-Zhang Data Library (*Hou-Xue-Zhang q-Factors Data Library*, 2022). In comparison, these data are available and provided for the US market.

Third, we have collected the S&P 500 Index, Russell 2000 Index, Russell 3000 Value Index, and Russell 3000 Growth Index from the CRSP database. These data are, of course, representative of the US market.

The benchmark and factor portfolios used as proxies in the factor models are not perfect explanatory variables. However, we find these portfolios the most appropriate as they roughly match the markets of the ETFs' asset allocations. In addition, these portfolios have been used extensively in academia to estimate whether assets and funds generate alpha.

Table 3: Correlation matrix of factor portfolios.

	Mkt	*SMB*	*HML*	*RMW*	*CMA*	*WML*	*IA*	*ROE*	*SP500*	*(R2-SP500)*	*(R3V-R3G)*
Mkt	1										
SMB	0.24	1									
HML	0.15	0.02	1								
RMW	0.04	-0.35	-0.49	1							
CMA	-0.10	-0.16	0.77	-0.27	1						
WML	-0.40	-0.07	-0.64	0.25	-0.37	1					
IA	-0.13	-0.06	0.57	-0.29	0.83	-0.27	1				
ROE	-0.40	-0.54	-0.25	0.66	0.08	0.44	-0.02	1			
SP500	0.97	0.13	0.07	0.16	-0.15	-0.34	-0.16	-0.29	1		
(R2-SP500)	0.37	0.75	0.34	-0.59	0.01	-0.42	0.09	-0.72	0.27	1	
(R3V-R3G)	0.06	0.09	0.88	-0.42	0.75	-0.61	0.58	-0.22	-0.01	0.34	1

The factor portfolios utilized in the different asset pricing models are uncorrelated. Thus, there is likely no multicollinearity among the factor portfolios. Correspondingly, the slight correlation between factors implies that the choice of benchmark portfolios captures different sources of systematic risk. Moreover, the factors that substitute each other across models are highly correlated. Hence, these factors capture much of the same risk. Furthermore, Cremers et al., (2013) provide great substitutes for the constructed factor portfolios with truly traded portfolios.

4.3 Construction of Factor Portfolios

This section presents the construction of factor portfolios that we utilize in the regression-based benchmarks. All returns are denominated in USD, include dividends and capital gains, and are not continuously compounded.

4.3.1 The Market Factor

The market factor portfolio provided by Fama and French, $R_{Mkt} - R_f$, is constructed as a value-weighted portfolio of monthly returns in regions considered developed markets, in excess of the risk-free rate. We use the US one-month T-bill rate, as Fama and French do, to proxy the risk-free rate.

4.3.2 SMB and HML

The size factor portfolio provided by Fama and French is a hedged portfolio with long positions in small-capitalization stocks and short positions in big-capitalization stocks. The size factor returns are derived from the following portfolio equation:

$$SMB = \frac{1}{3}(Small\ Value + Small\ Neutral + Small\ Growth) - \frac{1}{3}(Big\ Value + Big\ Neutral + Big\ Growth) \tag{55}$$

Accordingly, the size factor portfolio is the equal weight average of the returns on three small-cap portfolios minus the average for the three large-cap portfolios.

The value factor portfolio provided by Fama and French is a hedged portfolio with long positions in high book to market equity ratio stocks and short positions in low book to market equity ratio stocks. The value factor returns are derived from the following portfolio equation:

$$HML = \frac{1}{2}(Small\ Value + Big\ Value) - \frac{1}{2}(Small\ Growth + Big\ Growth) \tag{56}$$

Accordingly, the value factor portfolio is the equal weight average of the returns on two value portfolios minus the average for the two growth portfolios.

4.3.3 WML

The momentum factor portfolio provided by Fama and French is a hedged portfolio with long positions in high positive momentum stocks and short positions in high negative momentum stocks. The momentum factor returns are derived from the following portfolio equation:

$$WML = \frac{1}{2}(Small\ High + Big\ High) - \frac{1}{2}(Small\ Low + Big\ Low) \tag{57}$$

Accordingly, the momentum factor portfolio is the equal weight average of the returns on two winner portfolios minus the average for the two loser portfolios.

4.3.4 RMW and CMA

The operating profitability factor portfolio provided by Fama and French is a hedged portfolio with long positions in robust operating profitability stocks and short positions in weak operating profitability stocks. The operating profitability factor returns are derived from the following portfolio equation:

$$RMW = \frac{1}{2}(Small\ Robust + Big\ Robust) - \frac{1}{2}(Small\ Weak + Big\ Weak) \tag{58}$$

Accordingly, the operating profitability factor portfolio is the equal weight average of the returns on two robust operating profitability portfolios minus the average for the two weak operating profitability portfolios.

The investment factor portfolio provided by Fama and French is a hedged portfolio with long positions in conservative investment stocks and short positions in aggressive investment stocks. The investment factor returns are derived from the following portfolio equation:

$$CMA = \frac{1}{2}(Small\ Conservative + Big\ Conservative) - \frac{1}{2}(Small\ Aggressive + Big\ Aggressive) \tag{59}$$

Accordingly, the investment factor portfolio is the equal weight average of the returns on two conservative investment portfolios minus the average for the two aggressive investment portfolios.

4.3.5 IA and ROE

The investment factor portfolio provided by Hou-Xue-Zhang is a hedged portfolio derived from the following portfolio equation:

$$IA = \frac{1}{6}\left(\sum_{i=1}^{2}\sum_{k=1}^{3} i, j = 1, k\right) - \frac{1}{6}\left(\sum_{i=1}^{2}\sum_{k=1}^{3} i, j = 3, k\right) \tag{60}$$

Where, i = 1, 2, j = 1, 2, 3, and k = 1, 2, 3. The investment factor portfolio is then constructed from the monthly returns of the value-weighted portfolio containing all the firms in the i^{th} size group, the j^{th} investment-to-asset group, and the k^{th} return on equity group.

Accordingly, Hou-Xue-Zhang constructs the investment factor portfolio by the equal weight average of the returns from six portfolios sorted in the lowest investment to asset group minus the average of the returns for the six portfolios sorted in the highest investment-to-asset group (Hou et al., 2022).

The return on equity factor portfolio provided by Hou-Xue-Zhang is a hedged portfolio derived from the following portfolio equation:

$$ROE = \frac{1}{6}\left(\sum_{i=1}^{2}\sum_{j=1}^{3} i, j, k = 3\right) - \frac{1}{6}\left(\sum_{i=1}^{2}\sum_{j=1}^{3} i, j, k = 1\right) \tag{61}$$

Where, i = 1, 2, j = 1, 2, 3, and k = 1, 2, 3. The ROE factor portfolio is then constructed from the monthly returns of the value-weighted portfolio containing all the firms in the i^{th} size group, the j^{th} investment-to-asset group, and the k^{th} return on equity group.

Accordingly, Hou-Xue-Zhang constructs the return on equity factor portfolio by the equal weight average of the returns from six portfolios sorted in the highest return on equity group minus the average of the returns for the six portfolios sorted in the lowest return on equity group (Hou et al., 2022).

4.3.6 The Real-World Substitutes

Cremers et al. (2013) substitutes the value-weighted market factor portfolio provided by Fama and French with the S&P 500 in excess of the risk-free rate, $R_{S\&P\ 500\ Index} - R_f$. The S&P 500 Index is constructed as a portfolio of the 500 biggest companies in the US. Again, we use the US one month T-bill rate as a proxy for the risk-free rate. Thus, it captures the systematic risk factor exposure to large-cap stocks.

Furthermore, Cremers et al. (2013) construct a hedged portfolio on the size factor with a long position in the Russell 2000 Index and a short position in the S&P 500 Index, $R_{Russell\ 2000\ Index} - R_{S\&P\ 500\ Index}$. The Russell 2000 portfolio is constructed as a portfolio of the 2000 smallest companies included in the Russell 3000 Index. Hence, the portfolio reflects the traditional size (SMB) factor portfolio constructed by Fama and French. Accordingly, it captures the systematic risk exposure to small-cap stocks.

Cremers et al. (2013) also substitute the traditional value factor portfolio (HML) constructed by Fama and French with a hedged portfolio consisting of a long position in the Russell 3000 Value Index and a short position in the Russell 3000 Growth Index, $R_{Russell\ 3000\ Value\ Index} - R_{Russell\ 3000\ Growth\ Index}$.

The Russell 3000 Value Index includes stocks with lower price-to-book equity ratios and expected growth rates. In contrast, the Russell 3000 Growth Index includes stocks with growth rates above average. The selection of assets comes from the Russell 3000 Index, which seeks to track the US stock market.

5. Results and Main Analysis

In this section, we analyze the obtained results to evaluate whether factor ETFs are a valuable investment. First, we examine whether the factor ETFs outperformed their regression-based benchmarks. Next, we evaluate whether factor ETFs add value to different investors. Finally, we end the analysis with a discussion of the most prominent style characteristics and provide a framework for how investors can exploit them through mean-variance portfolio optimization.

5.1 Regression-based Performance

The alpha is the part of a factor ETF's return that does not arise from exposures to systematic risk factors. Accordingly, it measures skill or ability to outperform return-attributing characteristics such as the market factor. However, alpha can be generalized by utilizing multifactor models that include more factors beyond the market factor of CAPM. The more systematic risk factors we include in the regression-based performance analysis, the models capture more of the variation in the funds' returns.

If alpha is more significant than zero, there are two main interpretations of the results. First, the alpha reflects the risk-adjusted and active return of the factor ETFs. Second, the factor ETFs might have exposure to additional systematic risk factors beyond what is captured by the asset pricing models' factor betas. However, if we assume that the different asset pricing models capture all relevant sources of systematic risk, then alpha is a good performance measure.

Table 4: Regression-based performance of factor ETFs by models.

	Positive Alpha		Negative Alpha	
	# ETF	%	# ETF	%
CAPM	92	27.54%	242	72.46%
FF3	172	51.50%	162	48.50%
FFC4	183	54.79%	151	45.21%
FF5	143	42.81%	191	57.19%
HXZ4	107	32.04%	227	67.96%
CPZ4	72	21.56%	262	78.44%
Average	128	38.37%	206	61.63%

Approximately 38% of factor ETFs generated positive alphas on average across the different models. These results imply that many funds outperformed the market and factor benchmarks. However, we must study the test statistic to determine whether these results are statistically significant. We use conventional levels of 15%, 10%, 5%, and 1% significance. Respectively, the test statistical significance thresholds are 1.44, 1.645, 1.96, and 2.576.

Table 5: Descriptive statistics of regression-based performance for factor ETFs. *For information, all estimates are presented in monthly terms and are not annualized.*

Panel A: Descriptive statistics of regression-based performance						
	CAPM α	FF3 α	FFC4 α	FF5 α	HXZ4 α	CPZ4 α
Mean	-0.2073 %	0.0235 %	0.0265 %	-0.0054 %	-0.0905 %	0.1468 %
(Average t-statistic)	(-0.5671)	-0.0089	-0.0284	(-0.1418)	(-0.3678)	-0.9323
Minimum	-3.6025 %	-1.8451 %	-1.8313 %	-2.0057 %	-2.3074 %	-1.9693 %
Median	-0.2373 %	0.0037 %	0.0143 %	-0.0379 %	-0.1462 %	0.1351 %
Maximum	3.7338 %	2.9299 %	2.6549 %	3.8879 %	4.4127 %	3.4268 %

Panel B: Rejection Frequencies of the null hypothesis in equation (35)						
Confidence	CAPM α	FF3 α	FFC4 α	FF5 α	HXZ4 α	CPZ4 α
85%	33	2	3	5	24	98
90%	26	1	3	1	20	74
95%	20	0	0	0	17	56
99%	13	0	0	0	1	34

The regression-based results show that the average factor ETF did not generate positive alpha. The average test statistic implies that the alphas for all models are statistically insignificant. Our analysis found that 38% of factor ETFs generated positive alphas on average, but most are not statistically significantly different from zero. Thus, we cannot reject the null hypothesis from equation (35) for most funds.

Moreover, the rejection frequencies imply that we can only reject the null hypothesis from equation (35) in a small portion of the total hypothesis tests conducted. E.g., 4.6% with 95% confidence. Accordingly, the average investor can obtain a greater risk-adjusted performance by holding the benchmark portfolio rather than the average factor ETF.

In addition, the results indicate that significant alphas arise when we substitute the value-weighted market benchmark with truly traded portfolios due to the increasing rejection frequency of the null hypothesis. However, the rejection frequency is not high enough to conclude that factor ETFs generate significant abnormal returns on average.

The multifactor models are constructed to explain the cross-section of expected returns. Above, we have estimated the risk-adjusted returns measured by time-series intercepts. However, the time-series intercepts are error terms in the cross-section with a mean of zero. Accordingly, we study the cross-section of factor ETFs to examine the link between factor characteristics and excess returns.

Thus, we research the sources of systematic risk and test whether the excess returns are significantly different from zero after allowing for the impact of these risks. The analysis shows that the cross-sectional intercept, λ_0, is insignificant at all conventional levels for each model except CAPM. The cross-sectional intercept for CAPM is relatively high and shows that the average excess returns are 69 bps after allowing for the impact of the market factor risk. However, we do not emphasize this result because CAPM fails to explain the cross-section of stock returns in general. Hence, the cross-sectional intercept can be inflated to a certain extent.

Table 6: Cross-sectional regressions of factor ETFs by models.

Panel A: CAPM								
	λ_α	λ_{Mkt}					R^2	$R^2_{adj.}$
Coeff.	0.0069	0.0032					0.1194	0.1168
(t-stat)	(2.05)	(0.46)						

Panel B: Fama-French 3 Factor Model								
	λ_α	λ_{Mkt}	λ_{SMB}	λ_{HML}			R^2	$R^2_{adj.}$
Coeff.	0.0015	0.0108	-0.0022	-0.0094			0.2837	0.2772
(t-stat)	(0.52)	(1.66)	(-0.84)	(-2.3)				

Panel C: Fama-French-Carhart 4 Factor Model								
	λ_α	λ_{Mkt}	λ_{SMB}	λ_{HML}	λ_{WML}		R^2	$R^2_{adj.}$
Coeff.	0.0014	0.0110	-0.0023	-0.0091	0.0047		0.3270	0.3188
(t-stat)	(0.47)	(1.69)	(-0.87)	(-2.18)	(0.79)			

Panel D: Fama-French 5 Factor Model								
	λ_α	λ_{Mkt}	λ_{SMB}	λ_{HML}	λ_{RMW}	λ_{CMA}	R^2	$R^2_{adj.}$
Coeff.	0.0016	0.0110	-0.0036	-0.0087	0.0017	-0.0030	0.4015	0.3924
(t-stat)	(0.57)	(1.71)	(-1.34)	(-2.09)	(0.65)	(-1.23)		

Panel E: Hou-Xue-Zhang 4 Factor Model								
	λ_α	λ_{Mkt}	λ_{SMB}	λ_{IA}	λ_{ROE}		R^2	$R^2_{adj.}$
Coeff.	0.0013	0.0105	-0.0034	-0.0089	-0.0002		0.3140	0.3056
(t-stat)	(0.44)	(1.63)	(-1.33)	(-2.26)	(-0.04)			

Panel F: Cremers-Petajisto-Zitzewitz 4 Factor Model

	λ_α	λ_{Big}	λ_{Small}	λ_{VMG}	λ_{WML}	R^2	$R^2_{adj.}$
Coeff.	0.0025	0.0113	-0.0042	-0.0087	0.0104	0.3289	0.3207
(t-stat)	(0.86)	(1.77)	(-1.01)	(-2.13)	(1.77)		

Our results show that none of the multifactor models have significant intercepts in the cross-section. Hence, we conclude that the average excess return of factor ETFs is not significantly different from zero after allowing for the impact from different sources of systematic risks. Accordingly, the regression-based benchmark portfolios seemingly explain the variation across factor ETFs' returns quite well. Moreover, the results show that low value, high growth, and high investment factors are significant systematic risk sources contributing to the excess returns.

5.2 Reward to Total Risk

The Sharpe ratio and M^2 are two renowned performance metrics, and these two metrics concern the first investor type we identified in section (*3.2*). These investors care about absolute risk and look for a single fund to hold their entire risky portfolio. Hence, both metrics are good evaluation metrics as they adjust the performance of the evaluated funds for total risk. The metrics are similar, with minor differences (*see sections 3.2.1 and 3.2.2*).

Moreover, both metrics are increasing for better risk-adjusted performance. Thus, the higher the metric, the better the performance. Usually, an annualized Sharpe ratio above one is considered good, but the metric also needs to be compared with the market benchmark. In comparison, M^2 outperforms the market benchmark if it is positive. Accordingly, a metric above zero is considered good.

Table 7: Descriptive statistics of reward to total risk exposure.

Panel A: Descriptive Statistics of the Sharpe Ratio and M^2

	SR (Monthly)	M^2 (Monthly)	SR (Annualized)	M^2 (Annualized)
Mean	0.1956	-0.3072 %	0.6775	-1.0640 %
Minimum	-0.3832	-2.8431 %	-1.3273	-9.8487 %
Median	0.1880	-0.3402 %	0.6514	-1.1785 %
Maximum	0.4197	0.6747 %	1.4538	2.3372 %
Benchmark	0.2657	------	0.9204	------

Panel B: Summary of higher Sharpe Ratio and positive M^2				
	Higher Sharpe Ratio		Positive M^2	
	# ETF	%	# ETF	%
Total	61	18.26%	61	18.26%

Our results show that the factor ETFs have mean and median metrics below the market benchmark. Thus, factor ETFs do not produce a greater risk-adjusted return than the market benchmark on average. Accordingly, the investors obtain a better risk-return tradeoff from investing in the value-weighted market portfolio than the average factor ETF. However, we find that 61 funds have a higher Sharpe ratio than the market benchmark and yield a positive M^2. Accordingly, 18.26% of the factor ETFs provide a better reward to total risk and outperform the market benchmark. Thus, we cannot reject the null hypotheses from equations (40) and (43) for most funds. The top 10 funds are listed and ranked in the following table.

Table 8: The top 10 performance rankings of factor ETFs measured by Sharpe Ratio and M^2.

ETF Name	Ticker	SR	M^2	Ranking
iShares Russell Top 200 Growth ETF	IWY	0.4197	0.6747 %	1
John Hancock Multi-Factor Technology ETF	JHMT	0.4075	0.6212 %	2
Vanguard Mega Cap Growth ETF	MGK	0.4037	0.6048 %	3
Schwab U.S. Large-Cap Growth ETF	SCHG	0.4005	0.5907 %	4
Vanguard Russell 1000 Growth ETF	VONG	0.4003	0.5900 %	5
iShares Russell 1000 Growth ETF	IWF	0.3989	0.5836 %	6
iShares Morningstar Growth ETF	ILCG	0.3986	0.5823 %	7
SPDR Portfolio S&P 500 Growth ETF	SPYG	0.3978	0.5788 %	8
Vanguard S&P 500 Growth ETF	VOOG	0.3967	0.5742 %	9
iShares S&P 500 Growth ETF	IVW	0.3965	0.5730 %	10

5.3 Reward for Additional Risk

The Information and Appraisal ratios determine the reward for the additional risk. Both metrics capture the risk-adjusted excess return of the factor ETFs above the market benchmark. The two are similar in style and concern the second investor type we identified in section (*3.2*). These investors hold a well-diversified portfolio, similar to the value-weighted market portfolio, and look for improved risk-adjusted performance by adding a factor ETF to their portfolio. A good Information and Appraisal ratio is above zero, but the higher the metric, the better the fund's risk-adjusted performance.

Table 9: Descriptive statistics of reward for additional risk exposure.

Panel A: Descriptive Statistics for the Information Ratio and Appraisal Ratio				
	IR (Monthly)	AR (Monthly)	IR (Annualized)	AR (Annualized)
Mean	-5.8707 %	-7.6253 %	-20.3368 %	-26.4147 %
Minimum	-40.0759 %	-41.9927 %	-138.8269 %	-145.4669 %
Median	-7.2016 %	-10.5934 %	-24.9472 %	-36.6967 %
Maximum	42.4427 %	40.8827 %	147.0257 %	141.6219 %

Panel B: Summary of Information Ratio and Appraisal Ratio				
	Positive IR		Positive AR	
	# ETF	%	# ETF	%
Total	104	31.14%	92	27.54%

Our analysis shows that both ratios have a negative median and mean. These results indicate that investors would not benefit from adding the average factor ETF to their portfolio. Consequently, investors are not rewarded for taking on such an additional risk. Although most factor ETFs do not yield a higher return for the additional risk, a few factor ETFs yield attractive risk-adjusted returns.

The Information ratio results imply that most factor ETFs do not consistently produce active returns. However, we find that 104 funds have a positive Information ratio. Accordingly, 31.14% of the factor ETFs provide a good reward for the additional risk measured by the Information ratio.

Moreover, the Appraisal ratio results show that there is little active return produced per unit of risk that the factor ETFs takes on in their style strategies. However, we find that 92 funds have a positive Appraisal ratio. Accordingly, 27.52% of the factor ETFs provide a good reward for the additional risk measured by the Appraisal ratio.

We cannot reject the null hypotheses from equations (46) and (50) for most funds. However, the top 10 funds are listed and ranked in the following table.

Table 10: The top 10 performance rankings of factor ETFs measured by Information Ratio (IR) and Appraisal Ratio (AR).

Panel A: Top 10 Information Ratio Rankings			
ETF Name	Ticker	IR	Ranking
John Hancock Multi-Factor Technology ETF	JHMT	42.4427 %	1
Schwab U.S. Large-Cap Growth ETF	SCHG	39.5301 %	2
iShares Russell Top 200 Growth ETF	IWY	39.4665 %	3
Vanguard Russell 1000 Growth ETF	VONG	39.0787 %	4
Nuveen ESG Large-Cap Growth ETF	NULG	39.0167 %	5
iShares Russell 1000 Growth ETF	IWF	38.6700 %	6
Vanguard Growth ETF	VUG	38.2660 %	7
Vanguard Mega Cap Growth ETF	MGK	38.2045 %	8
SPDR Portfolio S&P 500 Growth ETF	SPYG	35.5640 %	9
Vanguard S&P 500 Growth ETF	VOOG	35.5209 %	10

Panel B: Top 10 Appraisal Ratio Rankings			
ETF Name	Ticker	AR	Ranking
iShares Russell Top 200 Growth ETF	IWY	40.8827 %	1
Schwab U.S. Large-Cap Growth ETF	SCHG	39.6443 %	2
Vanguard Russell 1000 Growth ETF	VONG	39.1182 %	3
iShares Russell 1000 Growth ETF	IWF	38.7770 %	4
John Hancock Multi-Factor Technology ETF	JHMT	38.5669 %	5
SPDR Portfolio S&P 500 Growth ETF	SPYG	38.5589 %	6
Vanguard Mega Cap Growth ETF	MGK	38.4314 %	7
Vanguard S&P 500 Growth ETF	VOOG	38.3846 %	8
iShares S&P 500 Growth ETF	IVW	38.2718 %	9
Vanguard Growth ETF	VUG	37.7794 %	10

5.4 Reward for Systematic Risk

The Treynor ratio is the last metric we looked at and concerns the third investor type we identified in section (*3.2*). These investors care about compensation for exposure to systematic risk. The systematic risk is exposure to the market factor in this context and is captured by the CAPM beta. Thus, the funds evaluated are a potential subset of many assets in the average investor's overall portfolio.

Table 11: Descriptive statistics of reward for systematic risk exposure.

Panel A: Descriptive statistics for the Treynor Ratio		
	TR (Monthly)	TR (Annualized)
Mean	0.0104	0.1243
Minimum	-0.0024	-0.0286
Median	0.0093	0.1114
Maximum	0.1265	1.5178
Benchmark	0.0118	0.1420

Panel B: Summary of Treynor's Ratio		
	Higher Treynor Ratio	
	# ETF	%
Total	90	26.95%

Our analysis shows that factor ETFs score a lower Treynor ratio than the market benchmark on average. Hence, factor ETFs do not reward investors for exposure to systematic risk beyond the value-weighted market portfolio. Moreover, the results are quite similar to the Sharpe and M^2 results. These findings indicate that market risk is considered relatively similar to the total risk of the funds to a certain extent. Accordingly, the factor ETFs seemingly contain well-diversified portfolios.

In addition, the low Treynor ratio for factor ETFs implies that most funds are below the security market line (SML). Hence, the funds are overvalued and underperform the market benchmark in risk-adjusted terms. Therefore, investors can obtain a better risk-return tradeoff from holding the overall market rather than the average ETF.

However, the analysis also provides findings of a few funds that significantly outperform the market benchmark. We find that 90 funds have a higher Treynor ratio than the market benchmark. Accordingly, 26.95% of the factor ETFs provide a better reward to systematic risk and outperform the market benchmark. However, we cannot reject the null hypotheses from equation (53) for most funds. The top 10 funds are listed and ranked in the following table.

Table 12: The top 10 performance rankings of factor ETFs measured by Treynor's Ratio.

ETF Name	Ticker	TR	Ranking
WisdomTree Emerging Markets SmallCap Dividend UCITS ETF	DGSE	0.1265	1
RBC Quant European Dividend Leaders ETF	RPD	0.0344	2
RBC Quant Emerging Markets Dividend Leaders ETF	RXD	0.0249	3
iShares Global Monthly Dividend Index ETF CAD-Hedged	CYH	0.0246	4
Invesco DWA Technology Momentum ETF	PTF	0.0238	5
Invesco DWA Healthcare Momentum ETF	PTH	0.0230	6
BetaShares Managed Risk Global Share Fund	WRLD	0.0210	7
iShares Russell Top 200 Growth ETF	IWY	0.0205	8
John Hancock Multi-Factor Technology ETF	JHMT	0.0198	9
Vanguard Mega Cap Growth ETF	MGK	0.0195	10

5.5 Discussion of Style Characteristics

In the following discussion, we compare the style characteristics of the funds to the market benchmark and their peers.

Table 13: Descriptive statistics of average performance measures for factor ETFs sorted on style characteristics. *For information, all metrics are presented in monthly terms and are not annualized. In addition, performance metrics that outperforms the value-weighted market benchmark are written in bold text.*

	SR	M^2	IR	AR	TR
Size	0.2046	-0.2677 %	-6.1117 %	-0.2641 %	0.0093
Value	0.1698	-0.4201 %	-13.7972 %	-0.4347 %	0.0079
Growth	**0.2998**	**0.1497 %**	**19.6356 %**	**0.1981 %**	**0.0136**
Quality	0.2436	-0.0969 %	-21.1244 %	-0.0796 %	0.0108
Volatility	0.2202	-0.1992 %	-20.3356 %	-0.1359 %	0.0101
Dividend	0.1823	-0.3652 %	-24.4579 %	-0.3356 %	0.0083
Momentum	0.2553	-0.0455 %	**3.3305 %**	**0.0132 %**	**0.0118**
Blended / Multifactor	0.2293	-0.16 %	-17.17 %	-0.15 %	0.0102
Benchmark	0.2657	----	----	----	0.0118

We find that the average growth factor ETF adds value to investors who care about the reward to total risk. Moreover, we find that the average momentum factor ETFs add value to investors who care about an extra reward for additional risk and a reward for systematic risk.

Hence, growth is the only style characteristic that benefits investors who allocate all their capital to one asset. Furthermore, growth and momentum are the only style characteristics that benefit investors considering adding factor ETFs to their overall diversified portfolio. In addition, growth and momentum are the only style characteristics that benefit investors who seek to combine factor ETFs with multiple other investments to obtain a well-diversified portfolio.

Our results reveal that size, value, quality, volatility, and dividend are style characteristics that underperform the market benchmark for all performance metrics on average. In addition, we find similar results for blended style characteristics of multifactor funds. Accordingly, our study reveals that a combination of factor investing strategies has not been successfully implemented through ETFs.

Besides these results, we also find that the style characteristics of factor ETFs are highly correlated with each other and the market benchmark. Thus, there is a robust linear relationship between the style characteristics and the market portfolio.

Table 14: Correlation matrix of factor ETFs sorted on style characteristics.

	Size	Value	Growth	Quality	Volatility	Dividend	Momentum	Blended / Multifactor	Market Ptf.
Size	1								
Value	0.9829	1							
Growth	0.9602	0.9169	1						
Quality	0.9671	0.9591	0.9632	1					
Volatility	0.9630	0.9667	0.9247	0.9701	1				
Dividend	0.9747	0.9756	0.9193	0.9712	0.9793	1			
Momentum	0.9720	0.9323	0.9811	0.9473	0.9272	0.9264	1		
Blended / Multifactor	0.9828	0.9721	0.9681	0.993	0.9722	0.9787	0.9651	1	
Market Ptf.	0.9598	0.9446	0.9679	0.9883	0.9544	0.9588	0.9503	0.9872	1

5.6 Mean-Variance Optimized Portfolio

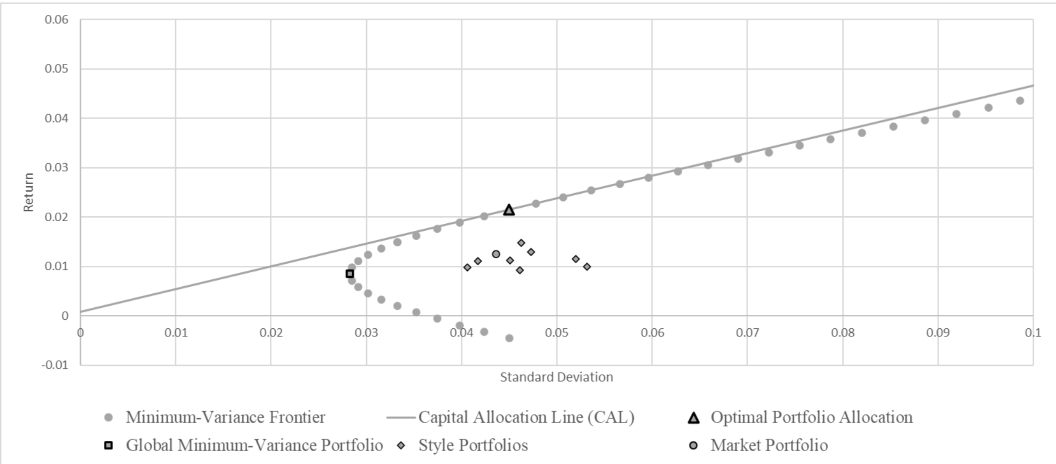
In order to exploit factor ETFs in combination with the market benchmark to diversify investors' portfolios, we utilize the variance-covariance matrix in table (15) below.

Table 15: Monthly variance-covariance matrix of equally weighted portfolios of factor ETFs sorted on style characteristics.

	Size	Value	Growth	Quality	Volatility	Dividend	Momentum	Blended / Multifactor	Market Ptf.
Size	0.0027								
Value	0.0027	0.0028							
Growth	0.0023	0.0022	0.0021						
Quality	0.0021	0.0021	0.0018	0.0017					
Volatility	0.0020	0.0021	0.0017	0.0016	0.0016				
Dividend	0.0023	0.0024	0.0019	0.0018	0.0018	0.0021			
Momentum	0.0023	0.0023	0.0021	0.0018	0.0018	0.0020	0.0022		
Blended / Multifactor	0.0023	0.0023	0.002	0.0018	0.0018	0.002	0.0020	0.0020	
Market Ptf.	0.0021	0.0022	0.0019	0.0018	0.0017	0.0019	0.0019	0.0019	0.0019

The statistics from table (14), show that all factor ETFs are highly correlated with each other and the market portfolio. This indicates that there might not be significant diversification benefits from exploiting factor ETFs in combination with the market. However, investors can short sell ETFs and thus obtain some diversification benefits.

Figure 3: Minimum-variance frontier and efficient frontier for factor ETFs sorted on style characteristics.



The best possible capital allocation line goes through the tangency portfolio that maximizes the Sharpe ratio. This portfolio, with the optimal asset allocation, is shown in table (16) below.

Table 16: Optimal Portfolio Allocation. *For information, performance metrics are denoted in monthly terms and are not annualized.*

Style Portfolios	Portfolio Weights
Size	-195.33 %
Value	19.52 %
Growth	319.84 %
Quality	-67.91 %
Volatility	162.71 %
Dividend	-21.98 %
Momentum	-25.63 %
Blended / Multifactor	-240.20 %
Market PTF	149.00 %
Sum (Σw)	100 %
Expected Return	2.15%
Standard Deviation	4.50%
Sharpe Ratio	0.4771

The monthly Sharpe Ratio of 0.4771 is the slope of the capital allocation line. Our analysis shows that the average investor who holds the market portfolio can obtain a better risk-adjusted return by exploiting factor ETFs. Investors benefit from the average growth, volatility, and value fund in addition to the value-weighted market portfolio. Moreover, investors can exploit the other style funds from short-selling to leverage their positions in the funds that maximize the Sharpe ratio.

5.7 Performance Persistence and Covid-19

This section divides the analysis into two periods (2017-2019) and (2019-2021). The latter is a unique period in our sample due to Covid-19. First, we evaluate the performance prior to and during Covid-19. Then, we examine whether the performance persisted or changed for different styles of factor ETFs. Unfortunately, many of the funds have been launched in recent years, and it can therefore be hard to tell whether the current performance up until today will persist or not. However, the performance impact of Covid-19 can tell us how the funds are impacted during periods of financial distress.

Table 17: Descriptive statistics of performance measures before and during Covid-19. *For information, all metrics are presented in monthly terms and are not annualized. In addition, performance metrics that outperforms the value-weighted market benchmark are written in bold text.*

Panel A: Performance metrics prior to Covid-19 (2017-2019)					
	SR	M^2	IR	AR	TR
Size	0.2037	-0.2361 %	-12.8903 %	-0.2204 %	0.0070
Value	0.1618	-0.3740 %	-19.8592 %	-0.3635 %	0.0057
Growth	**0.2925**	**0.0561 %**	**14.6147 %**	**0.1037 %**	**0.0100**
Quality	0.2540	-0.0705 %	-18.7871 %	-0.0593 %	0.0084
Volatility	0.2737	-0.0058 %	-14.2345 %	**0.0338 %**	**0.0095**
Dividend	0.2230	-0.1724 %	-25.5952 %	-0.1360 %	0.0076
Momentum	0.2415	-0.1115 %	-5.1363 %	-0.0534 %	0.0085
Blended / Multifactor	0.2417	-0.1111 %	-19.2675 %	-0.1044 %	0.0080
Benchmark	0.2754	----	----	----	0.0091

Panel B: Performance metrics during Covid-19 (2020-2021)					
	SR	M^2	IR	AR	TR
Size	0.2202	-0.2954 %	-0.9383 %	-0.2920 %	0.0130
Value	0.1881	-0.4781 %	-8.9442 %	-0.5036 %	0.0113
Growth	**0.3201**	**0.2743 %**	**25.6400 %**	**0.3256 %**	**0.0187**
Quality	0.2460	-0.1484 %	-24.5609 %	-0.1226 %	0.0142
Volatility	0.2011	-0.4043 %	-27.2493 %	-0.3366 %	0.0119
Dividend	0.1687	-0.5889 %	-26.5228 %	-0.5839 %	0.0100
Momentum	**0.2838**	**0.0676 %**	**12.4280 %**	**0.1424 %**	**0.0169**
Blended / Multifactor	0.2321	-0.2275 %	-17.3048 %	-0.2139 %	0.0134
Benchmark	0.2720	----	----	----	0.0155

We find that growth factor ETFs outperformed the market and added value for all types of investors prior to and during Covid-19. Moreover, our results show that volatility factor ETFs yielded a higher Treynor ratio than the market prior to Covid-19. In addition, we observe that volatility has a positive Appraisal ratio and a negative Information ratio in the same period. Accordingly, it is dependent on the investor which ratio they prefer to estimate their compensation for the additional risk. However, we would argue that the Information ratio is significantly negative. In contrast, the Appraisal ratio is slightly positive. Thus, we consider volatility factor ETFs to only provide a significant reward for systematic risk compared to the market benchmark.

Furthermore, we find that momentum factor ETFs outperformed the market benchmark for all risk-adjusted performance metrics during Covid-19. In comparison, volatility factor ETFs experienced a significant drawdown in all metrics.

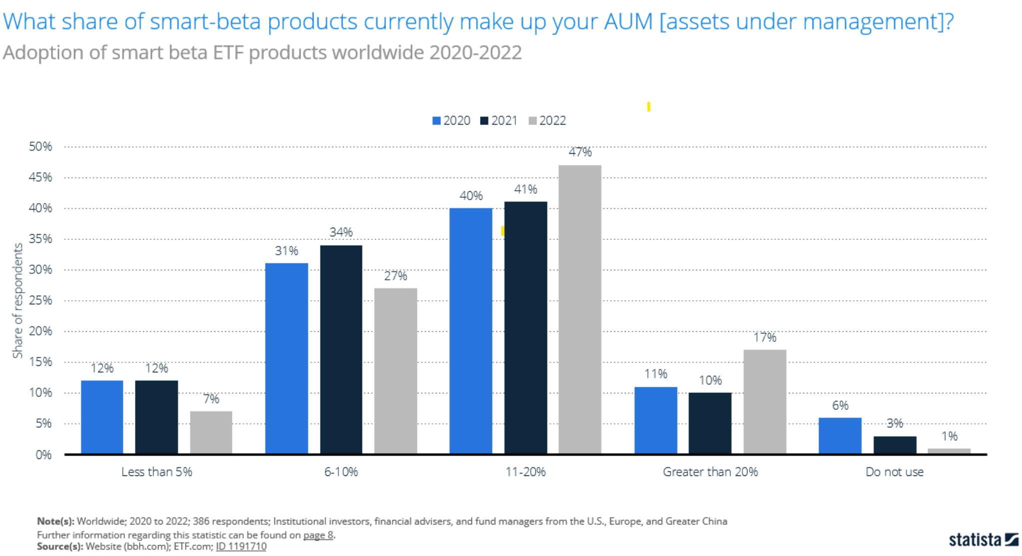
In addition, we observe that size, value, quality, dividend, and blended multifactor funds persistently underperformed for all performance metrics. Accordingly, growth factor ETFs are the only winner in both periods because growth is the only style characteristic that persistently outperformed the market benchmark. Next, we consider volatility ETFs as semi-winners in the first period and losers in the second. Then, we observe momentum factor ETFs arise from being losers in the first period and becoming winners in the second period. Our results show that the performance for some style characteristics persisted during Covid-19, with a few exceptions of volatility and momentum factor ETFs.

Furthermore, we observe an increase in the adoption of smart beta ETFs during Covid-19 (*see Figure 4 on the next page*). Hence, investors have increased their exposure to factor ETFs globally in recent years. Investors' main reasons for using factor ETFs during Covid-19 are respectively to seek returns above benchmark, generate income, mitigate risk, and reduce volatility of their overall portfolio (*see figure 5 on the next page*).

Kahn & Rudd, (1995) investigate whether the historical performance of mutual funds predicts future performance. Their findings indicate that investors need more than past performance numbers to pick future winners. Hence, historical performance does not indicate future performance. Moreover, our sample period might not represent the long-term due to market fluctuations and scenarios like Covid-19. Thus, it might not be fair to compare certain factors with each other under these market conditions.

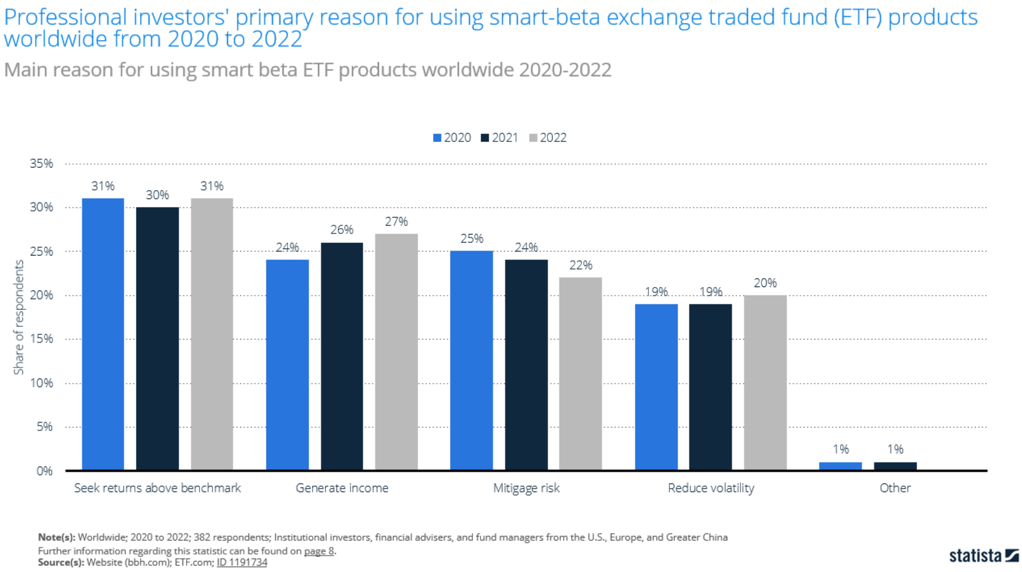
Too few observations might cause the lack of statistical significance for the regression intercepts because the mean and variance of expected returns are assumed to be constant. Consequently, we could have used shorter return periods such as weekly or daily returns to maximize the number of observations. However, the volatility might have been inflated by doing so, giving an inaccurate picture of the variance in returns.

Figure 4: Adoption of factor ETFs during Covid-19.



Website (bbh.com). (March 15, 2022). What share of smart-beta products currently make up your AUM [assets under management]? [Graph]. In Statista. Retrieved June 2, 2022, from https://www-statista-com.ezproxy.library.bi.no/statistics/1191710/etf-smart-beta-adoption-worldwide/

Figure 5: Investors' main reasons for using factor ETFs during Covid-19.



Website (bbh.com). (March 15, 2022). Professional investors' primary reason for using smart-beta exchange traded fund (ETF) products worldwide from 2020 to 2022 [Graph]. In Statista. Retrieved June 2, 2022, from https://www-statista-com.ezproxy.library.bi.no/statistics/1191734/etf-top-reason-smart-beta-use-worldwide/

6. Additional Considerations

In this section, we discuss additional considerations that might impact our conclusion. First, we examine the critical implications of our results. Then, we estimate the effect of these implications on our results and conclusion.

6.1 Survivorship Bias

Our analysis is prone to survivorship bias because we have based our analysis on a retrospective data collection process. Accordingly, we have included only ETFs that survived the whole sampling period in our sample, e.g., we have implicitly excluded funds that went bust during our sampling period. Knowing this survivorship bias is essential, as such instances might affect the results and the following conclusion.

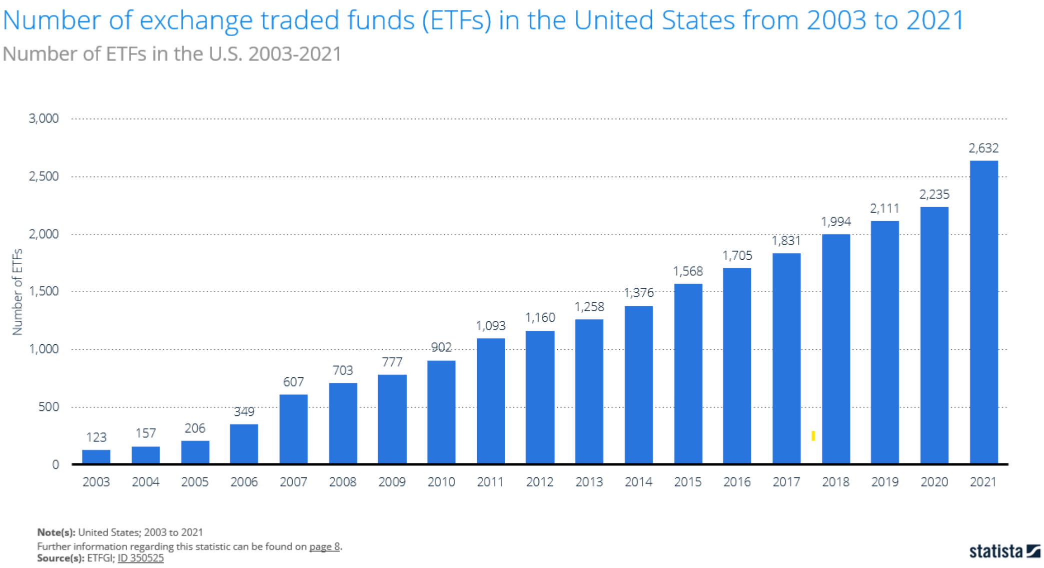
According to Investopedia, there are two main reasons for a fund to close (Chen et al., 2021). First, there may not be much demand for the fund, so asset inflows are not sufficient to justify keeping the fund open. Second, a fund may be closed by an investment manager due to a lack of performance. For our analysis, it is primarily the latter reason that would affect our results, as this would lead to skewed results if factor ETFs were closed for this reason.

To evaluate the level of skewness in our results, we examine the number of closed ETFs with respect to the total number of ETFs in the US. We find that 3.62% of ETFs are closed on average per year (*see figures 6 and 7 on the next page*). We assume that the US data on all ETFs are representative of factor ETFs both in and outside the US.

We estimate that the survivorship bias likely has minimal effects on our results for two main reasons. First, we estimate that factor ETFs make up a small part of the annual closed ETFs, and these results would have a minor impact on the average performance of style characteristics. Second, most funds underperformed the regression-based benchmarks and the market portfolio. Accordingly, our conclusion will not be significantly affected if the results are skewed towards better performance.

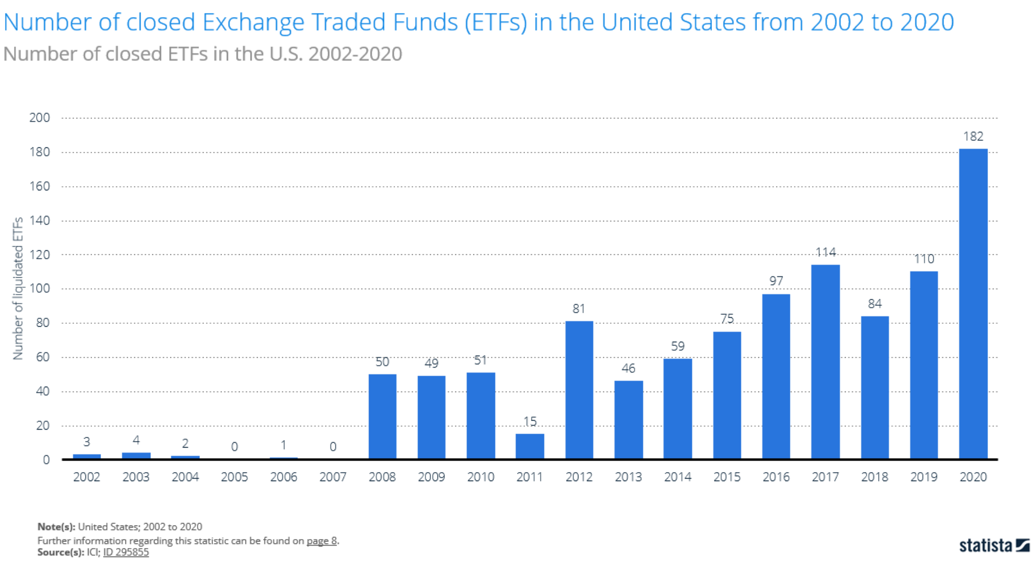
Only momentum and growth seemed to perform better than the benchmark after covid-19, and only growth performed better than the benchmark before covid-19. Thus, if we have left out some funds that were shut down for poor performance, it would have minimal effect on our overall results and conclusion. However, the presence of survivorship bias could affect the results of growth and momentum factor ETFs. Therefore, we would like to inform the reader that there is a presence of survivorship bias risk.

Figure 6: Number of ETFs in the US (2003-2021).



ETFGI. (March 11, 2022). Number of exchange traded funds (ETFs) in the United States from 2003 to 2021 [Graph]. In Statista. Retrieved June 2, 2022, from https://www-statista-com.ezproxy.library.bi.no/statistics/350525/number-etfs-usa/

Figure 7: Number of closed ETFs in the US (2002-2020).



ICI. (May 6, 2021). Number of closed Exchange Traded Funds (ETFs) in the United States from 2002 to 2020 [Graph]. In Statista. Retrieved June 2, 2022, from https://www-statista-com.ezproxy.library.bi.no/statistics/295855/number-of-liquidated-etfs-usa/

6.2 Expenses

In this thesis, we have evaluated the performance before expenses such as management fees and transaction costs. The transaction costs are challenging to estimate because they are subject to trading commissions, bid-ask spreads, and market impact costs. However, we have collected expense ratios from Bloomberg and calculated that the average management fee for factor ETFs is 50 bps annually.

Corresponding to Mateus et al., (2020) we assume that the current expense ratios (*available in June 2022*) have been constant since the factor ETFs' inception. We must make this assumption because the historical expense ratios are unavailable for our sample period.

We have neglected fees in our analysis because they would have a minimal effect on our results. The average monthly expense ratio would be the average of 50 bps divided by 12. Hence, monthly fees have an effect of approximately 4.17 bps on the expected returns after expenses. Subtracting this from the returns in our sample data has minimal effect on the returns and thus the risk-adjusted performance metrics.

Nevertheless, investors care about the performance after expenses and should consider the presence of an average of 50 bps in management fees annually. Moreover, induvial ETFs are subject to different bid-ask spreads. In addition, institutional investors will likely have lower trading commissions and higher market impact costs than individuals. Furthermore, there are also expenses concerning short-selling ETFs to exploit the factor characteristics. Hence, the effect of trading costs on performance after expenses are dependent on funds and investors respectively.

6.3 Benchmarks

Kothari and Warner, (2001) question the power of multifactor benchmarks to detect abnormal performance, and state that regression-based benchmarks will likely have lower power than characteristic-based benchmarks. Characteristic-based benchmarks are constructed portfolios using the information on a fund's holdings (Daniel et al., 1997).

This thesis uses regression-based benchmarks to estimate risk-adjusted returns for factor ETFs. However, Cremers et al., (2013), Chan et al., (2006), and Mateus et al., (2019) criticize this approach because the passive benchmark portfolios we apply might have non-zero alphas. Moreover, their research document the presence of embedded benchmark alphas due to the construction of factor portfolios. Respectively, they claim that the SMB and HML portfolios assign disproportionate weights to small-cap and high book-to-market value assets (*see equations* (55) *and* (56).

Another approach is proposed by Angelidis et al., (2013) who revisits mutual fund performance evaluation and suggests that the self-reported benchmark is more appropriate than passive portfolios with the same risk characteristics. Accordingly, many performance studies of factor funds use so-called AGT-adjusted benchmarks to detect alphas.

Hunter et al., (2014) argue that self-reported benchmarks do not provide insights into relative performance between peers of style characteristics. Accordingly, their research proposes an alternative methodology that applies active characteristic-based benchmarks for peer groups. Thus, some recent performance studies of factor funds use so-called APB-adjusted benchmarks to detect alphas.

Consequently, many approaches exist to choosing and constructing appropriate benchmarks for performance comparison purposes. Hence, choosing an appropriate benchmark that reflects the risk profile of factor ETFs is an additional consideration for investors to evaluate risk-adjusted performance.

7. Conclusion

This thesis aimed to evaluate the risk-adjusted performance of factor ETFs.

The regression-based performance analyses show that the average factor ETF had a negative risk-adjusted return. Accordingly, most funds underperformed their respective benchmarks.

When the funds evaluated are the investors' entire risky investment portfolio, we find that factor ETFs do not provide a greater reward to total risk compared to the market benchmark. Thus, investors are compensated more in risk-adjusted terms for holding the overall market than the average factor ETF. Consequently, factor ETFs do not add value by substituting the market portfolio.

When the value added by factor ETFs is evaluated as an inclusion to an overall broad and well-diversified market portfolio, the research shows that most funds do not yield improved risk-adjusted performance. Correspondingly, investors are not sufficiently compensated for taking on additional risk by including the average factor ETF in their well-diversified portfolios.

When the factor ETFs are evaluated as one of many other investments that in combination form investors' well-diversified portfolios, we also find that most funds do not improve the risk-adjusted performance. Thus, investors are not sufficiently compensated for taking on systematic risk concerning the average factor ETF.

Furthermore, our analysis of style characteristics shows that growth and momentum are valuable strategies for factor ETFs. However, there is a lack of certainty about whether the performance of these factor styles will persist in the future. Moreover, the mean-variance analysis of style characteristics provides a framework for how investors can exploit the different factors in combination with the market benchmark to improve the risk-return tradeoff of their portfolio.

Based on these findings, we conclude that the average factor ETF despite its popularity is not a valuable investment with respect to the overall market.

8. Further Research

In this thesis, we evaluate the historical risk-adjusted performance of factor ETFs in a limited period. Hence, the results may not be representative of long-term or future performance. Moreover, the continual launch of new factor ETFs will likely affect the means and variances of style characteristic performances. Therefore, we recommend future research on the longer-term performance of factor ETFs that include various factors and persistence.

Furthermore, it would be beneficial to conduct a style analysis of factor ETFs to examine how much they tilt towards different factors across different regions. Hence, we recommend future research on prominent factors in different financial markets. Future studies can also address the effects of survivorship bias, expenses, and potential non-zero benchmarks.

9. Reference List

Ang, A. (2014). *Asset Management: A Systematic Approach To Factor Investing*. Oxford University Press.

Angelidis, T., Giamouridis, D., & Tessaromatis, N. (2013). Revisiting mutual fund performance evaluation. *Journal of Banking & Finance*, *37*(5), 1759–1776. https://doi.org/10.1016/j.jbankfin.2013.01.006

Apergis, N., Poufinas, T., Panagakis, A., & Ritsios, I. (2022). How Well Do Factor Etfs Capture the Fama-French Factors? *Journal of Investment Management*, *20*(1), 48–69.

Arugaslan, O., Edwards, E., & Samant, A. (2008). Risk-adjusted performance of international mutual funds. *Managerial Finance*, *34*(1), 5–22. https://doi.org/10.1108/03074350810838190

Asness, C. S., Moskowitz, T. J., & Pedersen, L. H. (2013). Value and Momentum Everywhere. *The Journal of Finance*, *68*(3), 929–985. https://doi.org/10.1111/jofi.12021

Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics*, *9*(1), 3–18. https://doi.org/10.1016/0304-405X(81)90018-0

Barber, B. M., & Odean, T. (2000). Trading Is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors. *The Journal of Finance*, *55*(2), 773–806. https://doi.org/10.1111/0022-1082.00226

Bermejo, R., Figuerola-Ferretti, I., Hevia, T., & Santos, A. (2021). Factor investing: A stock selection methodology for the European equity market. *Heliyon*, *7*(10), e08168. https://doi.org/10.1016/j.heliyon.2021.e08168

Blitz, D., & Vidojevic, M. (2019). The Characteristics of Factor Investing. *The Journal of Portfolio Management*, *45*(3), 69–86. https://doi.org/10.3905/jpm.2019.45.3.069

Bodie, Z., Kane, A., & Marcus, A. J. (2020). *Investments* (Twelth Edition, International Student Edition). McGraw-Hill Education.

Brooks, C. (2019). *Introductory Econometrics For Finance* (4th Edition). Cambridge University Press.

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

Chan, L., Dimmock, S. G., & Lakonishok, J. (2006). *Benchmarking Money Manager Performance: Issues and Evidence* (Working Paper No. 12461). National Bureau of Economic Research. https://doi.org/10.3386/w12461

Chen, J., Rhinehart, C., & Velasquez, V. (2021, October 31). *Survivorship Bias Definition*. Investopedia. https://www.investopedia.com/terms/s/survivorshipbias.asp

Cremers, M., Petajisto, A., & Zitzewitz, E. (2013). *Should Benchmark Indices Have Alpha? Revisiting Performance Evaluation* (Working Paper No. 18050). National Bureau of Economic Research. https://doi.org/10.3386/w18050

Daniel, K., Grinblatt, M., Titman, S., & Wermers, R. (1997). Measuring Mutual Fund Performance with Characteristic-Based Benchmarks. *The Journal of Finance*, *52*(3), 1035–1058. https://doi.org/10.1111/j.1540-6261.1997.tb02724.x

DeFusco, R. A., Ivanov, S. I., & Karels, G. V. (2011). The exchange traded funds' pricing deviation: Analysis and forecasts. *Journal of Economics and Finance*, *35*(2), 181–197. https://doi.org/10.1007/s12197-009-9090-6

Durand, R. B., Lim, D., & Zumwalt, J. K. (2011). Fear and the Fama-French Factors. *Financial Management*, *40*(2), 409–426. https://doi.org/10.1111/j.1755-053X.2011.01147.x

Elton, E. J., Gruber, M. J., Brown, S. J., & Goetzmann, W. N. (2017). *Modern Portfolio Theory and Investment Analysis* (Ninth Edition). Wiley Custom.

Engelberg, J., Mclean, R. D., & Pontiff, J. (2018). Anomalies and News. *The Journal of Finance*, *73*(5), 1971–2001. https://doi.org/10.1111/jofi.12718

Engle, R. F., & Sarkar, D. (2006). Premiums-Discounts and Exchange Traded Funds. *The Journal of Derivatives*, *13*(4), 27–45. https://doi.org/10.3905/jod.2006.635418

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. https://doi.org/10.1016/0304-405X(93)90023-5

Fama, E. F., & French, K. R. (1998). Value versus Growth: The International Evidence. *The Journal of Finance*, *53*(6), 1975–1999. https://doi.org/10.1111/0022-1082.00080

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, *116*(1), 1–22. https://doi.org/10.1016/j.jfineco.2014.10.010

Fama, E. F., & MacBeth, J. D. (1973). Risk, Return, and Equilibrium: Empirical Tests. *Journal of Political Economy*, *81*(3), 607–636. https://doi.org/10.1086/260061

Filho, E. C., Moraes, F., & De-Losso, R. (2021). *Unskilled Fund Managers: Replicating Active Fund Performance With Few ETFs* (SSRN Scholarly Paper No. 3634701). https://doi.org/10.2139/ssrn.3634701

Frazzini, A., & Pedersen, L. H. (2014). Betting against beta. *Journal of Financial Economics*, *111*(1), 1–25. https://doi.org/10.1016/j.jfineco.2013.10.005

Glushkov, D. (2015). *How Smart are 'Smart Beta' ETFs? Analysis of Relative Performance and Factor Exposure* (SSRN Scholarly Paper No. 2594941). https://doi.org/10.2139/ssrn.2594941

Grinold, R. C. (1989). The fundamental law of active management. *Streetwise, The Best of the Journal of Portfolio Management*, 161–168.

Harvey, C. R., & Liu, Y. (2019). *A Census of the Factor Zoo* (SSRN Scholarly Paper No. 3341728). https://doi.org/10.2139/ssrn.3341728

Hou, K., Xue, C., & Zhang, L. (2022). *Technical Document: Factors*.

Hou, K., Xue, C., & Zhang, L. (2015). Digesting Anomalies: An Investment Approach. *The Review of Financial Studies*, *28*(3), 650–705.

Hou-Xue-Zhang q-factors data library. (2022). Retrieved 10 May 2022, from Global-q.Org. http://global-q.org/factors.html

Huij, J., Lansdorp, S., Blitz, D., & van Vliet, P. (2014). *Factor Investing: Long-Only versus Long-Short* (SSRN Scholarly Paper No. 2417221). https://doi.org/10.2139/ssrn.2417221

Hunter, D., Kandel, E., Kandel, S., & Wermers, R. (2014). Mutual fund performance evaluation with active peer benchmarks. *Journal of Financial Economics*, *112*(1), 1–29. https://doi.org/10.1016/j.jfineco.2013.12.006

Jacobs, P. D., & Levy, K. N. (2015). *Smart Beta: Too Good to Be True?* (SSRN Scholarly Paper No. 3083567). https://papers.ssrn.com/abstract=3083567

Jegadeesh, N., & Titman, S. (1993). Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *The Journal of Finance*, *48*(1), 65–91. https://doi.org/10.1111/j.1540-6261.1993.tb04702.x

Jensen, M. C. (1968). The Performance of Mutual Funds in the Period 1945-1964. *The Journal of Finance*, *23*(2), 389–416. https://doi.org/10.2307/2325404

Kahn, R. N., & Rudd, A. (1995). Does Historical Performance Predict Future Performance? *Financial Analysts Journal*, *51*(6), 43–52.

Kao, D.-L. (2002). Battle for Alphas: Hedge Funds versus Long-Only Portfolios. *Financial Analysts Journal*, *58*(2), 16–36. https://doi.org/10.2469/faj.v58.n2.2520

Kenneth R. French—Data Library. (2022). Retrieved 10 May 2022, from http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html#Developed

Kothari, S. P., & Warner, J. B. (2001). Evaluating Mutual Fund Performance. *The Journal of Finance*, *56*(5), 1985–2010. https://doi.org/10.1111/0022-1082.00397

Leippold, M., & Rueegg, R. (2021). Fama–French factor timing: The long-only integrated approach. *European Financial Management*, *27*(4), 666–700. https://doi.org/10.1111/eufm.12285

Lintner, J. (1965). Security Prices, Risk, and Maximal Gains From Diversification. *The Journal of Finance*, *20*(4), 587–615. https://doi.org/10.2307/2977249

Malkiel, B. G. (2014). Is Smart Beta Really Smart? *The Journal of Portfolio Management*, *40*(5), 127–134. https://doi.org/10.3905/jpm.2014.40.5.127

Markowitz, H. (1952). *Portoflio selection*. The Journal of Finance.

Mateus, C., B. Mateus, I., & Soggiu, M. (2020). Do smart beta ETFs deliver persistent performance? *Journal of Asset Management*, *21*(5), 413–427. https://doi.org/10.1057/s41260-020-00174-1

Mateus, C., Mateus, I. B., & Todorovic, N. (2019). Benchmark-adjusted performance of US equity mutual funds and the issue of prospectus benchmarks. *Journal of Asset Management*, *20*(1), 15–30. https://doi.org/10.1057/s41260-018-0101-z

Miffre, J. (2016). Long-short commodity investing: A review of the literature. *Journal of Commodity Markets*, *1*(1), 3–13. https://doi.org/10.1016/j.jcomm.2016.01.001

Modigliani, F., & Modigliani, L. (1997). Risk-adjusted performance. *Journal of Portfolio Management*, *23*(2), 45–54.

Odean, T. (1999). Do Investors Trade Too Much? *American Economic Review*, *89*(5), 1279–1298. https://doi.org/10.1257/aer.89.5.1279

Rosenberg, B., Reid, K., & Lanstein, R. (2021). Persuasive Evidence of Market Inefficiency (Spring 1985). In *Persuasive Evidence of Market Inefficiency (Spring 1985)* (pp. 48–55). Princeton University Press. https://doi.org/10.1515/9781400829408-007

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, *13*(3), 341–360. https://doi.org/10.1016/0022-0531(76)90046-6

Sharpe, W. F. (1964). Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk*. *The Journal of Finance*, *19*(3), 425–442. https://doi.org/10.1111/j.1540-6261.1964.tb02865.x

Sharpe, W. F. (1966). Mutual Fund Performance. *The Journal of Business*, *39*(1), 119–138.

Sharpe, W. F. (1992). Asset Allocation: Management Style and Performance Measurement. *Journal of Portfolio Management*, *18*(2), 7.

Simons, K. (1998). Risk-adjusted performance of mutual funds. *New England Economic Review*, *9*, 33–48.

The EY Global ETF Report 2017. (2017). Retrieved 10 June 2022, from https://eyfinancialservicesthoughtgallery.ie/wp-content/uploads/2017/12/The-EY-Global-ETF-Report-2017.pdf

Treynor, J. L., & Black, F. (1973). How to Use Security Analysis to Improve Portfolio Selection. *The Journal of Business*, *46*(1), 66–86.

Treynor, J. L., & Mazuy, K. (1966). *Can mutual funds outguess the market ?* Harvard Business Review, volume 43, p. 131-136.

Twenty years young. (2013, January). *The Economist*. https://www.economist.com/finance-and-economics/2013/01/26/twenty-years-young