UNIVERSITY OF VAASA

SCHOOL OF ACCOUNTING AND FINANCE

Daniel Lepistö

PERFORMANCE OF THE LOW BETA-TO-ETF STRATEGY AFTER ETF SELLOFFS

Master's Degree

Programme in Finance

VAASA 2020

TABLE OF CONTENTS

LIST OF GRAPHS

LIST OF TABLES AND FIGURES

UNIVERSITY OF VAASA
School of Finance
Author:	Daniel Lepistö
Topic of the Thesis:	Performance of the Low Beta-To-ETF Strategy After ETF Selloffs
Name of the Supervisor:	Janne Äijö
Degree:	Master of Science in Economics & Business Administration
Master's Programme:	Finance
Year of Starting the Degree:	2018
Year of Completing the Thesis:	2020 **Pages:** 96

ABSTRACT

This thesis investigates whether using the low beta-to-ETF measure introduced by Lynch, Page, Panariello & Giroux (2019) after high volume ETF selloffs produce investors abnormal returns on a short-term and long-term basis. This low beta-to-ETF strategy aims to capture the so-called ETF outsider stocks that are unintentionally co-moving with the rest of the ETF constituents. The key motivation is that the investors should buy these outsiders after a downward price-pressure from a selloff event and then capture the price reversion after the situation normalizes.

By applying Lynch et al. (2019) methodology, a set of both U.S. and European broad-index and factor ETFs are examined for selloff days from 01/2016 to 08/2020. Overall, 202 outsider ETF constituent portfolios are created, which are then each held for 40 days. These portfolios are then combined into long-term systematic strategies for each ETF and are backtested for the entirety of the sample period and estimated with the Fama-French Five-factor model with betting against beta expansion. Additionally, a fundamental proxy component of Piotroski F-Score is suggested to enhance stock-picking for the portfolios, as Lynch et al. (2019) discussed the benefits of applying fundamental analysis for their strategy.

The results of this thesis indicate that short-term outsider stock strategy for these sample ETFs produces an average cumulative abnormal return of 1.3% after a 40-day holding period. The long-term systematic strategy fails to generate statistically significant alpha estimated by the Fama-French Five-factor model but produces superior Sharpe ratios and reduces volatility compared to just passively holding the parent ETFs. Finally, the fundamental proxy suggested as an enhancing stock-picking factor does not improve the abnormal returns obtained in this thesis.

KEYWORDS: Exchange-Traded Funds, Co-movement, Selloffs, Low beta anomaly.

VAASAN YLIOPISTO
Laskentatoimen ja rahoituksen yksikkö
Tekijä:	Daniel Lepistö
Tutkielman nimi:	Performance of the Low Beta-To-ETF Strategy After ETF Selloffs
Ohjaaja:	Janne Äijö
Tutkinto:	Kauppatieteiden maisteri
Maisteriohjelma:	Rahoitus
Tutkinnon aloitusvuosi:	2018
Tutkinnon valmistumisvuosi:	2020 **Pages:** 96

TIIVISTELMÄ

Tämä tutkielma pyrkii selvittämään, voiko Lynch, Page, Panariello & Giroux (2019) esiintuomaa sijoitusstrategiaa hyödyntämällä saada epänormaaleja tuottoja. Strategiassa sijoitetaan matalan pörssinoteeratun rahaston kautta mitatun betan osakkeisiin sellaisten päivien jälkeen, kun pörssinoteerattua rahastoa myydään epätavallisen suurella volyymilla, ja rahastolla on samanaikaisesti negatiivinen päivätuotto. Perimmäisenä tarkoituksena on vangita sellaisten osakkeitten tuottokäänne, jotka ovat liikkuneet alas rahaston mukana.

Käyttämällä Lynch ym. (2019) menetelmiä, ja valitsemalla tutkimukseen USA- sekä Eurooppa pörssinoteerattuja indeksirahastoja sekä faktorirahastoja aikavälillä 01/2016 – 08/2020, yhteensä 202 matalan betan portfoliota saadaan luotua. Jokaista portfoliota pidetään 40 päivää, ennen myymistä. Tämän lyhyen ajan strategian tuloksen yhdistetään sen jälkeen yhdeksi pitkäksi strategiaksi jokaiselle rahastolle erikseen. Tämän pitkänajan strategian tuottoja arvioidaan Fama-French viiden faktorin mallilla ja Betting Against Beta lisäkomponentilla. Vielä lopuksi, portfoliot painotetaan tässä tutkielmassa ehdotetulla Piotroskin F-Score fundamenttikomponentilla, jonka tarkoitus on auttaa osakepoiminnassa portfolioiden luontivaiheessa.

Tämän tutkielman tulokset osoittavat, että lyhyen ajan matalan rahaston kautta mitatun betan osakepoimintastrategian epänormaalit tuotot kumuloituvat 1.30% asti, 40 päivän jälkeen. Pitkän aikavälin strategia ei tuota tilastollisesti merkittäviä epänormaaleja tuottoja yhdellekään rahastolle, mutta tehostaa tuottoja Sharpen luvulla mitattuna. Lopulta, fundamenttikomponentilla painotetut portfoliot eivät myöskään tuota oleellisesti suurempia epänormaaleja tuottoja.

1. INTRODUCTION

The popularity of passive investing has expanded significantly over the recent years, as passive index fund assets exceeded $10 trillion in 2020, and the trend is set to continue (Financial Times 2020). Similarly, Moody's estimated that the passive index funds will acquire over 50% share of total fund assets latest by 2024 (Reuters 2017). The preference of passive investing over active has ensured its increasing popularity from the lower costs, simplicity, and better long-run performance compared to active management (French 2008).

Especially the novel product of Exchange-Traded funds (ETFs) has gathered mainstream popularity among investors, as the total share of ETFs of the passive fund assets already surpassed 40% in 2017 (Reuters 2017). The continuous innovation and decreasing fund costs of ETFs have ensured that active management and traditional mutual funds keep losing market share. Morningstar's (2019a) annual fund fee study found out that the average passive fund expense ratio was just 0.15% compared to actively managed funds' 0.67%. This means that the average active fund investor pays 4.5 times the expenses compared to the passive investor, which are most often not backed with superior returns.

However, recent studies have shown increased concern on the effects of passive investing to the security level price discovery. Several studies have investigated the relationship between indexing and stock correlations and co-movements (Agarwal, Hanouna, Moussawi, Stahel 2018; Barberis, Shleifer & Wurgler 2005; Coles, Heath & Ringgenberg 2020; Glosten, Nallareddy & Zou 2016). The studies have suggested that the indexing tends to increase the constituent stock correlations, commonality in stock liquidity, and reduced information production. These findings open several new questions about the performance of individual stocks when the fundamental aspect of price-discovery might further decrease when index-linked investing becomes more common.

Furthermore, if investor limits their investment portfolio to passive index-following, it is impossible to beat the market as one captures only the followed market return, minus fund expenses. Naturally, not everyone can beat the market, and as Sharpe (1991) stated, financial markets are practically a zero-sum game. If there are winning managers, there must be losing managers, too. New, more dynamic ETF strategies have emerged to capture the gap between active and passive strategies. Smart-beta and strategic-beta ETFs

invest and diversify by different market risk factors found by academia while still beating active funds with lower fees (iShares 2020a).

This thesis is inspired to challenge the megatrend of passive investing by showing that using the active investment strategy suggested by Lynch et al. (2019) that exploits the ETF constituent co-movements by measuring the constituent stock betas against their parent ETFs and then investing to the lowest-beta or so-called, outsider stocks can yield investors abnormal returns. As previous studies on the subject of ETF constituent correlations inspect either a predetermined sample of sector SPDR (Standard & Poor's Depository Receipt) funds (Lynch et al. 2019) or a wide cross-section of ETFs (Da & Shive, 2018), this paper investigates the usefulness of similar strategies in a sample of popular broad-index, and factor ETFs. Motivation is to show further that active investing has its place in the mix of market participation and advance discussion about the suggested negative aspects of the ongoing trend of excessive passive investing.

1.1. Purpose of the study

The purpose of this thesis is to investigate whether the investment strategy proposed by Lynch et al. (2019) is still applicable and that exploiting the strategy provides abnormal returns. The sample in this paper tries to bring fresh insight compared to the U.S. broad-market and sector-focused funds that were investigated in the original study by selecting U.S. and European-based broad-index and factor ETFs for inspection. The factor ETFs include quality factor weighting and various ESG factor funds (Environmental, Social, and Governance).

This thesis contributes to the earlier literature by two key aspects. First, as Lynch et al. (2019) measure the short-term average cumulative abnormal returns by forming equal-weighted portfolios, these portfolios are expanded to value-weighted and fundamentally weighted alternatives in this thesis. The value-weighting is measured by each constituent's proportional holding weight in their parent ETFs. The motivation of value-weighting is based on the presumption that stocks with higher ETF ownership tend to experience more volatility during demand shocks (Ben-David, Franzoni & Moussawi 2018). This finding is not directly applied to this thesis, but instead used as a basis for setting the stage for the motivations of different portfolio weightings than in Lynch et al. (2019) paper.

Moreover, the fundamental weights are calculated from each constituent's fundamental rating proxied by Piotroski (2000) F-Score. Using fundamental weighting aims to solve the issue Lynch et al. (2019) discussed, where the lack of discretionary stock level analysis in the portfolio creation phase can limit the overall returns of the strategy.

Secondly, as Lynch et al. (2019) do not evaluate the systematic long-term performance of the low beta-to-ETF strategy in their paper, this thesis aims to backtest the strategy and investigate whether it can produce abnormal returns or better risk-adjusted returns for the investors on a long-term basis. The Fama-French Five-factor model is used to estimate the abnormal returns, and the risk-adjusted returns are measured with the Sharpe Ratio.

As the previous research discussing the ETF mispricings and ETF constituent co-movements has found that ETF constituent correlations are significant (Da & Shive 2018) and systematic strategies can be formed to capture abnormal returns from this informational inefficiency (Lynch et al. 2019), it is expected that similarly significant results can be obtained with the unique sample selected in this thesis. The studies mentioned earlier and in the literary review section are used as the main inspiration and motivator for this paper. Especially methods from Lynch et al. (2019) are utilized to capture a possible time-structure for the abnormal return after ETF selloff dates.

Inspired by Lynch et al. (2019), all the selected ETFs will be observed for any significant trading volume increases combined with negative ETF return for the same day. These captured days are treated as the ETF selloff days. A contrarian trading strategy is then utilized, where the ETF constituent stock sensitivities are measured against the parent ETF, creating each constituent a beta-to-ETF. From these betas, the decile of lowest betas-to-ETF form an outsider portfolio that is expected to outperform the ETF, by the theory that the outsider stocks are being unintentionally dragged down during selloffs with the rest of the fund and will rebound after the trading normalizes from the selloff day.

The fund selection in this thesis focuses mainly on singular region-based iShares ETFs. As for the outsider portfolio forming, comprehensive historical fund holding data is needed, which practically excluded most of the other fund providers apart from iShares, as their historical fund holding data is exhaustive compared to others. Thus, most of the funds used in this paper are either Europe or U.S.-based ETFs with a region or fundamental based strategy. This restriction is first to ensure comparability with the original (Lynch et al., 2019) study in which the authors used only U.S. broad-index and

sector SPDRs, and secondly to minimize portfolio creation and price staleness issues with different time-zones and market holidays, as the portfolio holding periods are relatively short.

The European fund selection mixed with U.S.-based funds provides a great updated cross-sectional look of the usability of the strategy perceived by Lynch et al. (2019) as especially systematic anomalies in the financial markets tend to disappear rapidly or are excessively challenging to reproduce (Hou, Xue & Zhang 2020). Additionally, this thesis provides much-needed contrast to ETF based studies since most of the research involving the topic are either from U.S. or Asian markets, even though the European ETF mass has grown almost 40% from the 2016 and is projected to hit $2 trillion in 2024 (ETFStream 2020).

However, as the U.S.-based factor ETFs and European ESG ETFs are not entirely comparable to the U.S. sector-based ETFs and broad-index funds used by Lynch et al. (2019), results might be unpredictable. In their paper, the authors investigated mainly sector-based SPRDs, SPY, the largest S&P 500 broad-index fund globally, and IJR, the S&P 500 small-cap ETF. Thus, this thesis is motivated to pursue whether the effects for the broad-index ETFs can be replicated in a wider, more diverse context and further applied to factor ETFs to prove the usability of the beta-to-ETF component as a convenient investment metric for contrarian investors to benefit from the continuously occurring selloff events.

1.2. Research Hypotheses

This thesis is motivated to investigate a total of three hypotheses. The first one is the main hypothesis of the thesis, and the second and the third are more auxiliary and tend to derive results from the findings of the main hypothesis. The first hypothesis of this thesis can be considered the base of this thesis, and it continues on the findings Lynch et al. (2019) captured in their paper. The first hypothesis is the following:

H_1: *The low beta-to-ETF stocks generate statistically significant abnormal returns after high volume ETF selloffs.*

While being auxiliary to the main hypothesis, the second hypothesis in this thesis provides insight into the lack of systematic backtesting provided for the low beta-to-ETF strategy

in the original (Lynch et al. 2019) paper. As the existence and size of the opportunity is captured, it is furthermore important to investigate whether the strategy proves its usefulness by the traditional portfolio evaluation metrics. In this thesis, the short-term portfolios are transformed into one long time-series of returns that examines whether systematically investing with the low beta-to-ETF selloff portfolio rule would generate abnormal returns in the long-term. Thus, the second hypothesis is stated as follows:

H_2: The low beta-to-ETF strategy generates statistically significant abnormal returns as a long-term systematic strategy.

Finally, the third and the last hypothesis presented in this thesis contributes towards solving the issue Lynch et al. (2019) mentioned in their paper, relating to the fundamental analysis in the portfolio creation process. As the ETF selloffs tend to push all the stocks in the ETF downwards, there should be stocks where this is not fundamentally warranted and is just overreacting with the rest of the index. Thus, in this paper, a discrete financial ratio of Piotroski F-Score (Piotroski 2000) is used as a fundamental proxy to investigate whether sorting outsider stocks by their fundamental attributes enhances the bounce-back effect of the outsider stocks:

H_3: The fundamental weighting provides better, statistically significant abnormal returns for low beta-to-ETF portfolios after high volume ETF selloffs.

2. THEORETICAL BACKGROUND

In this chapter, the main characteristics of passive investing and the Efficient Market Hypothesis are explained. Additionally, ETFs' key features and details about stock correlation and co-movement are introduced, as highly correlating assets without a fundamental basis can provide active investors possibilities to exploit this connection and predict future returns. Finally, methods and metrics to evaluate portfolio performance in the context of this thesis are briefly visited.

2.1. Passive investing

Passive investing has achieved an unquestionable megatrend status over the past decades. In the financial markets, capital has been moving towards passive, index-linked investments at an increasing rate. In the fall of 2019, assets allocated in index-linked U.S. equity funds overtook active U.S.-equity funds the first time in history (Bloomberg 2019a), as seen in Graph 1. The low costs of index tracking and continuous financial innovation, especially with exchange-traded products (ETPs), have accelerated the shift away from active investing.
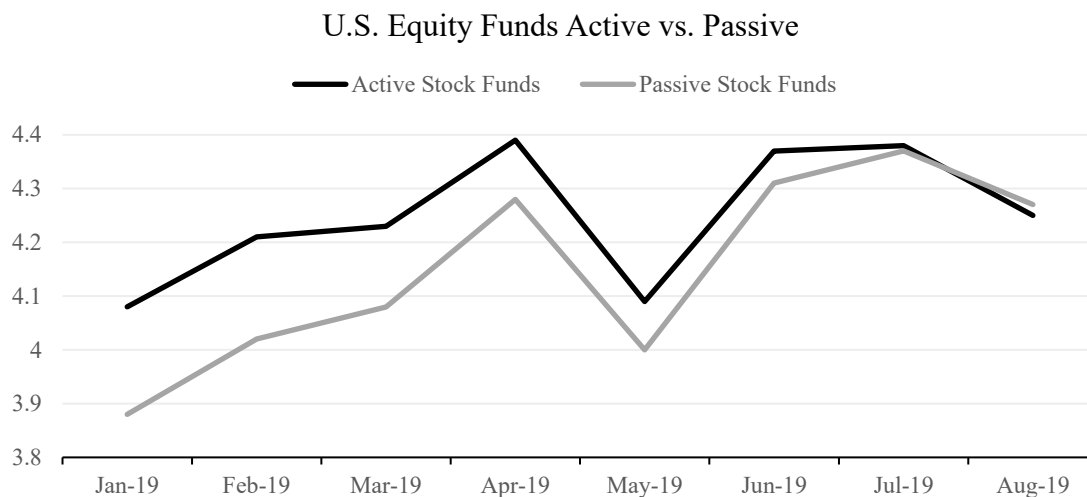
Graph 1. Active U.S. Equity funds versus Passive Equity Funds in Trillions of USD (Morningstar 2019b).

Overall, the trend of passive versus active assets has been developing towards passive biased a long time. As seen in Graph 2, from the year 2008 onwards, the amount of passive equity fund assets has been catching up with the amount of active equity fund assets and overtook them in August 2019, as mentioned.

U.S. Equity Funds Active Assets vs. Passive Assets
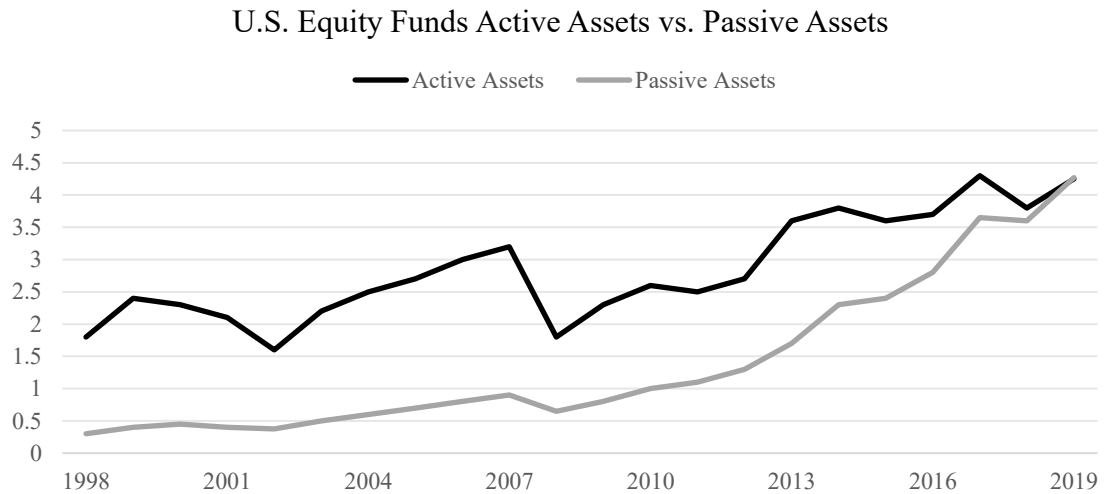
Graph 2. U.S. Equity Funds Active Assets versus Passive Assets in Trillions of USD (Morningstar 2019b).

Similarly, the equity fund flows tell the same story, as the actively managed equity funds have been experiencing consistent outflows from 2007 onwards. On the other hand, passive funds have experienced almost the reverse and gained significant inflows from 2007. Graph 3 illustrates this further:

U.S. Equity Fund Flows
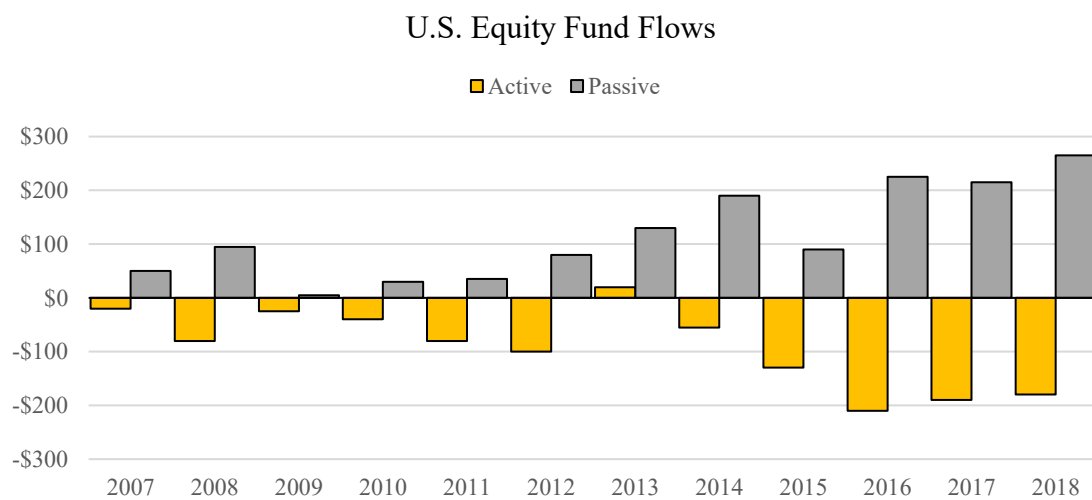
Graph 3. U.S. Equity Fund Flows in Billions of USD (Morningstar 2019b).

Theory and literature support this paradigm change. As Sharpe (1991) states, including costs in the equation, actively managed returns will be lower than passively managed ones in the long run. French (2008) instead explains that active investing is a negative-sum game. The passive market portfolio outperforms the aggregate of the active portfolios over time. As active investing seems unjustified by a purely mathematical and logical perspective, finding active investing entirely rational becomes more difficult.

However, some critique for increasing passive investing has recently surfaced, as it is argued that the price-discovery of equities has decreased by this increase in popularity. As the passive investing takes no standpoint in the valuation of the individual assets in the index, but the overall development of the followed market segment, it is possible that some constituents in indexes can co-move without reasonable fundamental reason (Lynch et al. 2019). Moreover, as even the small illiquid stocks in broad indexes are linked to trillions of dollars in indexed assets can further complicate the valuations for index constituents (Bloomberg 2019b).

The increasing popularity of passive and index-linked investing has created additional effects to the stock-index dynamic. Index inclusion or exclusion effect occurs when a stock passes the thresholds to be included in or excluded from an index. Petajisto (2011) shows that S&P 500 new additions have experienced an average +9% increase in price just around the inclusion event. Similarly, in the event of exclusion, the deleted stocks' prices have suffered as much as -15%. Additionally, when new stock is included in an index, it starts to move more closely with the other index constituents and less with the rest of the market.

Sharpe (1991) divides the stock market participants into active and passive investors. As the active investors deviate from the market index weights, passive investors settle for holding the market index. However, by combining all the active investors' holdings, they jointly hold the same whole market index as the passive investors. Conversely, the passive group receives the market return just by holding the index. This leads logically to the situation where active investing in aggregate loses to the passive investing by the transaction costs induced by their activity.

Heje Pedersen (2018) argues that the equation between active and passive investors is not as straightforward as Sharpe (1991) introduced. As Heje Pedersen (2018) suggests, even if the market portfolio or index is usually seen as constant, numerous company additions and deletions happen continuously. As the market portfolio changes over time, even

passive investors must regularly trade. Additionally, a hypothetical situation where everyone is investing passively would not work, as the market portfolio additions such as, initial public offerings, would be ultimately mispriced as no one would analyze the new company fundamentals. Eventually, anyone could list a company at any price, as the passive investors would be required to take it as a part of the market portfolio.

Finally, Coles et al. (2020) state that passive investors are somewhat freeriding the groundwork made by active investors by the form of fundamental analysis. Markets would not operate properly if everyone shifted to passive investing, as there must be active managers analyzing company fundamentals and making active investment decisions to buy undervalued and sell overvalued companies. In the context of this paper, to capture the effects of passive investing and co-movement in stock markets, ETFs are used as a convenient proxy.

2.2. The Efficient Market Hypothesis

Fama (1970) introduced the novel approach to inspect the functioning of the financial markets and security prices. As previously introduced, the EMH suggests that during any given moment, the market prices of securities must reflect all available information. The basis of the financial market is that it provides a platform for people to trade securities to each other to create wealth for themselves. In the markets, rational investors buy and sell securities by justifying their trades by their own expectations of the future price development of the security. The EMH essentially introduces a situation where the traders' profits are not necessarily dependent on their skill, but just coincidence.

Furthermore, the EMH explains the question of whether the security prices in the market reflect their true underlying fundamental values. As the EMH suggests, security prices move continuously and only when new fundamental information about a security is introduced. To illustrate the sentiment behind the EMH, a situation where traders become pessimistic about a stock when there has not been any new negative fundamental information would be immediately countered by optimistic rational investors to return to the equilibrium. However, if the negative sentiment was warranted by negative fundamental news, the stock price would decrease enough to gain a new equilibrium that incorporates the new negative information. (Barberis & Thaler 2002).

The EMH is often divided into three different subsets of information requirements, relating to market efficiency conditions. Firstly, the weak form of market efficiency states that security prices cannot be forecasted from past, historical information. In other words, any technical analysis of the historical stock prices should not yield a profit to investors. With technical analysis, investors study the historical security prices to predict the future development of the prices. As the conditions made by the weak form of EMH are relatively easy to test empirically, a significant amount of research has been made on the topic. (Burton, Shah & Shah 2013).

Secondly, the semi-strong version of the EMH suggests that security prices cannot be forecasted by using historical prices or any public fundamental information. Essentially, the semi-strong form implicates that active portfolio management should not provide superior returns in comparison to passive management. Previous research supports this, as the average equity portfolios underperform the average passive portfolios over a long period of time (Fama & French 2010). As the semi-strong form of EMH states, individuals have no superior means to gather information, as it is all publicly available, and any restrictions such as financial or positional, do not exist.

Finally, the strong form of EMH demonstrates a hypothetical situation where the security prices reflect all information, including private and public, all the time. In principle, the strong form means that even the silent information that has not been released to the public is already incorporated to the market prices of securities. The strong form of EMH is the least supported of the three subsets, but it provides a framework for further discussion. Especially, the effect of insider trading should be captured in the market prices if the strong form of EMH was true. (Burton et al. 2013).

Overall, the semi-strong form of EMH is usually the most supported subset (Bodie, Kane & Marcus 2009). However, as the core concept of the EMH is to explain price functions in the financial market, naturally, opposing research exists. Starting from the weak form of the EMH, studies have unveiled effects and price patterns that suggest return predictability in stock prices. Jegadeesh & Titman (1993) introduce an investment strategy called momentum, which provides returns by buying past winners and selling past losers in the market. In other words, securities that have fared poorly recently are being shorted and securities that have fared well are being bought. As the weak form of EMH states that the historical prices should not forecast future returns, the effects of momentum strategy are significant. Additionally, Bondt & Thaler (1985) show short-term

overreaction in the recent winners and recent losers' returns. This effect is more commonly known as the contrarian investment strategy.

This momentum and contrarian anomalies are exceptions to the weak form subset of the EMH. For the semi-strong subset of EMH, other significant anomalies have been studied. For instance, value stocks that are companies with high book-to-market ratios, systematically produce superior returns compared to other stocks (Lakonishok, Shleifer and Vishny 1994). Post-earnings announcement drift where the security prices continue increasing after positive earnings announcements is a situation where the information unveiled from the earnings announcement is not fully incorporated into the stock price immediately after the announcement (Bernard & Thomas 1990).

Furthermore, recent additional ways of thinking to oppose the EMH have lately emerged. Pedersen (2015), for example describes the financial markets as efficiently inefficient. This idea combines the observed market efficiency with the studied inefficiencies. Financial markets are perceived as nearly efficient with this school of thought due to high competition, but inefficient enough because of financial anomalies and investor behavioral biases. This market inefficiency allows skilled active portfolio managers to create wealth and showcase that operating in the markets is not just a game of chance.

To conclude, there are exceptions to both weak and semi-strong forms of EMH. As the EMH intends to prove that markets are efficient, it is problematic to hypothesize that the number of exceptions to the general rule is significant. Recent studies have introduced an additional viewpoint to market efficiency by combining efficient and inefficient markets (Pedersen 2015). Moreover, other factors negatively affect the applying of the EMH than just anomalies, as the EMH is built on the supposition of functioning arbitrage that will be next discussed further.

2.2.1. Arbitrage and limits to arbitrage

As this chapter thus far has explained the concept of the EMH, it is necessary to introduce the underlying process called arbitrage, that is behind the EMH. By the textbook, definition arbitrage is a simultaneous process of buying and selling a security, which captures a riskless profit. Theoretically, arbitrage does not require any capital as the arbitrageur conducts the buying and selling at the same time and secures the profit immediately. As a theorem, arbitrage is a simple concept, but it is difficult to reproduce in real life. (Shleifer & Vishny 1997).

As the basis of the EMH is that security prices in the market must continuously reflect all the available fundamental information, there must be a process that ensures that the security prices do not deviate from their fundamental values. The arbitrage prevents these price deviations. To demonstrate, if the price of one security does not represent its fundamental value, arbitrageurs should immediately buy or sell the security until the price equilibrium is obtained. In theory, the process sounds relatively simple, but several limits are perceived in this process in practice.

The critical problem with arbitrage is that by the textbook definition, it should be riskless. The perceived riskiness results in a situation where arbitrageurs hesitate to correct mispricing since the riskiness increases costs, and any costs to the arbitrage process would defeat its textbook purpose. Shleifer & Vishny (1997) show that in the real world, the possibility of forced liquidation before the security price converts back to its equilibrium creates a significant risk to the arbitrageurs since the resources are finite. Additionally, as the textbook arbitrage requires a simultaneous purchase and sale of a security, short selling must be available for the security. This is often not possible, and even when a short sale can be implemented, there are additional costs to this process (Barberis & Thaler 2002).

As the textbook arbitrage has been shown by studies to be both risky and costly, it is necessary to investigate what magnitude can it be implemented in real-life situations. Froot & Dabora (1999) inspect two dual-listed companies and compare their market prices to each other. These companies, Royal Dutch and Shell Transport, are independent companies that have merged their cash flows by a 60:40 ratio by agreement. This creates a situation where the Royal Dutch stock price should always be 1.5 times the price of the Shell Transport stock as these companies merge their cash flows in a predetermined manner.

Results of Froot & Dabora (1999) paper indicate that the relative pricing between the two stocks was significantly inefficient. For example, during the inspection period, the Royal Dutch stock price was as much as 35% underpriced, and sometimes over 15% overpriced. If the arbitrage in this situation would be riskless and costless, it should be simple to buy Royal Dutch when it is underpriced and short it later when it is overpriced. The difficulty here stems from the fact that these price deviations from the equilibrium last increasingly long periods when the risk of finite capital and early forced liquidation is nonzero (Barberis & Thaler 2002).

These limits of arbitrage create additional problems in financial bubbles. As the EMH suggests that the prices should converge to their fundamental values, the extended price deviations raise the problem of whether the arbitrage works in real life quickly enough. As Froot & Dabora (1999) investigate the Royal Dutch and Shell Transport, the relative price between the stocks converged to equilibrium in roughly 2002, when the inspection period started in 1980. The issue is whether the arbitrage process is efficient enough when eliminating this arbitrage opportunity took over 20 years or is there significant limits to this process, such as Shleifer & Vishny (1997) have suggested.

Consequently, even if the arbitrage was riskless, purchases and shorts were entirely costless, and there were no taxes to pay from the riskless profit, the final limit to arbitrage comes from the uncertainty that the financial model used to capture the fundamental value of the stock, is wrong. This model risk creates an additional issue that the model used indicates arbitrage opportunity, even when the market price is fundamentally correct. Being familiar with the EMH and arbitrage is crucial for understanding the underlying processes behind ETFs that will be introduced next.

2.3. ETF characteristics

An ETF can be explained as an investment vehicle that projects their pre-determined underlying basket of securities. ETFs function very similarly to mutual funds, but dissimilarities do also exist. The critical difference between an ETF and a mutual fund is that the ETF can be traded on the secondary markets. This is advantageous, as mutual fund investors who intend to cash out their positions must wait for the end of the day as the fund's net asset value (NAV) is calculated only once in the day (Greene, Hodges & Rakowski 2007). For ETF investors, such restrictions do not apply, and investors can freely trade their fund positions with other investors when the market is open. Price discovery for ETFs is thus a factor between the fund's NAV and its market price. For mutual funds and ETFs, NAV works similarly and describes the fund market value per share at a specific point in time. NAV can be calculated by dividing the fund's total assets minus liabilities by shares outstanding. NAV is based on the closing prices of the underlying securities.

In 1993 the first ETF, S&P Depositary Receipt (SPDR), was issued. As SPDR allowed the investors to invest in the S&P 500 index's contents with a single trade, invention soon gained more popularity. As the total number of ETFs worldwide was just over 200 in the

year 2001, the number of different funds has soared to over 6,500 funds in 2019. Additionally, total ETF assets worldwide have topped over 5 trillion dollars in 2019. (ETFGI, 2019).
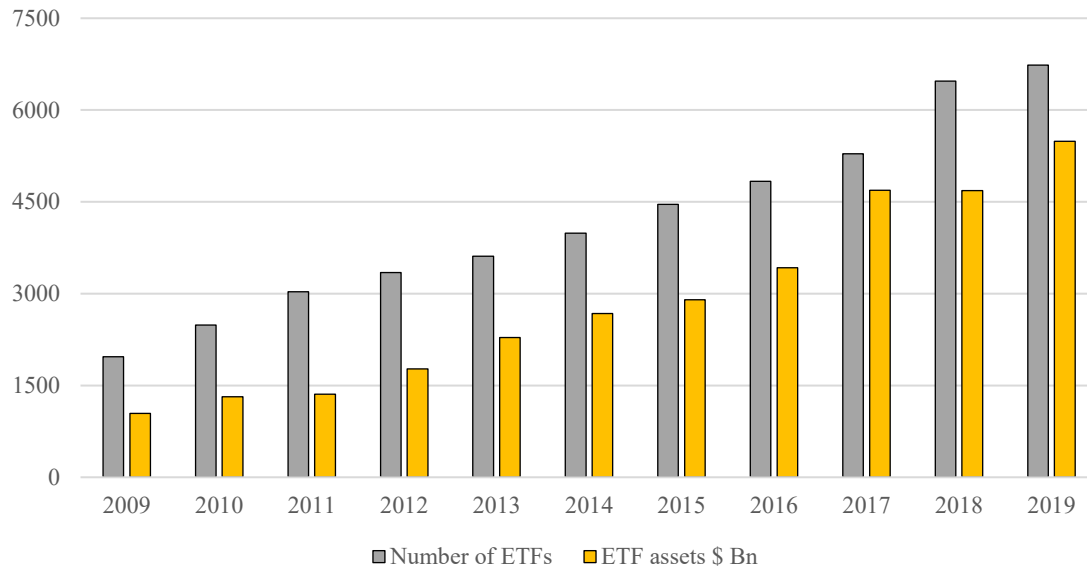
Graph 4. Total number of ETFs and ETF assets worldwide.

Furthermore, ETFs beat the traditional mutual funds at the fund expense ratios. Overall, decreasing fund expenses continues, illustrated by Morningstar's study (Pionline 2019), wherein 2018, the cheapest 20% of the funds saw larger net inflows than the remaining 80% together. The average ETF expense ratio was just 0.23% in 2016, and for comparison, the expense ratio for index-based mutual funds was 0.73% and 1.45% for actively managed mutual funds. As the mutual fund load, or the broker's commission, is absent in ETFs, expense ratios fall in favor of them. However, as the ETFs are traded in the secondary markets, there are additional commissions and transaction costs in ETF trading (Fidelity 2011).

As ETF products have become more and more popular, the supply of different fund styles has increased significantly. Especially factor investing with ETFs has gained popularity during recent years. As education about the benefits of passive investing increases, factor investing has become a useful intermediate between passive and active investing. Recent ETFs are attempting to capture the returns of fundamental factors observed by academia, such as value, size, market sentiment such as momentum or ESG factors.

As most of the other fundamental factors are well researched, ESG is a relatively new branch of categorization. As sustainable investing increases its popularity, naturally, new financial innovations and resources originate to satisfy the investor interest. UBS Research (2019) found out that the amount of completely ESG denialist asset owners have shrunk to just 5% in their 600 owner, $21 trillion asset sample. In the theme of the mentioned increase of passive investing, ESG ETFs have similarly gained a foothold in the financial industry. From the year 2016, the assets invested in the ESG ETF universe has expanded by over 400% to over 52 billion dollars (Nasdaq 2019). The rapid expansion of these thematical funds offers excellent motivation for including a sample of these funds in this paper.

2.3.1. ETF Creation-Redemption process

Behind every ETF product, the Creation-Redemption process illustrates the process happening in the primary market. This process involves several parties, from which the most important is the Authorized Participant (AP). AP's responsibility is to keep the ETF's market price as close to the ETF NAV as possible. However, due to this pricing mechanism's balancing nature, arbitrage opportunities in the ETF pricing appear. Large volumes and market fluctuations in the ETF supply and demand tend to make the ETF prices diverge from its NAV, which creates an arbitrage opportunity for AP to exploit (Ferri & Phillips 2009; Gastineau 2010). This continuous process creates constant additional price pressure for stocks that are included in any high-flow ETF. Thus, understanding the key process behind the ETF induced market activity is necessary for any study involving these financial instruments.

The creation process of a traditional open-ended stock ETF would be as follows. First, an ETF provider decides the ETF contents and benchmark index it follows and contacts an AP. The AP then acquires the necessary securities from the market by the replicated index weights. Finally, the AP delivers these securities to the ETF provider, and in exchange, the provider delivers the AP an ETF creation unit that represents blocks of equal value ETF shares. Usually, a creation unit consists of 50 000 to 250 000 individual ETF shares. The exchange happens on a fair-value basis, as the delivered underlying shares represent exactly the ETF share NAV, not the ETF secondary market value. (Gastineau 2010.)
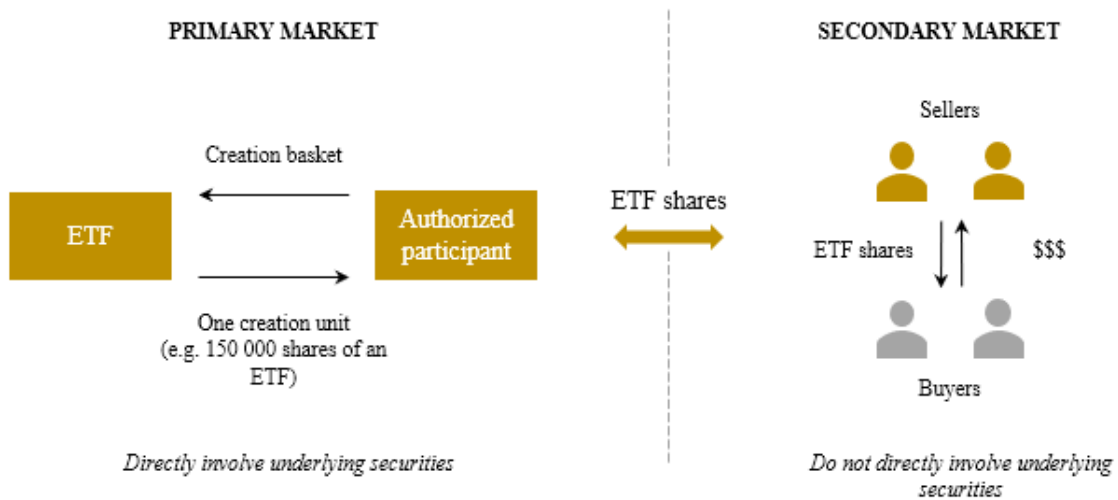
Figure 1. The ETF Creation-Redemption Process illustrated (Investment Company Institute 2020).

This process can be reversed and works as ETF share redemption. Here the AP can remove ETF shares from the market by submitting a full creation unit to the ETF provider. ETF provider in exchange gives AP the same value in original ETF underlying securities, in kind.

The ETF creation-redemption process and its effects on ETF constituents and ETF price efficiency has been studied widely. (Da & Shive, 2018) investigate the effects of ETF arbitrage activity to its underlying portfolio of securities and find that the arbitrage activity can transfer the non-fundamental shocks forward to the constituents and thus, increase co-movement between stocks. Additionally, (Petajisto 2017) finds economically significant mispricing between ETF market prices and their NAVs and creates a systematic trading strategy to exploit this effect.

2.3.2. The ETF arbitrage mechanism

As an AP's role is to keep the ETF market price matching its NAV, the question of whether the current ETF market price is over, under, or strictly at its NAV arises. As ETF creation-redemption process allows APs to create or redeem ETF shares constantly, the discrepancies between ETF market price and its NAV tend to be mostly moderate. Even though the creation-redemption ensures the relative efficiency of ETF market prices, it does not guarantee absolute efficiency between market price and the NAV. The ETF market price deviation from its NAV can be stated as the ETF premium or discount.

Meaning ETF trades at a premium if its market value is over its NAV and vice versa. (Ferri & Phillips 2009) In short, ETF creation-redemption is an arbitrage process conducted by APs to balance between ETF's NAV and market price.

As the markets can be considered semi-efficient by EMH, premium or discount in ETF prices should not be possible for extended periods or by great magnitude, as the arbitrageurs would immediately eliminate these opportunities from the markets. To capture the riskless profit, an arbitrageur could create overvalued (premium) ETFs by purchasing the underlying basket of securities from the market and creating the ETF unit. Currently, this process is controlled by APs, which capture the small arbitrage profits from the market. The competition between APs has been said to control the arbitrage profits, but recent studies have suggested that even with APs controlling the arbitrage, there are economically significant mispricing in some ETF classes that could be exploited by active trading strategies. (Petajisto 2017).

Moreover, even though the ETF pricing mechanism contains an arbitrage process, there can be additional limits to arbitrage. As (Shleifer & Vishny 1997) note, true arbitrage should not require any starting capital and contain any risk. This is often not the case in the real-world, and authors further discover that arbitrageurs tend to avoid more volatile positions that increase the fundamental risk. Additionally, arbitrage opportunities can often be appropriately exploited only by a handful of sophisticated investors; thus it is not an all market-wide phenomenon. As the real-life arbitrage is risky, it is debatable whether the relative value bets can be considered true arbitrage.

As the co-movement between the ETF constituent stocks is one of the key interests in this paper, it is necessary to briefly inspect the different replication methods for ETFs. In short, three main categories of ETF replication can be differentiated. The first method, the Physical, or full replication is the most theoretically simple. The underlying index is fully replicated by acquiring securities from the market by the exact index weights. Restrictions to the full replication method come mainly from possible rounding errors from fractional index weights for individual stocks. (Lyxor 2019).

The second method, the Sampling method, is often used when the replicated index consists of a large variety of different markets or illiquid securities. In these instances, full physical replication would be too expensive. In the sampling method, replication is optimized by quantitative models so that only the most liquid stocks are used for the replication, and thus transaction costs are minimized. In other words, instead of fully

replicating the index, only a representative subsample is needed that still delivers similar returns to the actual index. (Lyxor 2019).

The third method, Synthetic replication, uses derivatives to replicate the index. More specifically, a total return swap that delivers the index return. For the trade, ETF holds liquid securities and delivers the basket's profit to the swap counterparty. The Synthetic method can often replicate certain indexes such as commodities and money markets more efficiently than through the Physical or Sampling methods. However, when using derivatives, the risk that the swap deal's counterparty might go bankrupt and thus fail to deliver the return increases. This risk is called the counterparty risk. (ESRB 2019; Lyxor 2019)

To compare the efficiency of ETF replication methods between each other, a useful measure of tracking error is introduced. Tracking error reflects the discrepancy between the ETF NAV return and the return of its underlying index. The chosen replication method naturally affects tracking error. In theory, full physical replication should provide NAV return of zero to minimal error, compared to the sampling method that optimizes the holdings, thus not reflecting the underlying index perfectly but near enough to provide a similar return.

Piccotti (2018) measures the NAV tracking error as the standard deviation of the daily NAV return minus the underlying index's return. The tracking error of ETF j in the time t can be stated as follows:

(1) $$Tracking\ Error_{j,t} = \left(r_{j,t}^{NAV} - r_{i,t}^{INDEX}\right) * 100$$

where: $r_{j,t}^{NAV}$ = The daily arithmetic NAV return.
$r_{i,t}^{INDEX}$ = The daily arithmetic return of the underlying index.

Following this, the standard deviation for the obtained tracking error is calculated to obtain a comparable measurement of the relative deviation of the ETF return and index return. Furthermore, as the replication method has an impact on the ETF tracking error, Piccotti (2018) suggested that more segmented markets with higher entry barriers and higher illiquidity affect the ETF's ability to track its underlying index. Thus, region-based ETFs on emerging or frontier markets often suffer from higher tracking errors, similar to regional funds in more developed markets.

As the market segmentation is one factor for tracking errors, index constituent illiquidity has been suggested by several authors to be a magnifying factor for tracking error (Loviscek, Tang & Xu 2014; Ferri & Phillips 2009; Piccotti 2018). However, the market illiquidity is often a consequence of the segmentation in the target markets, in the form of increased trading costs or limited market depth.

For the replication methods, the synthetic replication ensures the smallest tracking error of the three. As trading frictions such as liquidity, bid-ask spread, and the optimization in sampling method all magnify the tracking error, the same return is guaranteed by the counterparty in synthetic replication. However, as mentioned previously, counterparty risk in swap-based replication exists. Additionally, there has been a regulatory concern for the swap-provider interconnectedness, which is a threat to financial stability in the event of financial crisis. (Vanguard 2013).

In the context of this paper, only funds with full physical replication are used, as replicating the underlying ETF portfolio created by the synthetic method is impossible for the majority of investors. Additionally, the sampling method is not relevant for the funds selected in this paper, as it is often used in more illiquid markets or asset classes, which are not in the scope of this study.

2.4. Stock co-movement

The significance of the stock correlations and co-movement has been studied widely. Barberis et al. (2005) investigate the common factors driving the returns of certain groups of stocks together. Logically, stock returns within similar industries or with similar fundamental attributes, such as value stocks and small-cap stocks, should be initially correlated to some extent. However, by assuming a frictionless economy with rational agents, the movement of stock prices should be attributed to its objectively forecasted cash flows and risk. This fundamental view of the stock co-movement explains some of the correlations but still leaves room for further investigation.

The term co-movement in the financial context describes a strong correlation between assets. The correlation measures the statistical relationship between two variables. Pearson's linear correlation coefficient measures the linear association of the inspected variables as follows:

(2)
$$\rho_{xy} = \frac{cov(x,y)}{\sigma_x \sigma_y}$$

Where $cov(x,y)$ is the covariance between x and y and σ_x and σ_y are the standard deviations of x and y.

The correlation coefficient is always between [-1,1], and the closer the coefficient is the limits, the stronger the linear relationship is. For the financial markets, beta coefficients are denoted by:

(3)
$$\beta = \rho_{xy} \frac{\sigma_x}{\sigma_y}$$

The correlation has thus a direct effect on the measurement of stock beta. It should be noted here that the beta typically measures the stock sensitivity with the market index, and thus increased stock-stock correlations should not affect the standard beta coefficient here. However, for managing equity portfolios, correlations, and beta coefficients between stocks provide information about diversification in the portfolio. Furthermore, as during high volatility, stock correlations tend to increase (Connolly, Stivers & Sun 2005), which further complicates and lessens the diversification benefits of pure equity portfolios.

Barberis & Shleifer (2003) argued that investors tend to categorize assets before forming portfolios. These categories would include sorting by, for instance, market cap, industry, and profitability. When several investors use these same categorizations and their collective sentiment changes, co-movement on even seemingly unrelated assets can be experienced, even though the fundamental aspects of these companies might be uncorrelated. Naturally, this raises a question of whether the increasing amount of indexing and new categorizations created by market demand, such as Quality or ESG themed indexes, further increase correlations between the constituent stocks (Index Industry Association 2019).

By forming these thematic indexes, index providers combine several different stocks with some predetermined common attributes to a basket that tracks the performance of each constituent as a single unit. As ETF providers create funds from these indexes, the ETF trading activity begins to affect the constituents' trading activity as well, since the ETF share creation and redemption process relies on buying and selling constituents' shares

on the market. This increases the correlation of trading volumes and co-movements between constituents (Greenwood & Sosner 2007).

As the number of ETFs and institutional ownership in stocks has increased, Staer & Sottile (2018) investigated whether the equivalent volume or the indirect trading caused by ETF creation-redemption arbitrage mechanism is a useful indicator for analyzing the co-movement between the ETF constituent stocks. The equivalent volume measures the ratio of weighted trading volumes of the inspected stock and its parent ETF. Authors found out that ETF trading volumes comprise up to 70% of inspected stock's volume. To further illustrate the significance, Apple Inc. (AAPL) has exposure to 323 different ETFs, and it is within the top 15 holdings for 179 of them. (ETF database 2019). Thus, the increasing ETF fund provider ownership in stocks will be an exciting topic to study further in the future.

Authors came to a similar conclusion than Wurgler (2010), with his paper examining implications for growing index-linked investing. There is substantial proof that index-linked investing affects fundamental prices of stocks and might distort them. However, the possible implications are currently difficult to measure, and scenarios where invested capital would unexpectedly mass shift from place to another, are hypothetical. Finally, Connolly et al. (2005) investigated the relationship between volatility and stock-bond and stock-stock correlations. The results indicated that higher levels of volatility, proxied by the VIX index, increased the stock-stock correlations. Additionally, more considerable absolute changes in VIX resulted in heightened correlations.

From a macroeconomic point of view, stock correlations on the international level have also been studied extensively. Solnik, Boucrelle & Yann (1996) find that international correlations increase when the overall market volatility is high. The correlations for stocks and bonds tend to fluctuate over time but are excessive during high market volatility times. Especially, the French and German markets began to correlate more after their leading roles in the EU. Thus, additional macroeconomic factors can link even different countries' stock markets to some extent with each other.

2.5. Portfolio evaluation metrics

As the essential activity in this thesis is to create portfolios continuously by the frequency of inspected events, the question how to evaluate the performance of these strategies

arises. In this section, the key metrics for portfolio evaluation in this paper are introduced. The most well-known and still broadly used pricing model, The Capital Asset Pricing Model (CAPM), implies the existence of a diversified market portfolio that all the investors should hold in combination with risk-free assets according to the risk appetite of the individual.

As CAPM attempts to portray the functioning of the real world, some pre-existing assumptions of the conditions surrounding the model must be made. These assumptions are presented in Figure 2 next:

THE ASSUMPTIONS OF THE CAPM

---

1. INDIVIDUAL BEHAVIOR

    a. Investors are rational, mean-variance optimizers.

    b. Their planning horizon is a single period.

    c. Investors have homogenous expectations (identical input lists).

2. MARKET STRUCTURE

    a. All assets are publicly held and trade on public exchanges, short positions are allowed, and investors can borrow or lend at a commong risk-free rate.

    b. All information is publicly available.

    c. No taxes.

    d. No transaction costs.

Figure 2. The Assumptions of CAPM (Bodie et al. 2009).

As quickly seen, the assumptions of the CAPM generate a somewhat restricted outlook of the financial markets. Naturally, these assumptions are necessary to provide a unitary formula to explain stock returns, which does not have an infinite amount of exceptions. However, by even briefly inspecting Figure 2, most of the assumptions for both individual behavior and market structure are not present in the real financial markets. Despite this, CAPM provides an excellent framework to understand expected returns for the securities. The following formula describes the CAPM:

(4)
$$E(r_i) = r_f + \beta_i \left[E(r_m) - r_f \right]$$

where: $E(r_i)$ = Expected return of the security i
          $E(r_m)$ = Expected market return
          r_f = Risk-free rate
          β_i = Beta of the security

The deduced risk exposure or the beta, captured here, moves us to the next measure of investment performance. The Jensen's alpha uses the CAPM market portfolio and treats the intercept as "alpha" or the abnormal return over the theoretical expected return of the inspected portfolio.

$$(5) \qquad \alpha_p = r_p - \left[r_f + \beta_p (r_m - r_f) \right]$$

where: r_p = Portfolio return
          r_m = Market return
          r_f = Risk-free rate
          β_p = Beta of the portfolio

After the generally used beta and alpha, also well-known risk-adjusted performance measure, the Sharpe ratio is introduced. The Sharpe ratio evaluates the portfolio performance by its return and the standard deviation or volatility of the portfolio return. Thus, it forms a ratio of how much the portfolio produces excess return compared to the risk-free rate per unit of risk.

$$(6) \qquad SR = \frac{(r_p - r_f)}{\sigma_p}$$

where: r_p = Portfolio return
          r_f = Risk-free rate
          σ_p = Standard deviation of the portfolio

As CAPM and its by-products have got their share of critique during their long history, it is necessary to visit more current methods used in this paper. Fama & French (1993) satisfy non-systematic risk sources by accounting for three identified risk factors. Following the model's popularity, authors expanded the model further to five relevant risk factors (Fama & French 2015).

The first three risk factors identified are the market excess return MKT, which is the value-weighted return of all U.S. CRSP (The Center for Research in Security Prices) firms listed in a major US stock exchange, minus the US one-month t-bill rate. SMB factor is the average return of nine small equity portfolios minus the average return of nine big equity portfolios:

$$(7) \quad SMB_{(B/M)} = 1/3(Small\ Value + Small\ Neutral + Small\ Growth) \\ - 1/3(Big\ Value + Big\ Neutral + Big\ Growht)$$

$$(8) \quad SMB_{(OP)} = 1/3(Small\ Robust + Small\ Neutral + Small\ Weak) \\ - 1/3(Big\ Robust + Big\ Neutral + Big\ Weak)$$

$$(9) \quad SMB_{(INV)} = 1/3(Small\ Conservative + Small\ Neutral \\ + Small\ Aggressive) - 1/3(Big\ Conservative \\ + Big\ Neutral + Big\ Aggressive)$$

$$(10) \quad SMB = 1/3(SMB_{(B/M)} + SMB_{(OP)} + SMB_{(INV)})$$

The HML factor is the return spread of high book-to-market minus low book-to-market stocks:

$$(11) \quad HML = 1/2(Small\ Value + Big\ Value) - 1/2(Small\ Growth \\ + Big\ Growth)$$

Additionally, for the five-factor model, the RMW factor describes the return spread of robust operating profitability minus weak operating profitability stocks:

$$(12) \quad RMW = 1/2(Small\ Robust + Big\ Robust) - 1/2(Small\ Weak \\ + Big\ Weak)$$

Furthermore, the CMA factor is the average return of stocks making investments conservatively minus the stocks making investments aggressively:

(13)
$$CMA = 1/2(Small\ Conservative + Big\ Conservative)$$
$$- 1/2(Small\ Aggressive + Big\ Aggressive)$$

By combining the five estimated risk factors to a single factor model, we get the following Fama French 5-factor model (FF5):

(14)
$$r_p - r_f = \alpha_p + \beta_{p,MKT}(r_m - r_f) + \beta_{p,SMB}SMB + \beta_{p,HML}HML$$
$$+ \beta_{p,RMW}RMW + \beta_{p,CMA}CMA + \varepsilon_i$$

where: r_p = Portfolio return
r_m = Market return
r_f = Risk-free rate
α_p = 5-factor alpha
$\beta_{p,MKT}$ = Sensitivity to the Market excess return
$\beta_{p,SMB}$ = Sensitivity to the SMB factor
$\beta_{p,HML}$ = Sensitivity to the HML factor
$\beta_{p,RMW}$ = Sensitivity to the RMW factor
$\beta_{p,CMA}$ = Sensitivity to the CMA factor
ε_i = Error term

The firm character variables, or the risk factors, were selected by the history of observations that these factors predict deviations from the average returns consistent with the CAPM, and by not accounting for them, all by-products of CAPM should be unreliable as the results contain these identified risk factors.

Finally, as this paper is motivated to provide solutions to the issue mentioned by Lynch et al. (2019) regarding the lack of fundamental analysis in the portfolio creation phase, in this thesis, Piotroski F-Score (Piotroski 2000) is used as a fundamental proxy to mimic fundamental analysis that separates fundamentally good firms from bad firms and provides easy, systematic alternative for fundamental analysis.

F-Score uses accounting-based historical company information and conveys this information as a score between 0 and 9. The score is totaled from succeeding in nine categories that award points by either 1 or 0. Each outsider stock gets their score from the most recent annual report published from the time of the portfolio formation. Thus, as with ratio analysis in general, there can be significant staleness with the relevancy of the

F-Score, if the portfolio formation date is close to the next annual report's release date. For the F-Score measurement, three major groups can be noted: profitability measures, leverage and liquidity measures, and operating efficiency.

Under these three groups, nine criteria are introduced. The different criteria and their groups are introduced in the table below:

	PROFITABILITY	success (1)	failure (0)
1	RETURN ON ASSETS	*If positive in current year*	*If negative*
2	OPERATING CASH FLOW	*If positive in current year*	*If negative*
3	CHANGE IN RETURN OF ASSETS	*If ROA higher in current year than last year*	*If lower*
4	ACCRUALS	*If OPERATING CASH FLOW/TOTAL ASSETS is higher than ROA in current year*	*If lower*
	LEVERAGE AND LIQUIDITY	**success (1)**	**failure (0)**
5	CHANGE IN LEVERAGE	*If ratio lower in current year than last year*	*If higher*
6	CHANGE IN CURRENT RATIO	*If ratio higher in current year than last year*	*If lower*
7	CHANGE IN NUMBER OF SHARES	*If no new shares were issued last year*	*If new shares were issued*
	OPERATING EFFICIENCY	**success (1)**	**failure (0)**
8	CHANGE IN GROSS MARGIN	*If margin higher in current year than last year*	*If margin lower*
9	CHANGE IN ASSET TURNOVER	*If ratio higher in current year than last year*	*If ratio lower*

Figure 3. The criteria for the Piotroski F-Score measurement.

As stated, the total F-Score is the sum of all successfully passed criteria for the company. Thus, the fundamentally best companies should possess an F-Score of 9 and the worst 0. Naturally, making fundamental analysis based on a single proxy can rarely be completely interchangeable with discretionary analysis. However, for this study, it should be sufficient to show that applying even some fundamental analysis for the case companies could improve the returns.

3. LITERATURE REVIEW

The amount of research discussing both the impact of ETFs on the financial markets and passive indexing has expanded rapidly. As the amount of U.S. passive equity assets passed the amount of active assets the first time in August 2019 (Bloomberg 2019a), it is expected that the trend of passive investing and simultaneously, the amount of research provided on the subject, will further increase. This chapter will proceed as follows: first, the research on passive investing in general and qualities and reasons for stock correlations. Secondly, studies investigating ETFs and their effects on the efficiency on the financial markets and their underlying stocks are presented. Finally, studies covering the low stock beta anomaly are presented, as this paper takes a very similar contrarian approach by sorting ETF constituents by their betas-to-ETFs.

3.1. Previous research on passive investing

The effect of indexing on the efficiency of stock prices has been studied widely. Already in 1986, Harris & Gurel (1986) examine the effect of stock inclusions and exclusions from the S&P 500 index. By the teachings of the EMH, stock prices should not be affected even from trading large blocks of stocks simultaneously if the seller does not possess private information about the stock. Thus, the stock prices should be elastic, and the sale itself should not affect the prices.

The changes in the composition of the S&P 500 index often induce large-scale purchases for the new constituents. The selection criteria for the index is based on a well-known criterion which should not reveal any new information about the future returns for the stocks selected in the index. In other words, no significant price impact should be expected from this event. However, the authors find excessive shifts in volume and a statistically significant increase of 3% in new constituents' prices. To explain this effect, the authors introduce a price-pressure hypothesis that suggests that the investors who provide liquidity in the steep demand shifts must be compensated for the costs and risks they bear in these fast-paced transactions.

Following the S&P 500 constituent changes theme, Vijh (1994) finds that the betas of stocks included in the S&P 500 index tend to be overstated, and for the stocks not included in the index, betas are understated. The author suggests that the increases in betas and

decreases in the positive autocorrelations the reason for the trading volume increases of the average S&P 500 stocks. The price pressure, or the excess volatility, is the price that needs to be paid to trade a large number of portfolio constituents with limited liquidity.

The issue linked with the inexplainable excessive stock correlations, or co-movements, is discussed by Barberis et al. (2005), who introduce three different views for stock co-movement. The first, "fundamentals" view can be considered the traditional position, where the correlations between different stocks can be explained by their fundamental attributes, such as cash-flows or discount rates. The second, "category-based" view, is an effect from investors grouping different securities, i.e., including stocks to a specific index, which increases their correlations as the inclusions and exclusions are executed systematically by constant measures.

The third view, "habitat-based" co-movement, is a product of investors systematically restricting themselves from trading individual stocks, which again increases the co-movement between these stocks. As the traditional view of stock correlation, the "fundamentals" view has been challenged by anomalies such as Siamese-twin stocks, other explanations for stock co-movement are required. The "category-based" view can be used as an explanation of why certain groups such as small-cap stocks, industry-related stocks, or bonds tend to correlate, even though their cash flows are mostly uncorrelated.

As investors tend to categorize and subconsciously form classifications for stocks, arbitrary links between stocks are created. Simplified investment process and systematic asset allocation rules of portfolio managers make the index constituent stocks even further correlated, which may create investing opportunities during category-based selloffs. In other words, even just the inclusion of a stock to an index may affect its returns, even though these returns might not be fundamentally validated. (Barberis et al. 2005).

In contrast, Chen, Singal and Whitelaw (2016) show that with robust, univariate regressions, the excess co-movement from Barberis et al. (2005) study almost entirely disappears. As regressions used to capture excessive co-movement are bivariate, before and after the inspected event, authors show that coefficients captured by the bivariate models are extremely sensitive to small changes in parameters. Thus, it is questionable whether one can interpret these coefficients as they might provide no new information in economic context.

Authors find that momentum has a significant effect on the beta patterns in the Barberis et al. (2005) study, which the authors did not control for. Additionally, using Dimson (1979) coefficient adjustment with leads and lags explains more than 50% of the effect from the original study, in which the adjustment accounted for only one third of the effect. The authors conclude that excessive co-movement is most likely due to beta changes for momentum winner stocks.

Macro factors and global events have also been studied as the source of excessive co-movements. Bekaert, Ehrmann and Fratzscher (2014) investigate the contagion of the 2007 – 2009 financial crisis to 415 different country and industry portfolios. Authors find evidence of contagion by estimating with factor model with global and domestic factors, but the model only explains 75% of the total return variation in the study. Authors also note that even though the 2007 – 2009 financial crisis originated from the United States, they find weak evidence of the contagion from US markets to global equity markets. Instead, they find a contagion link between the domestic markets and the individual domestic portfolios. Thus, the investors in the global crisis of 2007 – 2009 were more inclined to punish markets with weak economic fundamentals and poor sovereign ratings.

On the other hand, Parsley & Popper (2020) apply a simple gauge of return co-movement that decomposes the within-market market returns variance. The authors find that return co-movements are not necessarily tied to country-specific factors such as country risk or investor protections, as the Bekaert et al. (2014) suggest, albeit in the context of the financial crisis of 2007 – 2009. The return co-movements are instead related to shorter-term variables that portray international macroeconomic policy stability and suggest future firm-level research for properly linking return co-movement with firm-level variables such as foreign ownership, corporate culture, or corporate structure.

Returning to firm-level research, Coles et al. (2020) study the effect of index investing on single-stock prices. Authors demonstrate that index investing introduces noise into stock prices but does not directly affect long-run price efficiency. Stocks with higher proportion of index ownership have higher correlations with index movements and deviate more from random walk. Authors further show that index re-balancing causes significant changes in the company ownership composition and increased passive ownership effects in higher volatility. By investigating the set of stocks after assignment to the Russel 2000 index, the authors conclude that index inclusion has an immediate effect of increased correlations with other index members and increased volatility. Thus, the index inclusion changes the stock's price process.

The trend of proportionally increasing fund provider or institutional investor ownership in listed companies is especially intriguing, as the trend is relatively new and continuously strengthening. Some critique for the common ownership between competitors has been published by Azar, Schmalz and Tecu (2018), who investigate the U.S. airline industry and its concentrated ownership base. The paper inspects the common ownership mainly from an antitrust perspective, but the reduced competitiveness between the firms that the authors found raises the question of whether the common ownership has implications for the stock price processes.

By inspecting further, the index inclusions and firms' reactions to these events Yu (2008) finds that the stock's inclusion to S&P 500 index is an enhancing factor to combat earnings management due to the increased analyst coverage the inclusion brings. In other words, a higher number of analysts covering the firms ensures a lower amount of earnings management by the firms. The increased informational flow decreases the asymmetric information present and has positive implications for corporate governance as well.

Greenwood & Sosner (2007) investigate a unique sample of broad Nikkei 225 index redefinition in April 2000. A total of 30 stock additions and deletions happened simultaneously, and the authors found proof of excess index-linked co-movement from the subsequent stock returns. For the added stocks, correlations of trading volumes with Nikkei 225 constituent stocks increased, while for the deletions, they decreased, thus implying that added stocks became exposed to the trading shocks experienced by other members.

Moreover, Petajisto (2011) finds that the index additions and deletions have a significant economically material impact on the stocks included in or excluded from the index. By inspecting S&P 500 and Russel 2000 additions, the author finds a price impact of +8.8% and +4.7%, respectively. In contrast, exclusion announcements yielded -15.1% and -4.6% price effect on the event day. As the index inclusions in the investigated indexes are driven mainly by firm market cap and industry representation, the firm fundamentals are not drastically changing due to the index inclusion, which would explain the sudden event price shock. Additionally, the found effect on the stock prices is gradual, instead of being fully incorporated to the prices immediately after the announcement.

Finally, Belasco, Finke and Nanigian (2012) investigate the impact of passive investing in the corporate valuations in the S&P 500 index and find that the constituent valuations

increased by up to 167 bps relative to non-constituent stocks due to S&P 500 fund money flow. The impact additionally persists after the flow occurs and does not dissipate immediately. The paper's findings implicate that the preference shift towards passive index investing reduces the informational efficiency of stock prices, and these discrepancies are not combatted efficiently enough by arbitrageurs.

Overall, the body of literature for passive and index investing is vast and convincing. Index constituent changes have been shown to affect the price processes of index stocks in both inclusions and exclusion events Petajisto (2011). The fundamental properties of the stocks cannot explain these price-process changes, nor the information that the inclusion event generates (Barberis et al. 2005; Harris & Gurel 1986). However, the indexing has been studied to provide better price efficiency due to transformed ownership composition after index inclusion (Coles et al. 2020). Thus, while indexing has been shown to generate noise in the stock prices, it also increases efficiency by the heightened informational flow, corporate governance (Yu 2008), and steadier ownership base with large institutional investors.

3.2. Previous research on ETFs

Todorov (2019) studies the ETF induced price impact on the commodity and volatility asset classes. The author replicates and decomposes the value of the VIX futures contracts, extracts the non-fundamental component from the prices, and captures a return of 18.5% that is strongly related to the ETF rebalancing. Thus, the study results show that the passive (ETF) funds actively affect the prices of their underlying assets, commodities, and VIX in this instance. As the constant innovation for new ETFs to provide exposure to less traditional, alternative asset classes is continuous, these new investment vehicles tend to stay relatively under-researched. Thus, it is difficult to fully assess whether there are significant negative consequences of this constant trend of passive indexing.

Da & Shive (2018) provide evidence on the co-movement effect caused by arbitrageurs for ETFs. By inspecting a sample of U.S. equity ETF holdings, authors find a link between ETF activity and return co-movement on stock level. ETF arbitrage, i.e. creation-redemption process, allows the APs to balance ETF market prices for the trade for arbitrage profits. As discussed in the theoretical background section, APs can buy discount ETF shares, redeem the assets, and sell them to the market by the ETF's NAV price, thus making immediate profit.

Authors discover that if a stock has high total ETF ownership, it co-moves more with the market. The effect of ETF holding in stock is more substantial than the effect of mutual fund or other institutional holding, resulting 0.03 increase in beta for a 1% market cap increase in ETF ownership. Additionally, ETF turnover has a link for stock co-movement with the market, as the effect is augmented for small stocks with low turnover. The underlying question is whether the link between ETF activity and stock co-movement implies better incorporation of information or excessive noise from non-fundamental demand created by the ETF arbitrage mechanism.

The authors find negative autocorrelation in ETF daily returns, which is magnified if the ETF turnover is high. The findings support the assumption that ETFs may cause excessive co-movement by the non-fundamental shocks the ETF arbitrage causes. As the stocks with higher ETF activity have more often negative betas to lagged market returns, it is suggested that ETF activity affects excess co-movement.

Further inspecting the effects of ETF arbitrage on the underlying constituent prices Shim (2019) finds that increased ETF trading after 2008 has led to higher arbitrage co-movement. The author introduces the concept of arbitrage sensitivity, which combines the sensitivity of price impact and portfolio weight. The market price balancing creation-redemption process practiced by arbitrageurs spills over to constituent stocks and induces price pressure. As the arbitrage is executed by trading constituent stocks by their index weights and further trading for ETF shares (or vice versa), constituents are moving based on the arbitrage process, not fundamental demand.

Additionally, the author investigates whether ETF stock betas or the return of underlying constituent stock regressed against the ETF return. The relationship between arbitrage sensitivity and ETF betas are found to be positive, especially during higher trading volume periods from 2008 onwards. Changes in trading volume match the arbitrage sensitivity's ability to forecast ETF betas. To summarize, arbitrage co-movement is shown to impact underlying constituent prices, and as the stock covariances have significant implications for risk exposures, prevalent ETF arbitrage at large trading volumes might distort the perceived risk for the stock.

Ben-David et al. (2018) investigated the relationship between high ETF ownership and the underlying stocks' volatilities. The authors find a significant link between ETF ownership, higher volatility, and negative autocorrelation. The average ETF ownership in S&P 500 stocks has increased from 0.14% in 2000 to 7.05% in 2015. Thus, this

significant portion is passively owned by fund providers and depends entirely on the logic of each ETF's index-weighting methods. As the trend of passive investing and simultaneously higher ETF ownership in stocks keeps increasing, further studies on the effects of ownership commonality should be produced.

Petajisto (2017) inspects ETFs' pricing efficiency by analyzing the creation-redemption process and the amount of ETF premiums and discounts in the market. As the popularity of ETFs has increased heavily over recent years, problems with their relative efficiency should be highlighted. The author investigates a broad cross-section of ETFs and shows that even though the average ETF premium is only 6bps, in certain ETF classes such as U.S. municipal and high-yield bonds, international equity and international bond funds show significant average mispricing ranging from -13bps to 31bps.

The author further illustrates the significance of the ETF premiums and discounts by creating a trading strategy exploiting these mispricings. For individual investors, fund NAV should be used as a convenient measure to check whether the ETF inspected is currently cheap or expensive. Additionally, the cost-conscious investor should be aware that for some ETF classes, additional expenses can be caused in the form of relative mispricing between the NAV and market price. Overall, most of the ETFs are relatively efficiently priced and show minimal average premiums or discounts.

A redeeming factor for index-linked co-movement is introduced by Glosten et al. (2016), who argue that ETF activity might increase short-run informational efficiency for stocks in weak informational segments such as in situations of underlying low liquidity or short-sale constraints. As with ETF trading, the creation-redemption process allows investors to trade a basket of underlying securities to ETF shares, and vice versa, this feature could prove itself useful in the times of short-selling constraints. As in the financial crisis of 2008, short selling was banned for 797 stocks, ETFs were exempted from the ban. This process could have been used to circumvent this constraint. Thus, investors looking to take short positions against specific companies could use ETFs to short these stocks indirectly.

Additionally, as Boone & White (2015) study the effect of institutional ownership, in this instance, ETF institution ownership found out that more extensive institutional ownership results in better corporate governance and more careful investor scrutiny. Glosten et al. (2016) suggest that by this finding, increased ETF ownership should increase the informational efficiency for index members that do not possess these traits beforehand.

The finding that stock's index-inclusion increases its informational properties is sensible and more studies should investigate this topic to find out whether the effect is magnified with broader and more well-known indexes.

To provide evidence for ETF activity and information incorporation, Glosten et al. (2016) study the effects of the ETF activity and Post-Earnings Announcement Drift (PEAD) anomaly. The PEAD reflects a situation where stock prices experience a positive trend after positive earnings announcements and negative after negative announcements. This anomaly shows that the fundamental information from the earnings is not incorporated to the stock prices immediately, but only partially in the time of the announcement. Thus, the stock price keeps drifting after the announcement in the direction of the surprise. The authors further show that high ETF activity is associated with reduced PEAD strategy returns.

Brown, Davies and Ringgenberg (2020) further investigate this non-fundamental demand of stocks. They use ETFs as a tool to identify the non-fundamental demand, as ETFs have become an increasingly popular category-based asset and they are generally passively managed, which removes the possible fund manager skill factor from the equation. The ETF fund flows are observed, as the ETFs arbitrage pricing mechanism, or creation-redemption process makes ETFs vulnerable to non-fundamental demand shocks. When the ETF premium arises, the mechanism correcting the mispricing activates, and new ETF shares are created, which generates ETF fund flows.

ETF flows provide unique measurement for non-fundamental demand as authors show that flows are mostly independent from fundamental demand. Compared to other studied fund flows such as hedge fund or mutual fund flows, which have been shown to contain important information about investor demands, ETF flows are shown not to possess these attributes. The authors further show that these fund flows can predict future constituent asset returns. Authors categorize the ETFs in terms of the highest premium and highest discount to decile portfolios that go long on the largest discount ETFs and shorts the highest premium ETFs. The returns on these portfolios further show the ETF return predictability by its fund flows.

Ernst (2020) theoretically models the constituent stock and parent ETF trading in tandem and shows that the strategy leads to stock price-discovery from the ETF. The effect is dependent on the stock weight in the ETF and the level of asymmetric information. Furthermore, crucial stock-specific information such as earnings dates or news about the

company leads to increases in single-stock-ETF simultaneous trades, which can comprise up to 2% of daily ETF volume. Thus, when the investors have substantial stock-specific information, they tend to trade both the ETF and the constituent.

Finally, Lynch et al. (2019) inspect opportunities created by ETF selloffs using a contrarian strategy of buying oversold ETF constituent stocks. Authors justify their strategy by illustrating case studies where certain ETFs have experienced sudden selloff due to announcement considering their covered industry. Authors argue that even if new information about a particular market segment is published, the new fundamental information rarely affects every stock in the index on a fundamental basis. As several studies have shown (Da & Shive 2018; Shim 2019), excess co-movement in index-linked constituent stocks, during high volume selloff stocks that should not fundamentally experience a price decrease, often move in tandem with the other stocks in the index.

This creates opportunities for active investors to benefit from the co-movement caused by passive investors. Authors suggest finding "outsider" constituents from ETFs to capture the non-fundamental aspects of selloffs. The main questions with ETF selloffs are to investigate why this specific ETF category is being sold and should all its constituents be sold off with it. The authors group these outsiders by merely measuring constituent betas to the ETF they are part of. The least sensitive i.e., lowest ETF beta stocks, are then bought after high volume selloffs to capture the co-movement effect from the index inclusion, which should not be warranted by fundamental reasons.

The results show that this trading strategy proved profitable for 11 of the 12 inspected ETFs when held for 40 days. Authors note that this strategy should prove useful for individual investors as a form of screening and then conducting fundamental analysis on the highlighted companies. I.e., does the selloff warrant the price drop for this investigated stock. Additionally, during high volume selloffs, correlations among the index constituents increase significantly, which cannot be rationalized by sole fundamental reasons.

Altogether, studies examining the ETF induced stock co-movements and inefficiencies retell mostly the findings of the effects of the index investing. As the ETFs follow primarily indexes generated for these products and likewise, this is expected to produce similar inefficiencies to stock prices than indexing itself. Thus, the ETFs are just convenient proxies to investigate the index-induced effects. Incredibly intriguing is that

as the amounts invested to ETFs grow day by day, more convincing and reliable studies can be produced about the effects of indexing to asset price efficiency.

3.3. Previous research on the low beta anomaly

Finally, as this paper focuses heavily on measuring ETF constituent betas, which are in several instances with broad-index ETFs equivalent to the traditional beta measurements for these stocks, it is necessary to visit the research on the low beta anomaly. In a sense, the low beta anomaly works as a convenient theoretical framework behind the motivations in this paper and provides insight into how and why measuring the stock beta has been an effective way to capture originally unintended returns.

The expected return-beta relationship set by CAPM is that only beta, or the perceived risk of the company, affects the returns of the stock. This essentially means that stocks with higher beta can generate higher returns than low beta stocks by taking more risk. Bodie et al. (2009). However, it has been exhaustively studied that the CAPM derived security market line (SML) is flatter than presented by the original theory (Black, Jensen and Scholes 1972). Thus, the unconstrained investors should overweight low beta and underweight high beta stocks to capture this discrepancy.

Frazzini & Pedersen (2014) create a market-neutral factor strategy of Betting Against Beta (BAB) that invests in low beta assets and shorts high beta assets to capture the return spread between low and high beta securities. This strategy and its methods will be inspected more closely at the end of the next chapter, but the overall idea and the results of the BAB factor are necessary to present here for reviewing purposes. The authors find the same relative return flatness first investigated by Black et al. (1972) in 18 of the 19 international equity markets inspected.

Additionally, the authors propose that constrained investors tend to hold higher betas. Mutual funds and individual investors are leverage constrained and actively try to combat that disadvantage by investing in riskier assets. Conversely, funds with leveraged buyout capabilities and Warren Buffet's Berkshire Hathaway buy stocks with betas below one on average, as these investors are taking advantage of the BAB strategy and profiting from it.

Another takeaway from the high-beta stock underperformance against the low-beta stocks is by Baker, Bradley and Wurgler (2011). The authors suggest that the anomalous relationship with the betas can be partly explained by the institutional investors' motivations to beat their fixed benchmarks, which discourages arbitrage activity. The high-beta and high-volatility stocks underperformed the low-beta and low-volatility stocks by significant difference when divided to top (high-beta, high-volatility) and bottom (low-beta and low-volatility) quintiles:
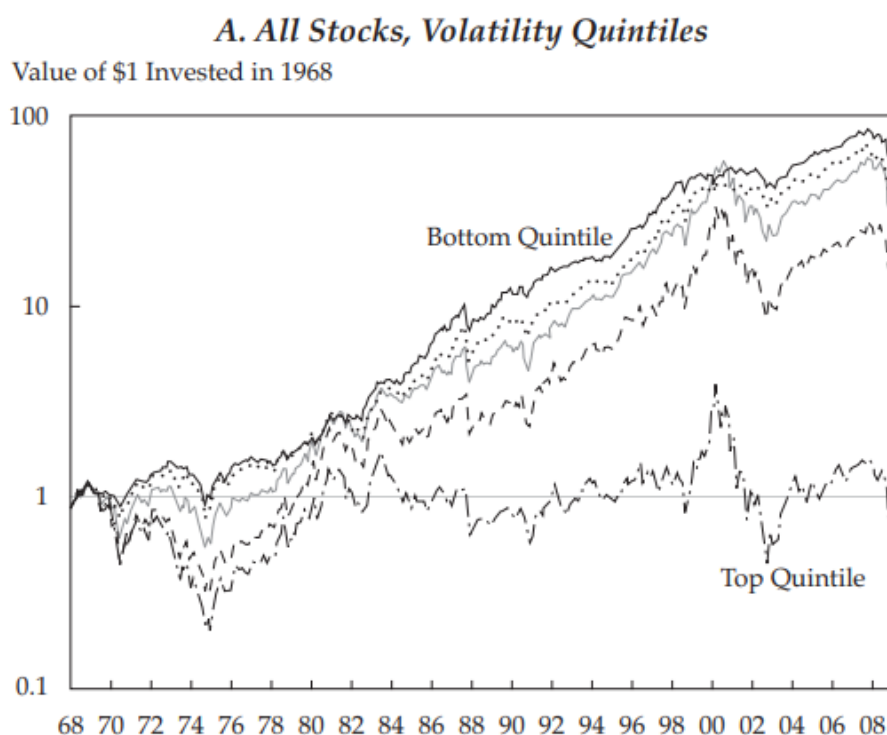
Graph 5. Results of the top and bottom beta + volatility quintiles (Baker et al. 2011).

By looking at a dollar invested in 1968 to both top and bottom quintile portfolios, the difference is staggering. One dollar invested to lowest beta and lowest volatility asset portfolio increased to $59.55, whereas the dollar invested to the top quintile portfolio was worth $0.58.

Following this finding, Christoffersen & Simutin (2017) further show that fund managers controlling large pension assets have higher exposures to high beta assets while keeping their tracking error low from their benchmarks. This finding implies that benchmarking may leave fund managers overly exposed to high-beta securities and leave low-beta

securities underweighted. This may further reinforce the beta related pricing anomalies and explain some of the negative alphas generated by high-beta stocks.

On the other hand, Cederburg & O'Doherty (2016) find that the conditional beta estimated for the high-minus-low beta portfolio negatively covaries with both equity premium and market volatility. This makes the generated alpha downwards biased and produces skewed results. The authors model the conditional market risk for the high-minus-low beta portfolios and further find that the conditional CAPM explains the betting against beta anomaly.

Authors provide five determinants of market exposures that may explain the beta-sorted portfolio dispersions from firm-level betas. These five characteristics are likely to forecast the shifts with the betas. The characteristics the authors found are economic considerations from the initial public offerings, heterogeneity of investment opportunities for the firms, heterogeneity in the used firm leverage, idiosyncratic risk, and economy-wide conditions for funding. Authors further analyze the relationship between the betas and these characteristics by regressing the time-series of portfolio betas against these five factors and showing that the firm investment opportunities, leverage, and idiosyncratic risk influence the high-beta portfolio betas.

Additionally, Hong & Sraer (2016) find that high-beta assets are prone to speculation and further overpricing. Due to investor disagreement of the high-beta assets payoff processes, these stocks experience more significant divergence of investor opinion. At the times when this aggregated disagreement is low, the estimated SML is upward sloping. With high disagreement, the SML flattens, and higher beta generates lower expected returns. The authors measure the market disagreement by calculating the standard deviation of the analyst forecasts for long-term EPS growth. When the deviation increases, the aggregate disagreement in the market increases.

Authors show that high-beta stocks are more speculative due to high sensitivity towards disagreement of their future cash flows. Because of this and the short sale constraints experienced by many investors, high-beta stocks are often overpriced compared to low-beta stocks. However, during low aggregate disagreement, the SML is upwards sloping as the CAPM suggests, but with high aggregate disagreement, the SML flattens and produces return patterns similar to (Black et al., 1972; Frazzini & Pedersen, 2014).

Finally, Schneider, Wagner & Zechner (2020) find that the low-risk anomalies are most often arisen by the co-skewness risk. The authors estimated that the option-implied

residual co-skewness factor explains most of the alphas generated by the BAB and betting against volatility factors. The authors explain that the investors demand compensation from the negative co-skewness linked to the asset's CAPM beta. As a result, the positive alphas generated by BAB and betting against volatility disappear when the co-skewness is controller for.

To summarize these bodies of research described above, several authors find passive indexing to have an effect on the prices of the underlying securities that they are following (Barberis et al. 2005; Coles et al. 2020; Greenwood & Sosner 2007; Petajisto 2011; Todorov 2019). Additionally, ETFs as a relatively new but increasingly popular asset class has similar but not yet very well quantified effects on the liquidity and price efficiency of their underlying assets (Brown et al. 2020; Da & Shive 2018; Glosten et al. 2016; Shim 2019). Finally, the low beta anomaly has been shown to yield abnormal returns in a wide cross-sectional context (Frazzini & Pedersen 2014). However, the freshness of the topic has generated opposing studies that have been able to explain some of the factors impacting the lower returns of the high-beta securities (Cederburg & O'Doherty 2016; Hong & Sraer 2016; Schneider et al. 2020).

4. DATA AND METHODOLOGY

This chapter introduces the collected data and methodologies used to create a systematic trading strategy utilizing ETF constituent stock co-movement after high fund trading volume events. Following the general description of the data collected and its properties compared to earlier studies, motivations for the sample restrictions and brief introductions of the ETFs used are presented. Finally, methodologies to capture abnormal volume days and trading strategies involving these events are introduced, and additional methods to assess the portfolio returns are presented.

4.1. Data description

As this paper primarily focuses on using systematic outsider beta-to-ETF stock strategy with broad-index and factor funds, most of the ETFs are chosen with as similar properties as possible compared to previous studies. Thus, the amount of assets under management (AUM) is one critical condition. As the funds Lynch et al. (2019) investigated in their paper were very large in terms of AUM and liquid in terms of trading volumes, funds in this paper have been chosen with similar factors, but with a focus on European broad-index and factor funds.

However, as the scope of this paper is to create outsider portfolios from fund holdings, some additional restrictions apply for the sample. Firstly, mainly one unitary region focused ETFs are inspected. This restriction is used to reduce the portfolio creation timing-based issues such as market holidays and price staleness induced from international diversification. As the price staleness distorts the measured ETF betas, most of the ETFs chosen in this paper are domestic. Other key restrictions were unreliable or lacking daily fund volume data for newer ETFs that were thus dismissed from the screening.

Secondly, creating portfolios from historical fund holdings requires fund providers to properly communicate the historical fund holding data to account for constant index rebalancing and reconstitution. By limitations of comprehensive database access and the lack of connections to fund providers resulted in the usage of only iShares provided ETFs. This is simply because iShares provides the most comprehensive historical fund holding data by far compared to other fund providers. In practice, this restriction should not have

any material effect on the results, as iShares, owned by BlackRock, is currently the largest ETF provider in the world and provides a broad range of funds spanning markets worldwide (iShares 2020b).

Thus, by performing these additional restrictions to prepare the sample for the next step in this paper, the total amount of ETFs that are selected forms to 8 funds. For every inspected ETF, the daily return and volume data are gathered to firstly capture the abnormal ETF volume days and secondly generate portfolios from outsider ETF beta stocks. The following ETFs are selected for further study:

DESCRIPTIVE STATISTICS OF THE SELECTED ETFs

ETF Symbol	Target country	ETF Name	Market Cap ($m)	Avg. daily volume*
IVV	USA	iShares Core S&P 500 ETF	$219 872	5 571 000
QUAL	USA	iShares MSCI USA Quality Factor ETF	$19 954	1 536 000
ISF	GBR	iShares Core FTSE 100 UCITS ETF	$10 122	7 055 000
EXS1	DEU	iShares Core DAX UCITS ETF	$6 953	484 000
SXRT	EUR	iShares Core EURO STOXX 50 UCITS ETF	$4 301	39 000
SUAS	USA	iShares MSCI USA SRI UCITS ETF	$2 765	294 000
IESG	EUR	iShares MSCI Europe SRI UCITS ETF	$1 629	3 000
IH2O**	INTL	iShares Global Water UCITS ETF	$1 028	10 000
			Median	Median
			$5 627	389 000

Notes: Market caps are denoted in ($m) millions of dollars, as of 08/2020. *Average daily volume during 08/2019 - 08/2020. **ETF contains globally diversified assets.

Table 1. Descriptive statistics of the selected ETFs.

By inspecting the descriptive statistics for ETFs selected for further study, the total market cap of the ETFs investigated is approximately $266 billion, with the median market cap of $5.6 billion. In Lynch et al. (2019) paper that investigated only U.S.-based sector SPDRs, the total market cap approximately $480 billion, and the median market cap of $15 billion. Even though the sample in this thesis is smaller in terms of AUMs, the difference can be explained by U.S.-markets still being way more massive than European markets. By comparing trading volumes, U.S.-based sector SPRDs are averaging 5 to 15

million as daily volume, which is an important factor to consider when discussing the comparability of the achieved results.

The motivation to inspect several ESG ETFs in addition to broad-index funds, even though their market caps and average trading volumes are on the lower end of the sample distribution, is the continually increasing relevancy of these instruments. ETFGI estimated the whole ESG ETF ecosystem to total over $67 billion in assets in February 2020. Additionally, BlackRock predicts that its European ESG ETF assets will grow to $250 billion by 2028. (Markets Media 2020).

By briefly focusing on the selected funds, IVV is equivalent to the SPY inspected in Lynch et al. (2019) study, with lower trading volumes and is used as a control group fund of sorts. ISF, EXS1, and SXRT are the largest European broad-index ETFs tracking FTSE 100, DAX 30 and EURO STOXX 50. SUAS, IESG, and IH2O are ESG ETFs that track socially responsible screening criteria of MSCI USA, MSCI Europe, and Global water infrastructure, respectively. Finally, QUAL is an MSCI USA tracking fund with a quality factor which combines return on equity, earnings variability, and debt-to-equity fundamental components.

4.2. Research Methodology

In this paper, each of the selected funds is investigated by a daily return and volume data, from which "abnormal volume days" are days when the ETF trading volume exceeds three standard deviation threshold of past trading volumes. These abnormal volume days accompanied by a negative ETF return of at least -0.01% for the same day are titled as ETF selloff days and are used as portfolio creation days. In the original Lynch et al. (2019) paper, the required negative return margin was not discussed further. Thus it is expected that the margin is not considerably strict for this strategy.

Some critique can be put on the description of the methods in the Lynch et al. (2019) paper, as the negative return criteria for the selloff day could have been motivated better. It should be noted that increasing the required negative return margin for selloff days decreases the total number of selloff days captured. For instance, changing the return requirement from -0.01% for the day to -0.02% decreases the total amount of selloff days captured from 243 to 213. Setting the return requirement for -1% for the day would amount to 113 selloff days in total for this sample. However, as Lynch et al. (2019) did

not motivate the negative return requirement more precisely, even the slightest negative return for the day with abnormal ETF volume is counted as ETF selloff day.

Abnormal volume days are identified with a Z-score of volume. Adapting Lynch et al. (2019) methods, a three-standard deviation threshold is used, where the abnormal volume day is recorded if the daily ETF volume inspected is three-standard deviations away from its past 40-day rolling mean. Mean, and standard deviation are exponentially weighted to bring relevancy in more recent events than historic. The Z-score on the day i is calculated the following way:

(15)
$$Z = \frac{x - \mu}{\sigma}$$

Where x is the ETF's trading volume on the day i, and μ and σ are the exponentially weighted 40-day rolling mean and standard deviation to day $i - 1$. The weightings are made to each observation with an exponential weighting of the half-life of one-quarter of annual trading days (63 trading days). The weightings placed on each observation are calculated as follows:

(16)
$$Weight = \left[\left(\frac{1}{2}\right)^{1/63}\right]^{n}$$

Where n is the number of observation dates before the starting date. In practice, a half-life of one quarter halves the weight of historical observations by a quarterly basis. Thus, volume observations one quarter away only weigh 50% of the most recent observation. When the abnormal volume days are captured, these dates are used to create the portfolios. The following table will highlight the amount of ETF selloff days and compare them to the total number of abnormal volume days in the sample:

DESCRIPTIVE STATISTICS OF THE SELLOFF DAYS

ETF Symbol	ETF Name	Total # Abnormal volume days	Total # selloff days	Selloff day prevalence of total
IVV	iShares Core S&P 500 ETF	75	50	67 %
QUAL	iShares MSCI USA Quality Factor ETF	68	28	41 %
ISF	iShares Core FTSE 100 UCITS ETF	70	39	56 %
EXS1	iShares Core DAX UCITS ETF	55	32	58 %
SXRT	iShares Core EURO STOXX 50 UCITS	65	24	37 %
SUAS	iShares MSCI USA SRI UCITS ETF	46	18	39 %
IESG	iShares MSCI Europe SRI UCITS ETF	50	21	42 %
IH2O	iShares Global Water UCITS ETF	59	31	53 %
	Median	62	30	47 %
	Total	488	243	

Notes: Full sample starts 01/2016 and ends 08/2020 resulting in a maximum total of 1165 sample days for each ETF. There are slight differences due to different market holidays in the US and European markets. Thus the number of selloff days between different ETFs is not fully comparable with each other. Abnormal volume days are days with ETF trading volume Z score over 3 from the past 40-day rolling volumes. Selloff days are these same days, but with accompanied negative ETF return of at least -0.01% for the day.

Table 2. Descriptive statistics of the selloff days.

As seen in the descriptive statistics, for most of the traditional broad-index funds abnormal volume day is more than 50% of the time accompanied by a negative ETF return. Especially IVV, the largest fund in the sample, experienced a negative return 67% of the time that abnormal volume event of $Z \geq 3$ was captured. On the other hand, the ETFs with fundamental factors tend to experience fewer selloff events compared to the total number of abnormal volume days, suggesting that the abnormal volume days these funds experienced were more often due to increased interest in the fund and thus accompanied with a positive return, instead of a selloff.

In total, 243 selloff days are captured for the whole ETF and further adjusted to reduce clustering for overlapping portfolios. If two consecutive selloff days were observed within five trading days from each other, the portfolio is formed only from the older event date. This restriction decreases the total number of formed portfolios from 243 to 202 and is applied to all the empirical tests in this paper.

The following graph will illustrate the cumulative development of the selloff days during the sample period of 01/2016 – 08/2020.

Cumulative number of selloff days (01/2016 - 08/2020)

■ Cumulative Selloff days   ■ Selloff days
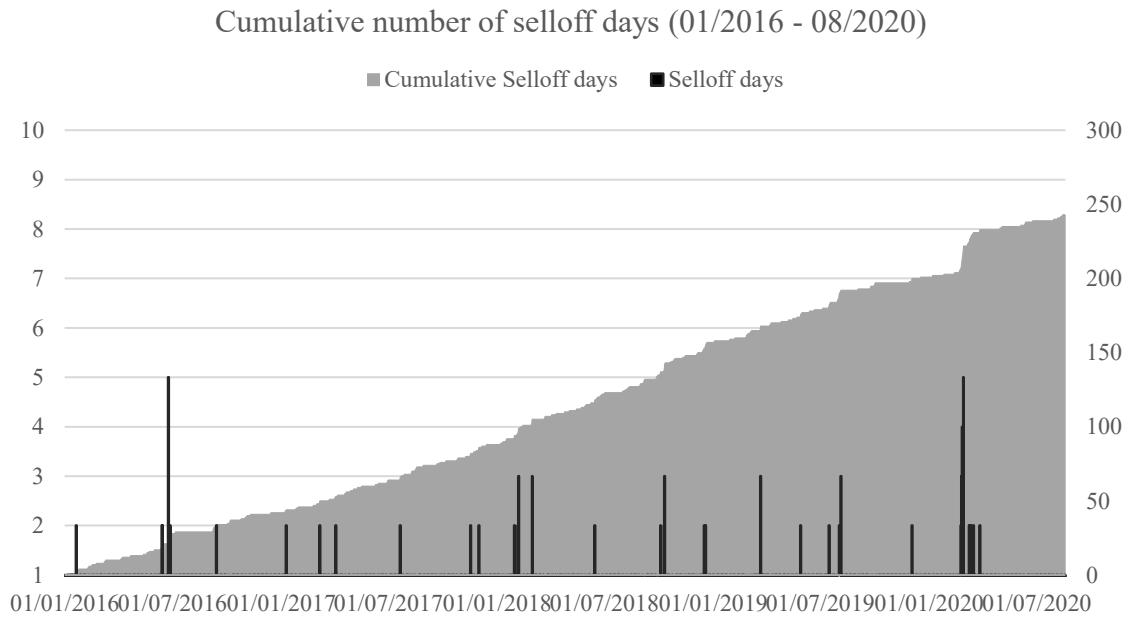
Graph 6. Cumulative number of selloff days.

In Graph 6, days with more than one event are plotted with the black bars, showing whether there are any periods with sample-wide clustering. As expected, a couple of important events can be captured immediately, mainly the 24.6.2016 Britain's Brexit vote, 2018 February brief market selloff, and the recent COVID-19 related market downfall in the turn of 2020 February-March. All the eight funds observed surpassed the three-standard deviation threshold for the selloff event 18 times in total during the six days of 21.2 – 28.2.2020. However, as this event was a global risk-off for the equity asset class as a whole, the results compared to Lynch et al. (2019) might be inconsistent and will be discussed with additional detail later.

4.2.1. Portfolio formation

After capturing the abnormal volume days, outsider stock portfolios from the ETFs are formed. Portfolio starting dates are t+1 after selloff day t, to combat the fact that historical volume data represents market closing volumes. However, if one were to use this strategy

on an ongoing basis, a portfolio could be formed even before the next day if it is apparent that the daily trading volume will break the required three-standard deviation threshold.

To begin the portfolio formation, historical holding details for the ETFs are fetched for each abnormal volume day. The fund provider iShares displays the historical ETF holdings data for its ETFs updated daily for the last four months, and before that, data is updated monthly. Thus, for the historical abnormal volume days occurring between the monthly ETF holdings updates, the older holdings data and weightings are used for the outsider portfolio formation. Some caveats with the ETF portfolio replication would include the fact that in some instances, later privatized companies in the ETF holdings lacked the historical price data even before the privatization. Similarly, some ETFs had transition shares or rights in their holdings during the time of the event. These shares often represented a tiny fraction of the total portfolio and were thus eliminated.

Next, daily historical return data for all the ETF holdings and the ETF itself are acquired. As the ETFs investigated in this paper are purely equity-based funds with physical replication, not much has to be done to clean the ETF holdings for the beta analysis. One additional detail for the ETF holdings replication process would be that cash and money market instruments are eliminated from the original ETF holdings, which often represent $0.01 - 0.05\%$ of the total ETF weight. This is done to ease the next ETF beta estimation step.

The stock beta-to-ETF (β_{ETF}) is estimated for every stock in the ETF after the abnormal volume event. Beta regression is exponentially weighted and is estimated from the event date t to t – 252, thus calculating one-year historical betas from daily data as follows:

(17)
$$\beta_{ETF} = \frac{Cov(R_S R_{ETF})}{Var(R_{ETF})}$$

Where R_S is the exponentially weighted stock return, R_{ETF} is the exponentially weighted ETF return, and Cov and Var are covariance and variance, respectively. The exponential weighting is done with a similar method than with the abnormal volume calculation in formula (16). The beta-to-ETF estimation differs from traditional beta estimation only because instead of using broad market index return, ETF return is used for each constituent stock (Lynch et al. 2019).

After calculating the β_{ETF} for all the ETF holdings, the portfolio is formed from the lowest 10% of the estimated β_{ETF}. Thus, the sensitivity of this portfolio (β_{PORT}) with respect to the ETF should be the following:

$$(18) \qquad \beta_{PORT} = \sum_{i=1}^{n} w_i \beta_{ETF} \,|\beta_{ETF} = 1st\ decile$$

Where w_i is the assumed weighting for the outsider stock in the created portfolio. In the Lynch et al. (2019) paper, the authors used only equally weighted portfolios. In this paper, however, a total of three different portfolio weightings are used. Equal weighted, value-weighted by proportional constituent weights in the fund, and finally fundamentally weighted. The motivation for value weighting portfolios is to capture possible additional effects of selloffs to outsiders with high weighting on the inspected ETF.

The value weighting considers the original ETF holding weight for the constituent and uses it as in proportion against the other constituents selected in the portfolio. The fundamental weighting is measured by Piotroski F-Score for each outsider portfolio stock and weights the stocks with better historical fundamentals over worse. Table 3 illustrates a hypothetical weighting of 5 outsider stocks with different methods:

	Stock 1	Stock 2	Stock 3	Stock 4	Stock 5	**TOTAL**
ETF holding weight	10.0 %	2.5 %	2.0 %	5.0 %	0.1 %	
Outsider Portfolio value weight	51.0 %	12.8 %	10.2 %	25.5 %	0.5 %	100 %
Outsider Portfolio equal weight	20.0 %	20.0 %	20.0 %	20.0 %	20.0 %	100 %
Piotroski F-score	3	4	9	5	1	
Outsider Portfolio fundamental weight	13.6 %	18.2 %	40.9 %	22.7 %	4.5 %	100 %

Table 3. Hypothetical weightings of 5 outsider constituents by different weighting methods.

As seen, value weighting by original ETF holding weights can overweight single stock ownership in the outsider portfolio significantly. However, as the nature of this paper is to try to capture unintentional co-movement induced from ETF stock ownership, it is not necessarily unwanted to overweight these constituents that therefore, should experience co-movement on a more extensive degree due to their high weights in the original ETFs. Additionally, weighting portfolios by the fundamental Piotroski F-score factor

hypothesized that the fundamentally solid companies could be overweighted compared to poorly performing ones.

A portfolio of 10% lowest β_{ETF} stocks is generated after the day of the abnormal volume event. All stocks are bought long and held for 40 days. Portfolio returns are calculated alongside with the abnormal returns (ARs), cumulative abnormal returns (CARs), and cumulative average abnormal returns (CAARs) adapted from (Lynch et al., 2019) with the following formulas:

(19)
$$AR_{i,t} = R_{ETF} - \left[R_f + \beta_{PORT}\left(R_{ETF} - R_f\right)\right]$$

where the R_{ETF} is the ETF return, R_f is the risk-free rate and β_{PORT} is the portfolio beta calculated earlier:

(20)
$$CAR_i = \sum_{t=T_1+1}^{T_2} AR_{i,t}$$

CARs are calculated for each formed portfolio. The lowest number of abnormal volume events recorded and portfolios created for one ETF was for SUAS with 17 events. The highest number of events was recorded for IVV with 36 total acceptable events. After solving the abnormal returns for these portfolios, a final CAAR for the ETF strategy is calculated:

(21)
$$CAAR = \frac{1}{N} \sum_{i=1}^{N} CAR_i$$

The significance of the possible returns and abnormal returns generated by these portfolios are then estimated by Wilcoxon signed-rank statistical test, as the short 40-day holding period implies non-normality in the sample. To recap the three key points of interest in this paper introduced in the first chapter, the three hypotheses set are presented again more explicitly. Hence, the first hypothesis was stated as follows:

H₁: The low beta-to-ETF stocks generate statistically significant abnormal returns after high volume ETF selloffs.

For this hypothesis, the two-tailed test with Wilcoxon signed-rank test is utilized. The signed-rank test is a traditional non-parametric test for event study abnormal return settings. This method was used by Lynch et al. (2019) and is formed as follows:

(22)
$$H_o: abnormal\ return\ median = 0$$
$$H_1: abnormal\ return\ median \neq 0$$

1. Differences (d_i) between each observation (x_i) and the hypothesized median $d_i = x_i - 0$ are calculated.
2. Each $|d_i|$ is given a rank from 1 to n.
3. Ranks are labeled with their signs ($+$ or $-$).
4. Calculate the sum of the ranks with positive d_i, W^+ and the sum of the ranks with the negative d_i, W^-
5. Test the statistic $W = \min(W^+, W^-)$ against the Wilcoxon critical values.

For the long-term strategy evaluation, the formed portfolios are backtested from 01/2016 to 08/2020 systematically. In practice, the strategy is treated as it was utilized continuously from 2016 onwards. The obvious caveat with the selloff event clustering, i.e., a situation where several portfolios returns overlap, makes the perceived effects from the original events unreliable. Lyon, Barber & Tsai (1999) discuss the impact of overlapping event returns in their study and find the only solution to counter the cross-sectional dependence of the sample is to remove the overlapping observations.

To enhance the real-world functionality and minimize the cross-correlation problems in the sample, the overlapping event portfolios are treated so that the next event portfolio is only started after the previous has ended. This is the method Lyon et al. (1999) suggested to control for the cross-sectional dependence. This decreases event clustering and increases the significance of the statistical methods used. As the time-structure (Lynch et al. 2019) found for the cumulative average abnormal returns in their paper was consistent towards the end, starting the strategy between the 40 day holding period should not cause limitations to the overall returns.

As the nature of this study differs slightly from traditional event studies, no clearly correct way of estimating long-term returns can be immediately motivated. Most of the existing literature debate between buy-and-hold abnormal returns (BHARs) and the calendar time portfolios (CTP), which both have their caveats. (Fama 1998) argue for using either short-term abnormal returns (AARs and CARs) or the CTP method for preventing a cross-

correlation problem in the returns. In this paper, CTP is used in combination with the Fama-French 5-factor model to estimate long-run abnormal returns.

As the sample initially uses daily returns to create short-term portfolios in the first section, Calendar-time portfolios are daily equally weighted instead of monthly. Furthermore, the nature of this study prevents estimating similarly different holding periods per event induced, as in traditional event studies, since each "event" is used as a trigger to create a unique portfolio that is sold after 40 days. Thus, this long-term strategy section does not repeat the analysis of the first 40-day abnormal return section but focuses on the performance of the strategy if it was systematically reproduced for the duration of the whole sample. For the analysis of the long-term strategy, the second hypothesis was formed:

H$_2$: The low beta-to-ETF strategy generates statistically significant abnormal returns as a long-term systematic strategy.

Thus, for the second hypothesis, raw buy-and-hold returns and Sharpe ratios for both the ETF and the strategy are presented. Then, the excess returns of the ETF strategies are estimated by the Fama-French 5 factor model, but with a supplemental measure to investigate the usefulness of the beta-to-ETF.

Frazzini & Pedersen (2014) find that high beta assets are generally associated with low alpha. Authors then constructed portfolios that long low-beta securities and short-sell securities with high beta, which produces significant positive returns. The Betting-against-beta or BAB factor is interesting as this paper takes a somewhat similar approach with the difference of equity betas being calculated by their parent ETF. However, especially for broad-index funds, the estimated "betas-to-ETF" are equivalent to traditional beta estimates for these stocks.

Another key difference between the BAB factor and low beta-to-ETF strategy is that this strategy does not take the opposite, short-sell bet for high beta-to-ETF stocks. As Lynch et al. (2019) disregard this as the point of the strategy is to capture the outsider stocks that are less sensitive towards the index, or in this instance, ETF, they are part of. The BAB betas are estimated by the formula (23) with a 1-year rolling standard deviation for volatilities with a five-year correlation estimate. One-day log-returns are used to estimate volatilities, and three-day log-returns are used with correlations to control for non-

synchronous trading. Additionally, time-series estimates for the betas are shrunk towards the cross-sectional mean (β^{XS}) by:

(23)
$$\beta = w\beta^{TS} + (1-w)\beta^{XS}$$

where: β^{TS} = Time-series estimates for the betas
β^{XS} = Cross-sectional means of the betas
w = weight that is set at 0.6

BAB portfolios are then constructed by going long to low-beta securities and shorting high beta securities. Aggregate country/region portfolios are then weighted by the region's total lagged market capitalization. Estimated betas are ranked and assigned to low beta and high beta categories. The portfolios are rebalanced by every calendar month as to where $z = n$ x 1 is a vector of beta ranks $z = rank(\beta)$ at the portfolio formation, and $\bar{z} = 1'_n z/n$ is the average rank, where n is the number of securities and the 1_n is an n x 1 vector of ones.

The portfolio weights of the high beta and low beta portfolios are given by $w_H = k(z - \bar{z})^+$ and $w_L = k(z - \bar{z})^-$, where k functions as a normalizing constant $k^{-1} = 1'_n |z - \bar{z}|/2$ and x^+ and x^- are the positive and negative elements of the vector x. To construct the BAB factor, the acquired high and low portfolios are rescaled to beta 1 at the time of the formation. As the BAB factor short sells and longs simultaneously, it is a self-financing zero-beta portfolio by design. The full BAB factor is constructed as follows:

(24)
$$r_{t+1}^{BAB} = \frac{1}{\beta_t^L}(r_{t+1}^L - r^f) - \frac{1}{\beta_t^H}(r_{t+1}^H - r^f)$$

where: $r_{t+1}^L = r'_{t+1} w_L$
$r_{t+1}^H = r'_{t+1} w_H$
$\beta_t^L = \beta'_t w_L$
$\beta_t^H = \beta'_t w_H$

For this paper, the BAB factors are obtained from the AQR Capital Management's (AQR 2020) dataset library, which provides the BAB factors for US, European and Global equities. To ensure the usefulness of the BAB factor, correlations between traditional FF5 model factors are presented:

CROSS-CORRELATIONS BETWEEN THE FF5 + BAB FACTORS

	Mkt-RF	SMB	HML	RMW	CMA	BAB
Mkt-RF	1					
SMB	0.22	1				
HML	0.20	0.35	1			
RMW	-0.03	-0.07	0.19	1		
CMA	-0.19	-0.02	0.47	0.12	1	
BAB	0.00	-0.28	-0.25	-0.02	-0.04	1

Notes: Correlations between factors are calculated from the sample period 01/2016 to 08/2020 and Fama-French 2x3 US factors.

Table 4. Cross-correlations between the FF5 + BAB factors.

As seen in Table 4, the correlations follow mostly the pattern of (Fama & French, 2015) apart from the HML factor with small positive correlations with the Mkt-RF and SMB factors, which were -0.30 and -0.11 respectively, in the original study. The inspected BAB factor correlates as a zero-beta market portfolio, as clarified earlier. Additionally, the correlations with the other factors are either negative or close to zero, ensuring that the BAB factor provides unique uncorrelated information to the model.

To further show that BAB is a useful addition to the FF5 model for this sample, a right-hand side test is used to prove that the other factors do not span the factor. If the BAB factor adds to the explanatory power of the FF5 model, BAB returns for the sample period are regressed against the other FF5 factors. If the spanning regression yields a nonzero, statistically significant intercept, the factor adds to the model's explanatory power for this sample period (Fama & French 2018). In the following table, the spanning regression for each BAB factor are presented:

BAB FACTOR SPANNING REGRESSION FOR FF5 FACTORS

	α	Mkt-RF	SMB	HML	RMW	CMA	R²
BAB US	0.0003***	-0.19***	0.01	-0.38***	0.32***	0.57***	0.35
	(0.00)	(0.00)	(0.74)	(0.00)	(0.00)	(0.00)	
BAB EU	0.0003**	-0.16***	0.41***	-0.12	0.00	0.21**	0.19
	(0.04)	(0.00)	(0.00)	(0.15)	(0.99)	(0.03)	
BAB INTL	0.0002*	-0.17***	-0.26***	-0.18***	0.26***	0.07	0.27
	(0.07)	(0.00)	(0.00)	(0.00)	(0.00)	(0.16)	

Notes: BAB US, Europe, and International factors are regressed against the Fama-French factors for four years, from 01/2016 to 01/2020. P-values are in parentheses.

*** Denotes two-tailed significance at the 1% level.

** Denotes two-tailed significance at the 5% level.

* Denotes two-tailed significance at the 10% level.

Table 5. BAB factor spanning regression for FF5 factors.

The results in Table 5 show that all the intercepts obtained are nonzero, although relatively close to zero, and statistically significant at least on the 10% level. Therefore, adding the BAB factor to the FF5 is useful for the model's full explanatory power. Thus, the FF5 model is expanded with the BAB factor to further capture the long-term abnormal return and the sensitivity of this low beta-to-ETF strategy to this factor. To prove the second hypothesis set in this thesis, the strategies should yield positive alpha by these measures.

Finally, the last area of interest in this paper is the usefulness of the fundamental proxy for the low beta-to-ETF strategy. As described earlier, Piotroski F-Score is used as a proxy to model discretionary, fundamental analysis, and forms the third hypothesis was stated as follows:

H3: The fundamental proxy provides better, statistically significant abnormal returns for low beta-to-ETF portfolios after high volume ETF selloffs.

For the usability of the fundamental proxy, 1-40 day low beta-to-ETF portfolios are weighted with calculated F-Score for each stock, instead of fund holding value or equal weighting. To fulfill the third hypothesis, abnormal returns of the 1-40 day strategy should prove to increase with fundamental sorting. As the Piotroski F-Score requires past financial statement information about the portfolio companies, four variables are created

for the ORBIS database to efficiently fetch this information. The used variables can be found in the Appendix 8.

The F-Score is measured for each stock at the time of the portfolio creation date by looking at the last annual statement during that period. This restriction is made to combat the significant look-ahead bias that would otherwise arise. For instance, if the portfolio was created on 01/2020, the latest annual report available at the time was not the 2019 report, but the 2018 report as the preparation of the 2019 financial statement was in progress. Thus, there can be significant lag with the fundamental component for specific portfolios, which is typical for any financial ratios that require comprehensive annual statement-based data (Breitschwerdt 2015).

5. EMPIRICAL RESULTS

In this section, the performance and properties of the created strategy portfolios are analyzed. First, following the methodology of Lynch et al. (2019), proof of constituent cross-correlations during abnormal volume periods in the sample are presented. Secondly, the acquired portfolio cumulative average abnormal returns (CAARs) are estimated by the periods of selloff day t+5, t+10, t+20, and t+40 to gather whether there is a consistent time structure behind the portfolio abnormal returns. Following this, the strategy is applied in a systematic, long-term fashion, and the annualized returns of the selected ETFs and implemented strategies are analyzed with conventional portfolio evaluation metrics.

Next, the long-term strategies are estimated with the Fama-French 5-factor model to determine the abnormal returns generated by the strategies. The factor model is applied with and without the BAB factor expansion introduced earlier to assess whether actual abnormal returns can be captured long-term. Finally, the usefulness of the fundamental proxy suggested is tested by comparing the returns of the traditional portfolio weighting methods against the fundamentally weighted alternative.

5.1. Evaluation of the short-term strategy

By examining selected ETF constituent behavior during the ETF abnormal volume days, a similar pattern to Lynch et al. (2019) can be captured by the broad index funds. However, the funds with a fundamental stock selection criterion seem to have a slight inverse factor on the constituent correlations after an ETF volume spike. Mainly funds QUAL, SUAS and IESG showed, on the one hand, heightened constituent correlations, but also, on the other hand, substantial negative correlations between some constituents. Thus, averaging the overall cross-constituent correlations towards zero.

Cross-constituent correlations (ρ) for ETF constituent stocks are calculated as follows:

$$(25) \qquad \rho = f(x) = \begin{cases} corr(x,y), & All \\ corr(x,y), & Z \geq 3, Abnormal\ volume \end{cases}$$

Where x and y are respectively the time series of daily returns for two ETF constituents, and Z is the Z-score of daily ETF trading volume. The Z-score is calculated by the

methods of formulas (15) and (16). Average cross-constituent correlations from 01/2019 to 08/2020 are as follows:

Average cross-constituent correlations 01/2019 - 08/2020
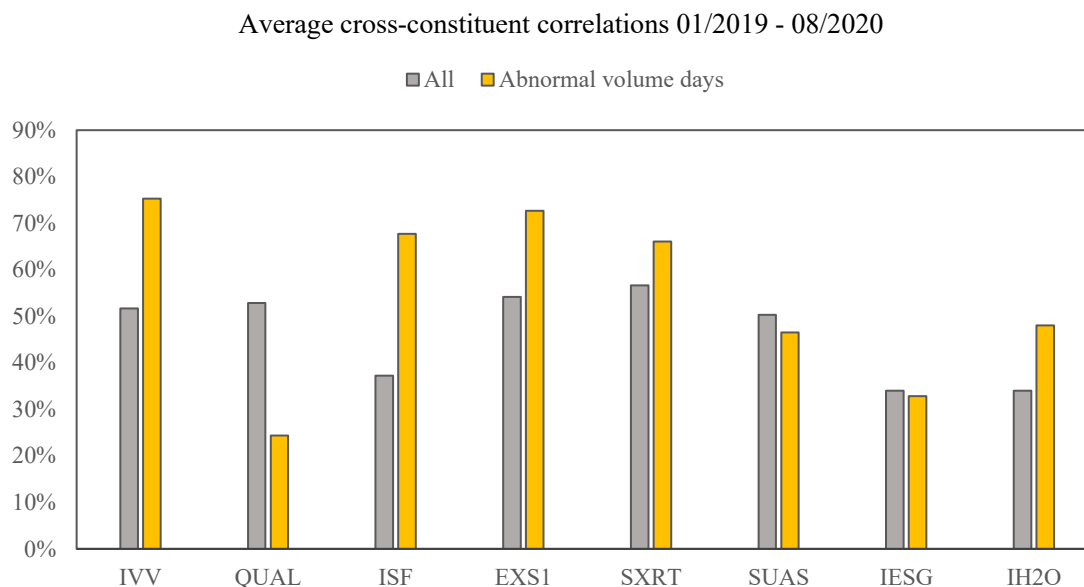
Graph 7. Average cross-constituent correlations at abnormal volume days versus all days.

As seen in Graph 7, most of the selected ETF constituents follow the pattern identified by (Da & Shive 2018; Lynch et al. 2019). Nevertheless, as mentioned earlier, ETF constituents with fundamental selection criteria did not experience a unified pattern in cross-constituent correlations. Thus, for the funds QUAL, SUAS and IESG, the cross-correlations averaged towards zero for the abnormal volume days. This inverse relationship is curious and suggests that some stocks in these thematical funds could possess flight-to-safety attributes in terms of abrupt market sentiment changes. For example, stocks that emphasize corporate social responsibility have been found to be less sensitive to crash risk (Kim, Li & Li 2014).

Next, the CAARs acquired for each ETF by incorporating the low beta-to-ETF strategy are presented. The CAARs are calculated from the generated event portfolios with the methods introduced in formulas (19), (20), and (21). The sample period is 01/2016 – 08/2020 to provide convenient continuity to the Lynch et al. (2019), who investigated the alphas for U.S.-based SPDRs from 2010 to 2017. In the first table, results of the equally weighted outsider stock portfolios are shown:

EQUAL WEIGHTING

	Days after volume spike t			
	t+5	t+10	t+20	t+40
IVV	-0.003	-0.006	-0.011	0.000
	(0.54)	(0.16)	(0.50)	(0.79)
QUAL	-0.001	0.006	-0.003	0.007
	(0.97)	(0.15)	(0.13)	(0.14)
ISF	-0.001	0.003	0.008	0.026**
	(0.51)	(0.87)	(0.12)	(0.01)
EXS1	-0.001	0.001	0.001	0.008
	(0.95)	(0.95)	(0.75)	(0.44)
SXRT	0.003	-0.003	0.000	0.007
	(0.76)	(0.97)	(0.73)	(0.55)
SUAS	0.014**	0.015**	0.019**	0.024**
	(0.02)	(0.01)	(0.01)	(0.02)
IESG	0.000	0.004	0.009	0.015
	(0.41)	(0.46)	(0.26)	(0.13)
IH2O	-0.008	-0.016	-0.008	-0.003
	(0.26)	(0.12)	(0.43)	(0.34)

Notes: This table presents the CAARs for each selected ETF by the periods of t+5, t+10, t+20 and t+40, t being the abnormal ETF volume event. The ETF-beta estimation period starts from 01/2015 and the full sample consists of 01/2016 - 08/2020. The Wilcoxon signed-rank test indicates significance levels of 0.05 (**) and 0.01 (***). Please see table 1 for ETF symbol explanations.

Table 6. Cumulative average abnormal returns for outsider portfolios created after ETF selloff day, Equal weighted.

Similarly to Lynch et al. (2019) paper, some evidence of abnormal returns obtained from the strategy can be captured. Especially the funds ISF and SUAS presented statistically significant 2.6% (p = 0.01) and 2.4% (p = 0.02) CAARs after 40 days, p-values in parentheses. However, similarly consistent and significant abnormal return patterns that Lynch et al. (2019) found could not be confirmed immediately. In their paper, authors found that approximately after t+5, most investigated funds produced already positive CAARs. At the t+40, most of the CAARs ranged from 2% to over 7%, while most being statistically significant.

Especially the control group fund IVV produced disappointing returns. The time-structure of the returns was mostly negative and ended up to zero at the t+40. Even though total CAARs were positive for most of the funds, apart from the fund IH2O, the significance of the returns could not be verified by the statistical methods used. For Table 7, portfolio

weighting methods are changed from Lynch et al. (2019) used equal weighting to value weighting by proportional constituent weights in ETF.
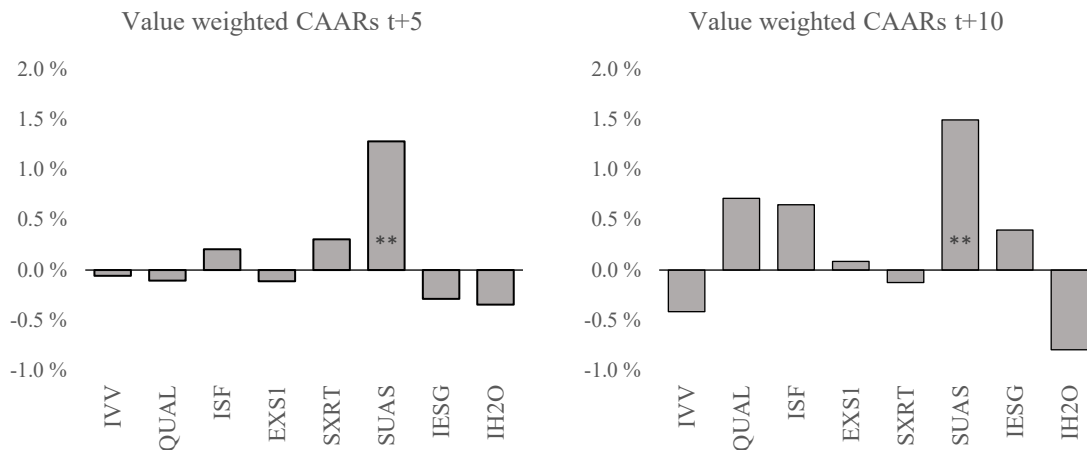
VALUE WEIGHTING			
Days after volume spike t			
t+5	t+10	t+20	t+40
IVV			
-0.001	-0.004	-0.010	0.004
(0.89)	(0.46)	(0.53)	(0.35)
QUAL			
-0.001	0.007	-0.002	0.009
(0.99)	(0.14)	(0.40)	(0.20)
ISF			
0.002	0.007	0.013**	0.028***
(0.91)	(0.44)	(0.04)	(0.00)
EXS1			
-0.001	0.001	0.001	0.009
(0.99)	(0.79)	(0.66)	(0.35)
SXRT			
0.003	-0.001	0.002	0.010
(0.48)	(0.88)	(0.65)	(0.31)
SUAS			
0.013**	0.015**	0.018**	0.025**
(0.03)	(0.01)	(0.01)	(0.03)
IESG			
-0.003	0.004	0.007	0.007
(0.53)	(0.48)	(0.48)	(0.24)
IH2O			
-0.003	-0.008	0.008	0.012**
(0.40)	(0.26)	(0.07)	(0.03)

Notes: This table presents the CAARs for each selected ETF by the periods of t+5, t+10, t+20, and t+40, t being the abnormal ETF volume event. The ETF-beta estimation period starts from 01/2015, and the full sample consists of 01/2016 - 08/2020. The Wilcoxon signed-rank test indicates significance levels of 0.05 (**) and 0.01 (***). Please see table 1 for ETF symbol explanations.

Table 7. Cumulative average abnormal returns for outsider portfolios created after ETF selloff day, Value weighted.

The results captured here follow the pattern of Table 6 with the difference that most of the p-values of the results strengthened, albeit most of them still being insignificant. Similarly, to Table 6 ISF and SUAS produced the highest CAARs after 40 days, 2.8% (p = 0.00) and 2.5% (p = 0.03) respectively. Additionally, the fund IH2O produced a significant 1.2% (p = 0.03) return, an absolute increase of 1.5% from the equally weighted alternative -0.3% (p = 0.34). As seen, the value-weighting method proves superior here with this sample and will be used as the baseline in the following long-term performance evaluation sections. The idea of higher stock volatility due to higher ETF ownership (Ben-David et al. 2018) may have some basis here, as the largest owners of the selected outsider stocks were almost unanimously asset management companies.

However, as mentioned earlier, overweighting single stocks in this outsider strategy is not an unwanted attribute and can even capture the ETF induced constituent co-movement better. Even the control group ETF IVV experiences enhanced returns from the value weighting with 0.4% (p = 0.35) positive CAAR at t+40. Overall, the CAAR time-structure is smoother and ends up positive for every ETF inspected. By plotting the value-weighted CAARs to graphs of t+5, t+10, t+20, and t+40, a better overview of the returns can be captured:

Graph 8. T+5 and T+10 CAARs for each ETF strategy, Value weighted. Significance levels of 0.05 are denoted by (**) and 0.01 by (***).
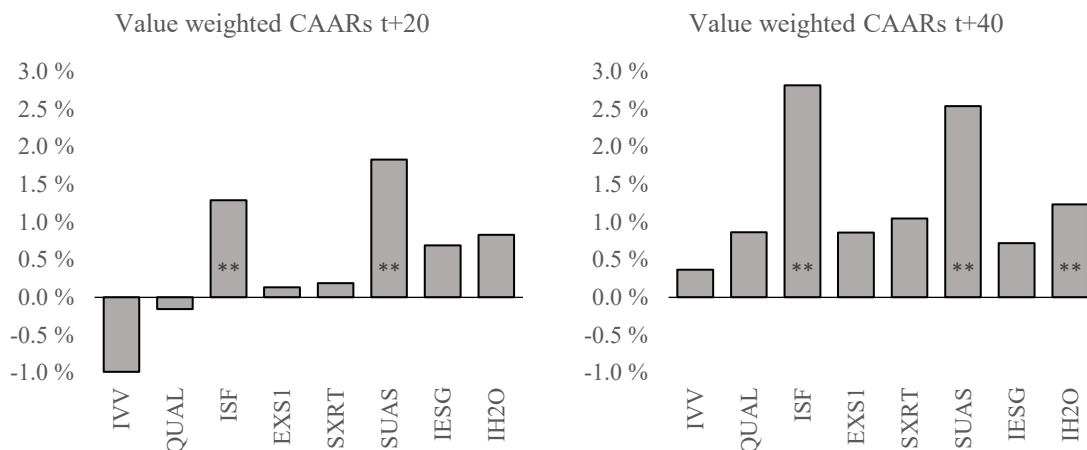
Graph 9. T+20 and T+40 CAARs for each ETF strategy, Value weighted. Significance levels of 0.05 are denoted by (**) and 0.01 by (***).

As briefly discussed earlier, the 2020 February-March COVID-19 related market crash is a complicating factor when comparing these results to (Lynch et al., 2019), as during their sample period (01/2010 – 12/2017) starting just after the financial crisis of 2008, no

significant full economy-wide market crashes occurred apart from the European debt crisis in 2011, that had limited effects to the U.S. markets in its entirety. To counter this disparity, a COVID-19 adjustment for the portfolios is made for backtesting purposes. In this adjustment, all the selloff events happened between February – March 2020 are omitted, in combination with the value-weighted portfolios, to produce striking results:

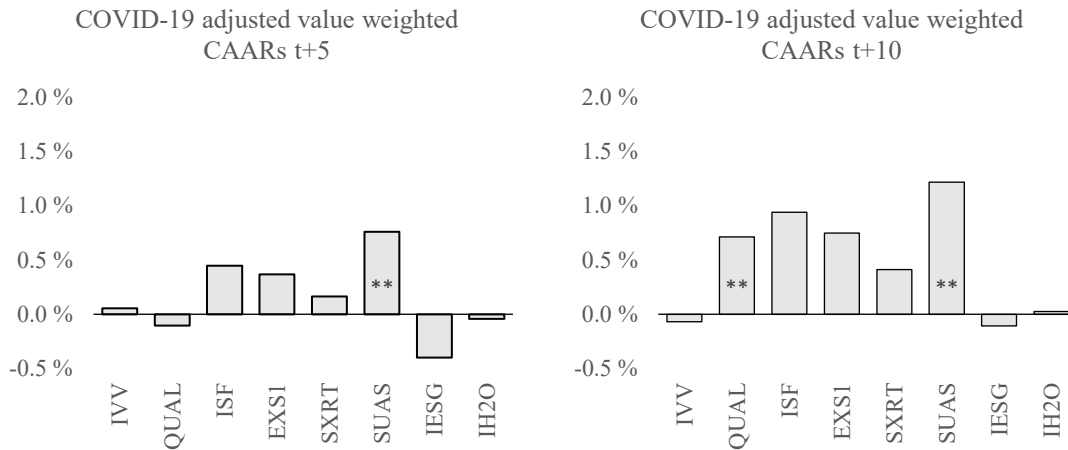|	COVID-19 ADJUSTED VALUE WEIGHTING			
	Days after volume spike t			
	t+5	t+10	t+20	t+40
IVV	0.001	-0.001	-0.002	0.009
	(0.64)	(0.92)	(0.73)	(0.10)
QUAL	0.001	0.008**	0.006	0.012**
	(0.36)	(0.05)	(0.10)	(0.05)
ISF	0.004	0.009	0.016***	0.025***
	(0.22)	(0.10)	(0.01)	(0.00)
EXS1	0.004	0.007	0.009	0.016
	(0.26)	(0.15)	(0.12)	(0.07)
SXRT	0.002	0.004	0.005	0.009
	(0.29)	(0.23)	(0.17)	(0.14)
SUAS	0.008**	0.012**	0.016***	0.023**
	(0.05)	(0.02)	(0.01)	(0.03)
IESG	-0.004	-0.001	0.003	0.010
	(0.16)	(0.28)	(0.31)	(0.20)
IH2O	0.000	0.000	0.014**	0.013**
	(0.48)	(0.49)	(0.02)	(0.03)

Notes: This table presents the CAARs for each selected ETF by the periods of t+5, t+10, t+20 and t+40, t being the abnormal ETF volume event. The ETF-beta estimation period starts from 01/2015 and the full sample consists of 01/2016 - 08/2020. The Wilcoxon signed-rank test indicates significance levels of 0.05 (**) and 0.01 (***). Please see table 1 for ETF symbol explanations.

Table 8. Cumulative average abnormal returns for outsider portfolios created after ETF selloff day, COVID-19 adjusted, Value weighted.

By adjusting to the February - March COVID-19 related market turmoil, the CAARs of the strategy are strengthened significantly. Following the earlier two Tables 6 and 7, ISF and SUAS still produce the most impressing CAARs, 2.5% (p = 0.00) and 2.3% (p = 0.03) respectively after 40 days. However, two additional funds experience statistically significant CAARs as QUAL and IH2O show CAARs of 1.2% (p = 0.05) and 1.3% (p = 0.03). Additionally, IVV with 0.9% (p = 0.10) and EXS1 1.6% (p = 0.07) are just shy of

being statistically significant on the 0.05 level. Overall, all the p-values decrease greatly towards statistical significance, and acquired CAARs are improved.

Thus, the acquired results imply that the outsider strategy is still functional but does not bring protection or superior returns in terms of whole market-wide risk-off scenarios experienced in February – March 2020. Again, by plotting these value-weighted CAARs to graphs, the overall improved time structure of the CAARs can be noticed:

Graph 10. T+5 and T+10 CAARs for each ETF strategy, COVID-19 adjusted, Value weighted. Significance levels of 0.05 are denoted by (**) and 0.01 by (***).
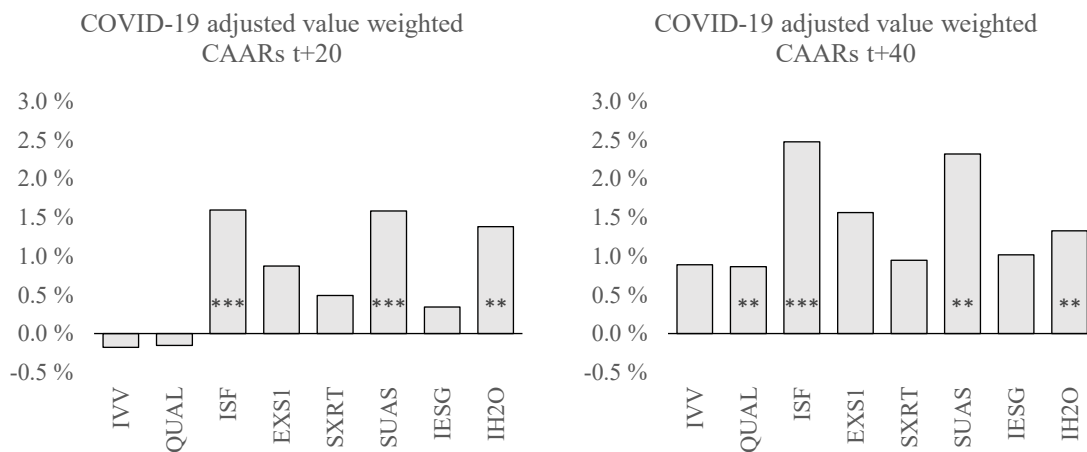
**Graph 11.** T+5 and T+10 CAARs for each ETF strategy, COVID-19 adjusted, Value weighted. Significance levels of 0.05 are denoted by (**) and 0.01 by (***).

As mentioned, the overall time structure of CAARs is enhanced, and the significance of the acquired results is much improved. All the eight selected ETFs produced positive CAARs after 40 days with the COVID-19 adjustment. Funds ISF, EXS1, SXRT, and SUAS generated the most consistently positive returns starting from t+5. The control group fund IVV is still slightly disappointing compared to the CAARs Lynch et al. (2019) acquired (6% after 40 days, statistically significant by 0.01 level). However, different time-period for the sample that the fact that IVV and SPY trading volume patterns are not identical impacted the results.

Finally, for the short term CAAR analysis, value-weighted portfolio returns with and without the COVID-19 adjustment are plotted in a graph illustrating the average CAARs acquired for the whole sample:
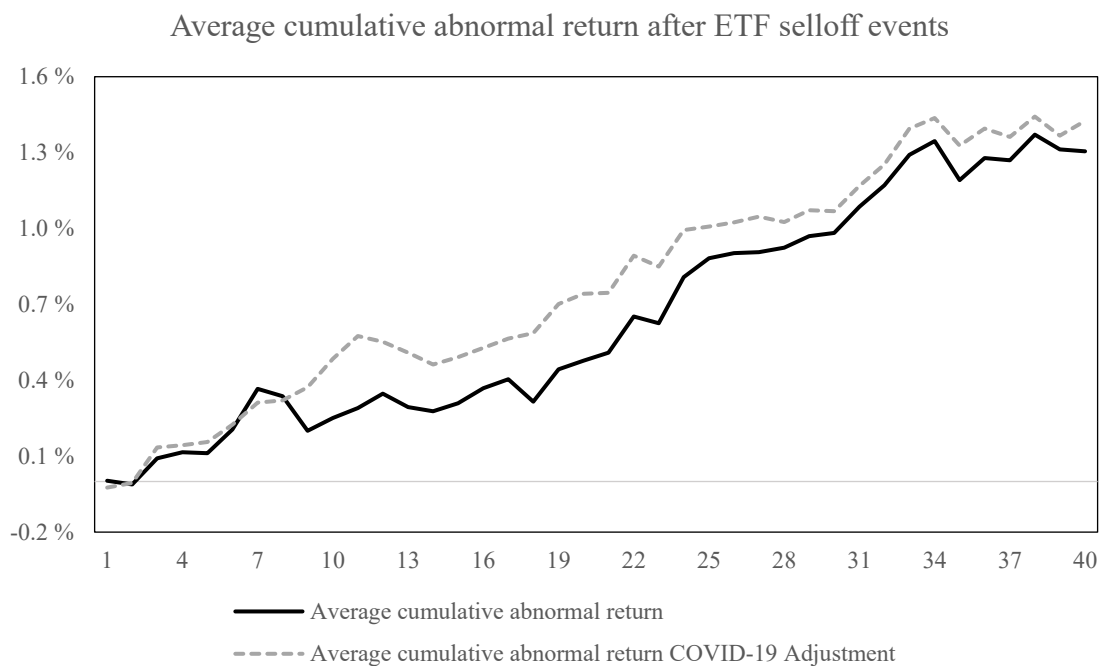
Graph 12. Average CAARs after ETF selloff days Value weighted against COVID-19 adjusted Value weighted.

As Graph 12 illustrates, COVID-19 adjustment generates superior, more consistent returns sample-wide. The average CAAR received in the sample is 1.3% after 40 days, and with the COVID-19 adjustment, 1.43%. Even though by absolute means, the difference is just 0.13%, the overall improved statistical significance of the COVID-19 adjusted results advocates that the adjustment is necessary for full comparability to

previous studies. Overall, the amount of statistically significant findings with the value-weighted portfolios were similar to the Lynch et al. (2019) equal-weighted findings, who found 5 of the total 11 CAARs to be statistically significant at the end of the 40 days. However, when comparing the magnitude of figures obtained to the previous Lynch et al. (2019) paper, it is essential to consider that the qualities of the sample used in this paper differed not only regionally but with smaller median AUM and lower trading volumes. Thus, it can be expected that the CAARs captured may not be as aggressive as with larger funds with higher trading volumes. After all, the positive CAARs obtained in this paper further confirm the low beta-to-ETF metric's usefulness combined with ETF selloffs as a universal trading rule to consider. In the next section, the strategy is adapted for systematic long-term practice and measured against traditional asset pricing models.

5.2. Evaluation of the long-term strategies

For the long-term strategies, the non-adjusted value-weighted 40-day portfolio holding periods are used and backtested consistently from 01/2016 to 08/2020. The strategies include the same low beta-to-ETF portfolio creation process, but the non-event periods are filled with either the return of the parent ETF or the risk-free rate. The first strategy with ETF ownership between the event dates shows the performance opportunity present with the low beta-to-ETF strategy compared to just passively holding the ETF. However, the risk-free return strategy dismisses owning the parent ETF and demonstrates the raw return acquired from the outsider stock-picking strategy.

First, descriptive statistics of the selected ETFs and the strategies incorporated are presented, and in the table, average annual returns, standard deviations, and Sharpe ratios are presented for the parent ETF, denoted as the "X ETF" where X indicates the ETF ticker in question. The first strategy, where the non-event period days are filled with the return of the parent ETF, is denoted as the ("X Strategy"), where X indicates the ETF ticker in question and finally, the second strategy with non-event period days filled with the risk-free rate is denoted as ("X RF Strategy"), where the X indicates the ETF ticker in question. The descriptive statistics are as follows:

DESCRIPTIVE LONG-TERM STRATEGY RETURNS

	Average Annual Return	Average Annual Stdev	Sharpe Ratio	#Obs
IVV ETF	10.80 %	0.195	0.55	1155
IVV STRATEGY	10.64 %	0.184	**0.58**	1155
IVV RF STRATEGY	9.20 %	0.162	**0.57**	1155
QUAL ETF	10.17 %	0.191	0.53	1155
QUAL STRATEGY	**10.54 %**	0.180	**0.58**	1155
QUAL RF STRATEGY	6.46 %	0.164	0.39	1155
ISF ETF	-4.25 %	0.209	-0.20	1159
ISF STRATEGY	**8.40 %**	0.153	**0.55**	1159
ISF RF STRATEGY	**7.31 %**	0.130	**0.56**	1159
EXS1 ETF	3.45 %	0.211	0.16	1160
EXS1 STRATEGY	**6.09 %**	0.185	**0.33**	1160
EXS1 RF STRATEGY	**4.79 %**	0.144	**0.33**	1160
SXRT ETF	3.92 %	0.205	0.19	1174
SXRT STRATEGY	**5.78 %**	0.183	**0.32**	1174
SXRT RF STRATEGY	**6.19 %**	0.117	**0.53**	1174
SUAS ETF	15.66 %	0.170	0.92	1023
SUAS STRATEGY	12.91 %	0.181	0.71	1023
SUAS RF STRATEGY	4.34 %	0.142	0.30	1023
IESG ETF	6.71 %	0.178	0.38	1184
IESG STRATEGY	**7.95 %**	0.178	**0.45**	1184
IESG RF STRATEGY	-1.67 %	0.111	-0.15	1184
IH2O ETF	8.29 %	0.171	0.49	1194
IH2O STRATEGY	**10.43 %**	0.180	**0.58**	1194
IH2O RF STRATEGY	5.84 %	0.151	0.39	1194

Notes: This table presents the raw returns from just holding the investigated ETF and two different strategy scenarios where in "ETF STRATEGY" the ETF is held in between the event dates, and in "ETF RF STRATEGY" only risk-free rate is received in between the event dates. Please see table 1 for ETF symbol explanations. Sharpe ratios and annual returns for strategies are **bolded** when they exceed the parent ETF annual returns and Sharpe ratios.

Table 9. Descriptive long-term strategy returns.

Examining the most conventional portfolio performance evaluation metrics, utilizing the low beta-to-ETF strategies has a positive effect against the parent ETF returns. By combining the passive ETF ownership with the active strategy of buying outsider constituents after ETF selloff days, investors receive superior returns by absolute measure for the QUAL, ISF, EXS1, SXRT, IESG, and IH2O funds. For the rest of the funds, most of the relative return underperformance is countered with increasing Sharpe ratios, as only the SUAS ETF Strategies underperform the parent ETF by all measures. However, the underperformance could be partially explained by the relative lack of total selloff dates captured for the SUAS ETF.

As SUAS and IESG ETFs experienced the least amount of selloff days, their RF Strategies underperform most against the parent ETFs, as the number of days where the risk-free, or in practice, zero-rate is received increases significantly. For the other ETFs, the RF Strategy significantly lowers the standard deviation of the returns compared to the parent ETFs and the ETF Strategy, as expected. The following graphs will illustrate the best performing long-term strategies. All the graphs from the strategies can be found at the appendices 1-6 at the end of this paper:
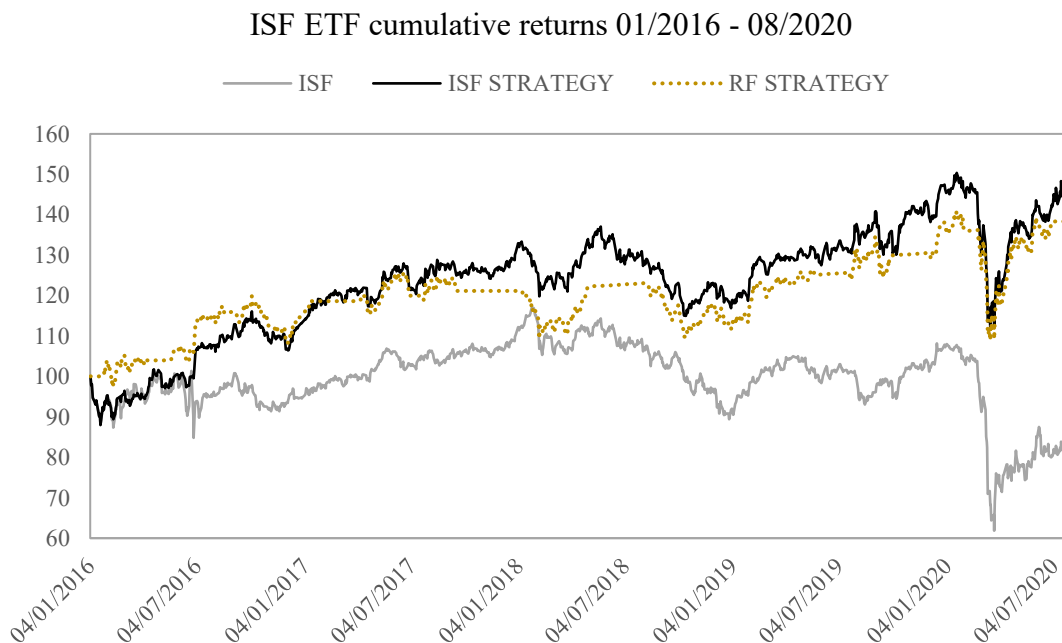
Graph 13. Cumulative returns for the ISF ETF, from 01/2016 to 08/2020.

As seen in Graph 13, both strategies overperform against the FTSE 100 following ISF ETF almost the entire sample period, and most of the overperformance was gained during the UK Brexit vote in the summer of 2016. It seems to be that for that specific event, being invested in low beta-to-ETF stocks in the UK region proved to be a useful hedge against the Brexit-induced market turmoil in the UK. Next, the DAX 30 tracking EXS1 ETF is analyzed:

EXS1 ETF cumulative returns 01/2016 - 08/2020
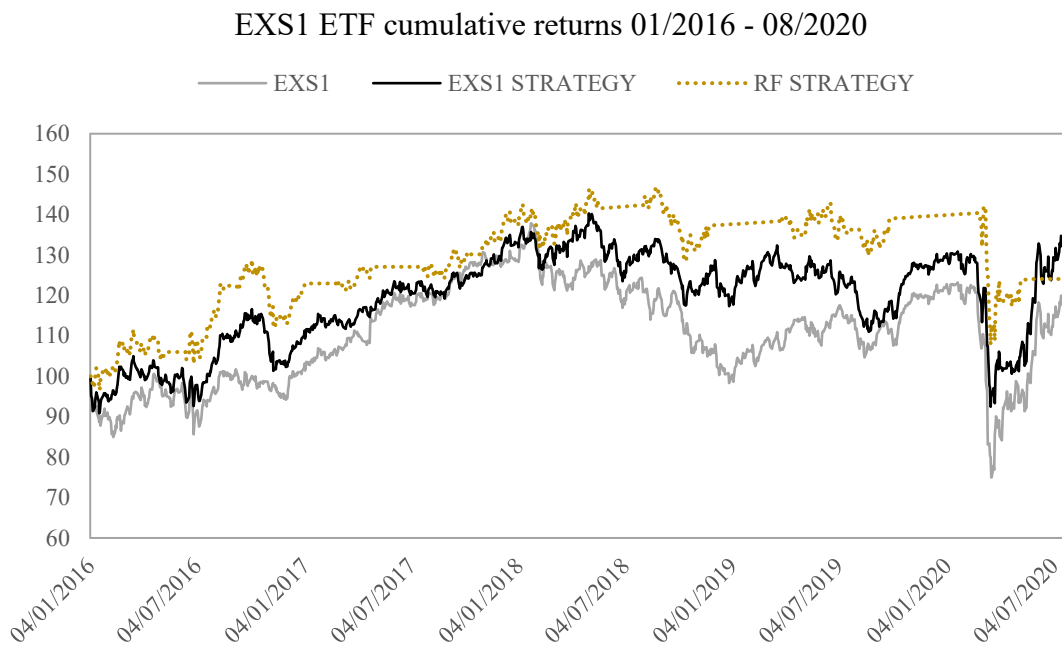
Graph 14. Cumulative returns for the EXS1 ETF, from 01/2016 to 08/2020.

As Graph 14 shows, the EXS1, or DAX 30 index following ETF strategy returns, show superior performance against the parent ETF. Especially the RF Strategy overperforms the ETF Strategy and the parent ETF to the February – March 2020 COVID-19 crash, but in the recovery phase, the ETF Strategy passes it.

As a mandatory caveat, it must be stated that performing these outsider strategies in real markets would require active rebalancing with the created event portfolios and the parent ETF, which would accumulate the transaction costs significantly. However, these costs are difficult to quantify as the switches between the event portfolio's contents can be major or minor. Especially for the ETFs with consistently occurring selloff days, it is possible that no full portfolio rebalances, or a situation where all the outsider stocks are

sold and repurchased after the next event, are needed. Table 10 shows the excess returns for the strategy with the ETF ownership between the event dates:

EXCESS RETURNS FOR THE ETF STRATEGY (FF5 + BAB)

	α	βMKT	βSMB	βHML	βRMW	βCMA	βBAB	df	R²
IVV STRATEGY	0.0001	0.66***	-0.03	0.09**	0.24***	0.27***	0.33***	1155	0.56
	(0.72)	(0.00)	(0.44)	(0.02)	(0.00)	(0.00)	(0.00)		
	0.0001	0.69***	-0.12***	0.02	0.25***	0.33***		1155	0.52
	(0.65)	(0.00)	(0.00)	(0.65)	(0.00)	(0.00)			
QUAL STRATEGY	0.0000	0.79***	-0.08**	-0.02	0.31***	0.47***	0.32***	1155	0.76
	(0.86)	(0.00)	(0.01)	(0.40)	(0.00)	(0.00)	(0.00)		
	0.0000	0.82***	-0.17***	-0.09***	0.32***	0.53***		1155	0.72
	(0.98)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)			
ISF STRATEGY	0.0002	0.50***	-0.51***	-0.50***	0.21*	0.47***	-0.05	1159	0.45
	(0.26)	(0.00)	(0.00)	(0.00)	(0.06)	(0.00)	(0.15)		
	0.0002	0.50***	-0.53***	-0.49***	0.22**	0.46***		1159	0.45
	(0.30)	(0.00)	(0.00)	(0.00)	(0.05)	(0.00)			
EXS1 STRATEGY	0.0002	0.46***	-0.13	-0.11	0.10	-0.27*	-0.21***	1160	0.26
	(0.45)	(0.00)	(0.11)	(0.33)	(0.52)	(0.08)	(0.00)		
	0.0001	0.49***	-0.23***	-0.05	0.14	-0.30*		1160	0.25
	(0.62)	(0.00)	(0.01)	(0.66)	(0.39)	(0.05)			
SXRT STRATEGY	0.0001	0.75***	-0.79***	-0.29***	-0.17*	0.13	-0.06**	1174	0.77
	(0.38)	(0.00)	(0.00)	(0.00)	(0.05)	(0.13)	(0.02)		
	0.0001	0.76***	-0.82***	-0.27***	-0.16*	0.12		1174	0.77
	(0.46)	(0.00)	(0.00)	(0.00)	(0.07)	(0.16)			
SUAS STRATEGY	0.0001	0.72***	-0.09**	-0.08**	0.13**	0.66***	0.34***	1022	0.67
	(0.54)	(0.00)	(0.02)	(0.02)	(0.02)	(0.00)	(0.00)		
	0.0002	0.75***	-0.17***	-0.15***	0.11*	0.73***		1022	0.62
	(0.48)	(0.00)	(0.00)	(0.00)	(0.08)	(0.00)			
IESG STRATEGY	0.0002	0.63***	-0.36***	-0.64***	-0.16	0.58***	0.00	1184	0.52
	(0.35)	(0.00)	(0.00)	(0.00)	(0.14)	(0.00)	(0.91)		
	0.0002	0.63***	-0.36***	-0.64***	-0.16	0.58***		1184	0.52
	(0.35)	(0.00)	(0.00)	(0.00)	(0.14)	(0.00)			
IH2O STRATEGY	0.0001	0.79***	0.12*	0.01	0.32**	0.39**	0.30***	1194	0.39
	(0.62)	(0.00)	(0.09)	(0.91)	(0.01)	(0.01)	(0.00)		
	0.0002	0.94***	0.45***	-0.12*	0.08	0.79***		1194	0.58
	(0.39)	(0.00)	(0.00)	(0.09)	(0.43)	(0.00)			

Notes: Excess return of the strategy is regressed on Fama-French five factor model as follows: Rpi-Rfi=αp+β(p,MKT)*(RmiRfi)+β(p,SMB)*SMB+β(p,HML)*HML+β(p,RMW)*RMW+ β(p,CMA)*CMA+εp, where Rpi is the value-weighted combined event portfolio or parent ETF return on day i, Rfi is the 1-month T-bill rate converted to daily rate, Rmi is the market return on day i. The intercept α measures the average daily abnormal return of the strategy. P-values are in parentheses. FF5 daily factors are used for IVV, QUAL and SUAS. FF5 European daily factors are used for ISF, EXS1, SXRT and IESG. FF5 developed daily factors are used for IH2O. Please see table 1 for ETF symbol explanations. *** Denotes two-tailed significance at the 1% level, ** Denotes two-tailed significance at the 5% level, * Denotes two-tailed significance at the 10% level.

Table 10. Excess returns for the ETF Strategy, estimated with FF5 (+BAB).

By interpreting the results obtained, one can immediately notice that not a single ETF Strategy provided statistically significant alpha by the FF5 and FF5+BAB models for the sample period. However, all strategies provided positive alpha by sign, but not by statistical measures when the low beta-to-ETF strategy was implemented. Primarily the ISF Strategy provided 2bps (p = 0.26) daily alpha, which still missed fairly the barrier of 10% significance. Thus, the absolute alpha generating ability of the ETF Strategy cannot be confirmed by the FF5 + BAB estimates.

Furthermore, as the strategy returns deviate from the parent ETF returns, the sensitivity to the MKT factor deviates from one with all the strategies. One additional interesting detail to cover is that the strategies often take, large negative position against the SMB factor. The negative SMB loading is magnified for most of the strategies, apart from the IH2O strategy. This is explained that most of the funds examined in this paper are either broad-index or fundamental factor funds with a slight bias towards large-cap stocks. For the IH2O, most of the constituents are either mid or small-cap infrastructure stocks. However, it is curious that the low beta-to-ETF strategy further magnifies this bias in most cases, implying the outsider stocks in these ETFs tend to be larger in general.

For US and International ETFs, IVV, QUAL, SUAS, and IH2O, strategies are positively sensitive towards the BAB factor with statistical significance on the 1% level. As the low beta-to-ETF strategy does not take the opposite short high beta stocks approach, factor loadings closer to 1 were not expected. The fact why the European-based funds ISF, EXS1, SXRT, and IESG experienced either negative or close to zero loadings for the BAB factor can be explained by the lack of country-specific BAB factors used in this thesis.

Finally, Table 11 presents excess returns of the low beta-to-ETF strategy combined with the risk-free return in between the event periods (RF Strategy):

EXCESS RETURNS FOR THE RF STRATEGY (FF5 + BAB)

	α	βMKT	βSMB	βHML	βRMW	βCMA	βBAB	df	R²
IVV RF STRATEGY	0.0001	0.54***	-0.04	0.07**	0.19***	0.26***	0.46***	1155	0.49
	(0.80)	(0.00)	(0.33)	(0.04)	(0.00)	(0.00)	(0.00)		
	0.0001	0.57***	-0.16***	-0.03	0.20***	0.34***		1155	0.40
	(0.71)	(0.00)	(0.00)	(0.51)	(0.00)	(0.00)			
QUAL RF STRATEGY	-0.0001	0.65***	-0.11***	0.00	0.29***	0.46***	0.38***	1155	0.66
	(0.51)	(0.00)	(0.00)	(0.89)	(0.00)	(0.00)	(0.00)		
	-0.0001	0.67***	-0.21***	-0.08**	0.29***	0.53***		1155	0.59
	(0.65)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)			
ISF RF STRATEGY	0.0002	0.35***	-0.39***	-0.44***	0.26**	0.41***	-0.07**	1159	0.35
	(0.27)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.04)		
	0.0002	0.36***	-0.42***	-0.42***	0.27**	0.40***		1159	0.34
	(0.32)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)			
EXS1 RF STRATEGY	0.0002	0.25***	0.10	-0.28***	-0.11	-0.08	-0.11**	1160	0.09
	(0.53)	(0.00)	(0.17)	(0.00)	(0.41)	(0.54)	(0.01)		
	0.0001	0.26***	0.05	-0.25**	-0.10	-0.10		1160	0.09
	(0.64)	(0.00)	(0.46)	(0.01)	(0.49)	(0.46)			
SXRT RF STRATEGY	0.0002	0.39***	-0.38***	-0.41***	0.11	0.59***	-0.09***	1174	0.43
	(0.24)	(0.00)	(0.00)	(0.00)	(0.24)	(0.00)	(0.00)		
	0.0002	0.40***	-0.42***	-0.38***	0.13	0.57***		1174	0.43
	(0.32)	(0.00)	(0.00)	(0.00)	(0.18)	(0.00)			
SUAS RF STRATEGY	-0.0001	0.59***	-0.11***	-0.04	0.10*	0.54***	0.43***	1022	0.65
	(0.42)	(0.00)	(0.00)	(0.22)	(0.05)	(0.00)	(0.00)		
	-0.0001	0.63***	-0.21***	-0.13***	0.07	0.63***		1022	0.55
	(0.59)	(0.00)	(0.00)	(0.00)	(0.21)	(0.00)			
IESG RF STRATEGY	-0.0001	0.37***	-0.09*	-0.55***	-0.06	0.52***	-0.04	1184	0.28
	(0.44)	(0.00)	(0.08)	(0.00)	(0.56)	(0.00)	(0.21)		
	-0.0002	0.37***	-0.11**	-0.53***	-0.05	0.52***		1184	0.28
	(0.39)	(0.00)	(0.03)	(0.00)	(0.60)	(0.00)			
IH2O RF STRATEGY	0.0000	0.58***	0.10	0.06	0.36***	0.33**	0.46***	1194	0.38
	(0.88)	(0.00)	(0.13)	(0.48)	(0.00)	(0.02)	(0.00)		
	0.0002	0.79***	0.14*	-0.16*	0.37***	0.62***		1194	0.38
	(0.48)	(0.00)	(0.05)	(0.08)	(0.00)	(0.00)			

Notes: Excess return of the strategy is regressed on Fama-French five factor model as follows: $Rpi-Rfi=\alpha p+\beta(p,MKT)*(RmiRfi)+\beta(p,SMB)*SMB+\beta(p,HML)*HML+\beta(p,RMW)*RMW+\beta(p,CMA)*CMA+\varepsilon p$, where Rpi si the value-weighted combined event portfolio or risk-free return on day i, Rfi is the 1-month T-bill rate converted to daily rate, Rmi is the market return on day i. The intercept α measures the average daily abnormal return of the strategy. FF5 daily factors are used for IVV, QUAL and SUAS. FF5 European daily factors are used for ISF, EXS1, SXRT and IESG. FF5 developed daily factors are used for IH2O. P-values are in parentheses. Please see table 1 for ETF symbol explanations. *** Denotes two-tailed significance at the 1% level, ** Denotes two-tailed significance at the 5% level, * Denotes two-tailed significance at the 10% level.

Table 11. Excess returns for the ETF RF Strategy, estimated with FF5 (+BAB).

Similar to the previous strategy results, the RF Strategy provides no statistically significant alpha for any of the ETFs. However, in comparison to the ETF Strategy alphas, QUAL, SUAS and IESG fail to produce a positive sign for the alpha estimate, but as with all the previous intercepts, none of these are statistically significant. ISF and EXS1 provide statistically insignificant 2bps (p = 0.27) and 2bps (p = 0.53) of daily alpha. These results do not differ notably from the earlier ETF Strategy implemented.

As expected, the MKT factor loadings further decrease, as the ETF RF Strategy incorporates the risk-free return in between the event periods instead of equity (ETF) return. Like the previous strategy with the added ETF ownership, most of the RF Strategies take positive loadings for the CMA factor. BAB factor is again relatively large and statistically significant at the 1% level for the U.S.-based IVV, QUAL, and SUAS ETFs.

Overall, both strategies failed to produce statistically significant alpha with the FF5 estimation or the FF5 + BAB factor expansion. Thus, it is justified to reject the second hypothesis as statistically significant abnormal returns could not be captured long term. As a caveat, however, Lynch et al. (2019) paper investigated the amount of alpha "left on the table" by passive investors. Thus, this low beta-to-ETF strategy is not necessarily tailored to produce absolute alpha but to overperform relative to their parent ETFs. By these measures, with the higher Sharpe ratios produced, these strategies can prove to be useful for investors that are looking to capture more efficient risk-adjusted performance relative to these inspected ETFs, at least.

5.3. Evaluation of the fundamental weighting

In this final section of the empirical results, the usefulness of the fundamental proxy introduced earlier is evaluated. The motivation behind the fundamental weighting is to use a simple proxy for discretionary fundamental analysis for the low beta-to-ETF strategy. As Lynch et al. (2019) discussed, the low beta-to-ETF strategy could be enhanced with discretionary stock picking during the ETF selloffs. However, testing systematic trading strategies with combined discretionary analysis would provide dishonest results due to severe look-ahead bias. Thus, using a fundamental proxy of Piotroski F-Score by Piotroski (2000) is suggested in this paper. In the next table, the

results from the 1-40 day CAARs is presented for the low beta-to-ETF strategy with F-score portfolio weightings:

FUNDAMENTAL WEIGHTING

	Days after volume spike t			
	t+5	t+10	t+20	t+40
IVV	-0.003	-0.006	-0.011	0.001
	(0.54)	(0.15)	(0.47)	(0.68)
QUAL	-0.001	0.006	-0.004	0.006
	(0.99)	(0.15)	(0.23)	(0.16)
ISF	-0.002	0.002	0.006	0.022**
	(0.41)	(0.89)	(0.18)	(0.02)
EXS1	-0.002	0.001	0.000	0.004
	(0.75)	(0.97)	(0.93)	(0.67)
SXRT	0.003	-0.002	0.003	0.009
	(0.71)	(0.76)	(0.62)	(0.29)
SUAS	0.013**	0.015**	0.019***	0.026**
	(0.04)	(0.01)	(0.01)	(0.02)
IESG	0.000	0.002	0.008	0.015
	(0.36)	(0.38)	(0.27)	(0.12)
IH2O	-0.005	-0.012	-0.003	0.000
	(0.41)	(0.21)	(0.34)	(0.27)

Notes: This table presents the CAARs for each selected ETF by the periods of t+5, t+10, t+20 and t+40, t being the abnormal ETF volume event. The ETF-beta estimation period starts from 01/2015 and the full sample consists of 01/2016 - 08/2020. The Wilcoxon signed-rank test indicates significance levels of 0.05 (**) and 0.01 (***). Please see table 1 for ETF symbol explanations.

Table 12. Cumulative average abnormal returns for outsider portfolios created after ETF selloff day, Fundamental weighted.

Compared to the value-weighted results of the 1-40 day strategy, by weighting the selloff date low beta-to-ETF portfolios by the fundamental factor provides unsatisfactory results. Most of the 40-day CAARs decrease compared to the results in Tables 6 and 7, apart from IESG ETF, and the fundamental component brings no better statistical significance to the overall results. The only consistently positive CAAR with statistical significance is with the SUAS ETF, which had these same attributes with value-weighted portfolios.

The other well-performing outsider portfolio with value-weighted results was the ISF ETF which CAARs get slightly worse and less statistically significant. As seen, results thus far are disappointing and do not advocate for the F-Score weighting for forming the

outsider portfolios. However, the statistic in Table 12 suffers from the same COVID-19 related adversity as the 1-40 day statistics presented for the value and equal-weighted outsider portfolios in Tables 6 and 7.

Furthermore, making a similar COVID-19 adjustment to the F-Score weighted portfolios increases the generated cumulative 40-day average abnormal returns, but the results still lack the statistical significance, prevailing in the value-weighted CAARs (Table available in appendix 7). This means that despite the cumulative abnormal returns on average are larger, there still is enough randomness in the results that stating these average positive abnormal returns as meaningful is not justifiable.

To further contrast the F-Score weighted portfolios against the value-weighted portfolios, the following Graphs 15 and 16 present the CAARs of both portfolio weightings with and without the COVID-19 adjustment. Graph 15 shows the CAAR differences between the full samples:
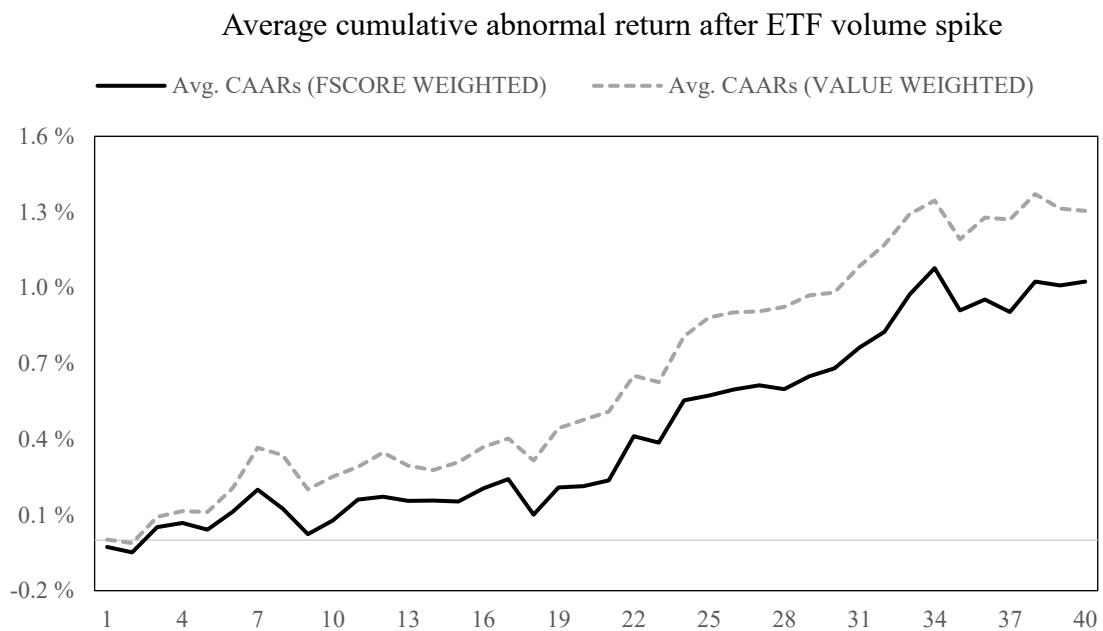
Graph 15. Average CAARs after ETF selloff days, F-Score weighted against Value weighted.

As seen in Graph 15, the value-weighted average CAARs are approximately 30 bps greater than the value-weighted portfolios after 40 days. Even the COVID-19 adjustment

does not change the setting, as F-score weighted average CAARs underperform the value-weighted results by 18 bps. Graph 16 demonstrates the difference between the COVID-19 adjusted portfolios:

Average cumulative abnormal return after ETF volume spike
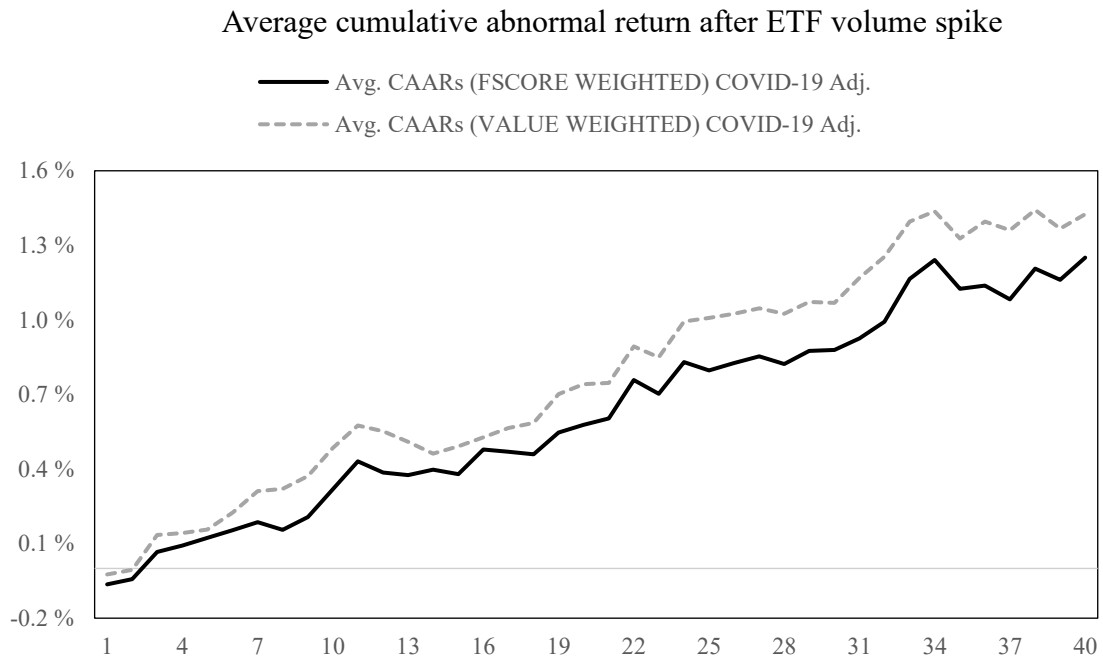
Graph 16. Average CAARs after ETF selloff days, COVID-19 adjusted F-Score weighted against COVID-19 adjusted Value weighted.

To summarize the findings and assess the third hypothesis, the F-Score weighting, or the fundamental component, does not enhance the stock-picking ability for the low beta-to-ETF strategy. Thus, the third hypothesis is rejected. The discretionary equity analysis suggested by Lynch et al. (2019) for enhanced returns is still the favored approach for investors looking for an added fundamental approach to the low beta-to-ETF strategy, as the backward-looking F-Score does not provide better results as a systematic proxy.

6. CONCLUSIONS

In this paper, the Lynch et al. (2019) study inspired low beta-to-ETF systematic trading strategy that exploits the ETF constituent stock co-movement after ETF selloffs, is tested with a unique and updated sample of U.S., European and International market tracking broad-index and fundamental factor ETFs. This paper investigated eight different ETFs with a combined AUM of $266 billion. The ETFs were investigated for the sample period of 01/2016 to 08/2020 to provide continuity and further look for the Lynch et al. (2019) study.

In the low beta-to-ETF strategy, stock betas are estimated against their parent ETFs and used to create a portfolio from the bottom 10% beta stocks. These portfolios are created after the trigger events of ETF selloff days, measured with increased daily ETF trading volume with combined negative ETF return for that day. With the value-weighted portfolios, the average abnormal returns produced by this strategy cumulated to 1.30% after 40-days. By excluding the February – March 2020 COVID-19 related events from the sample, the cumulative average abnormal returns further increased to 1.43% after 40-days.

Curiously, when investigating the behavior of the sample ETF constituent correlations during the abnormal ETF volume days in Graph 7 with the methods of Lynch et al. (2019), a similar pattern of heightened constituent cross-correlations could be captured for the broad-index funds, but not with the ESG and quality factor ETFs in the sample. These findings show that some constituents in these fundamentals themed funds stay largely uncorrelated even during abnormal volume events. This instead implies that these uncorrelated stocks could possess crash-risk protecting qualities during uncertain times (Kim et al. 2014).

Additionally, this paper further tests the usability of the low beta-to-ETF strategy by analyzing its performance when systematically utilized in the long-term. Thus, a single return time series with consistent outsider portfolio creation is made for each of the ETFs in which excess returns are then estimated with the Fama-French 5-factor model, with the additional Betting Against Beta (BAB) factor. These long-term strategies fail to generate statistically significant alpha but overperform almost unanimously against their parent ETFs when measured by Sharpe ratio. Consequently, systematically utilizing the low

beta-to-ETF strategy can yield investors better risk-adjusted performance in contrast to just passively investing with ETFs, but the strategies fail to generate pure alpha when measured by the FF5 model.

Finally, the advantages of the fundamental component to function as a proxy for discretionary stock analysis in the outsider portfolio creation phase are tested. The returns obtained with the Piotroski (2000) F-Score weighted low beta-to-ETF portfolios are almost without exception the same or worse than the value-weighted portfolio returns. This underperformance persists in the long-term analysis, and thus the Piotroski F-Score does not function as a useful fundamental proxy for the creation of these outsider portfolios and does not replace the discretionary fundamental equity analysis.

Suggestions for future research are numerous. As the time-structure of the low beta-to-ETF approach has been found to increase consistently towards the end of the holding period in this paper and Lynch et al. (2019) study, the starting points for the outsider portfolios could be lead forward and held longer after. The caveat for the longer held portfolios is that the clustering increases significantly, as selloff dates tend to repeat themselves regularly. This could be mitigated by increasing the selloff date requirement threshold either by demanding more than three standard deviation departure from the usual trading volume or adding a more significant negative return requirement for the high-volume days, as the return requirement was not appropriately motivated by the original Lynch et al. (2019) paper. Additionally, it could be analyzed whether the magnitude of the negative return for the selloff date impacts the profitability of the strategy portfolios.

Moreover, as the outsider strategy in half resembles the BAB strategy by Frazzini & Pedersen (2014), the strategy could be converted even more towards it by taking the opposite position on the high beta-to-ETF stocks. Moreover, focusing the ETF sample solely on the largest and most liquid ETFs without regional restrictions would prove to be an exciting addition to further research. Finally, as the AUMs and the trading volumes with the ESG factor ETFs are poised to increase in the future, the methods in this thesis and Lynch et al. (2019) paper should be used to investigate these thematical funds more closely, as the constituents in these ETFs can possess unique traits in terms of cross-correlations.

REFERENCES

Agarwal, V., Hanouna, P., Moussawi, R., & Stahel, C. (2018). Do ETFs Increase the Commonality in Liquidity of Underlying Stocks? *SSRN Electronic Journal*.

AQR (2020). *Betting Against Beta: Equity Factors, Daily*. Retrieved from: https://www.aqr.com/Insights/Datasets/Betting-Against-Beta-Equity-Factors-Daily

Azar, J., Schmalz, M. C., & Tecu, I. (2018). Anticompetitive Effects of Common Ownership. *The Journal of Finance*, *73*(4), 1513–1565.

Baker, M., Bradley, B., & Wurgler, J. (2011). Benchmarks as Limits to Arbitrage: Understanding the Low-Volatility Anomaly. *Financial Analysts Journal*, *67*(1), 40–54.

Barberis, N., & Shleifer, A. (2003). Style investing. *Journal of Financial Economics*, *68*(2), 161–199.

Barberis, N., Shleifer, A., & Wurgler, J. (2005). Comovement. *Journal of Financial Economics*.

Barberis, N., & Thaler, R. (2002). *A Survey of Behavioral Finance* (Working Paper No. 9222; Working Paper Series). National Bureau of Economic Research.

Bekaert, G., Ehrmann, M., Fratzscher, M., & Mehl, A. (2014). The Global Crisis and Equity Market Contagion. *The Journal of Finance*, *69*(6), 2597–2649. JSTOR.

Belasco, E., Finke, M., & Nanigian, D. (2012). The impact of passive investing on corporate valuations. *Managerial Finance*, *38*(11), 1067–1084.

Ben-David, I., Franzoni, F. & Moussawi, R. (2018). Do ETFs Increase Volatility? *Journal of Finance*, 73(6), pp. 2471-2535.

Bernard, V. L., & Thomas, J. K. (1990). Evidence that stock prices do not fully reflect the implications of current earnings for future earnings. *Journal of Accounting and Economics*, *13*(4), 305–340.

Black, F., Jensen, M. C., & Scholes, M. (1972). The capital asset pricing model: Some empirical tests. *Studies in the Theory of Capital Markets* 79–121.

Bloomberg (2019a). *End of Era: Passive Equity Funds Surpass Active in Epic Shift.* Retrieved from: https://www.bloomberg.com/news/articles/2019-09-11/passive-u-s-equity-funds-eclipse-active-in-epic-industry-shift

Bloomberg (2019b). *The Big Short's Michael Burry Explains Why Index Funds Are Like Subprime CDOs.* Retrieved from: https://www.bloomberg.com/news/articles/2019-09-04

Bodie, Z., Kane, A., & Marcus, A. J. (2009). *Investments* (8th ed.). McGraw-Hill/Irwin.

Bondt, W. F. M. D., & Thaler, R. (1985). Does the Stock Market Overreact? *The Journal of Finance*, *40*(3), 793–805. JSTOR.

Boone, A. L., & White, J. T. (2015). The effect of institutional ownership on firm transparency and information production. *Journal of Financial Economics*, *117*(3), 508–533.

Breitschwerdt, E. (2015). Point-In-Time vs. Lagged Fundamentals, This time it's different? *Quantamental Research*, 08/2015. S&P Capital IQ.

Brown, D. C., Davies, S., & Ringgenberg, M. (2020). *ETF Arbitrage, Non-Fundamental Demand, and Return Predictability* (SSRN Scholarly Paper ID 2872414). Social Science Research Network.

Burton, E., Shah, S., & Shah, S. (2013). *Behavioral Finance: Understanding the Social, Cognitive, and Economic Debates*. John Wiley & Sons, Incorporated. http://ebookcentral.proquest.com/lib/tritonia-ebooks/detail.action?docID=1158343

Cederburg, S., & O'Doherty, M. S. (2016). Does It Pay to Bet Against Beta? On the Conditional Performance of the Beta Anomaly. *The Journal of Finance*, *71*(2), 737–774.

Chen, H., Singal, V., & Whitelaw, R. F. (2016). Comovement revisited. *Journal of Financial Economics*, *121*(3), 624–644.

Christoffersen, S. E. K., & Simutin, M. (2017). On the Demand for High-Beta Stocks: Evidence from Mutual Funds. *The Review of Financial Studies*, *30*(8), 2596–2620.

Coles, J. L., Heath, D., & Ringgenberg, M. (2020). *On Index Investing* (SSRN Scholarly Paper ID 3055324). Social Science Research Network.

Connolly, R., Stivers, C., & Licheng Sun. (2005). Stock Market Uncertainty and the Stock-Bond Return Relation. *Journal of Financial & Quantitative Analysis*, *40*(1), 161–194.

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

Dimson, E. (1979). Risk measurement when shares are subject to infrequent trading. *Journal of Financial Economics*, *7*(2), 197–226.

Ernst, T. (2020). *Stock-Specific Price Discovery From ETFs*. 74.

ESRB (2019). *Can ETFs contribute to systemic risk?* Reports of the Advisory Scientific Committee No 9 / June 2019.

ETF database (2019). *ETFs with Apple Inc (AAPL) Exposure.* Retrieved from: https://etfdb.com/stock/AAPL/

ETFGI (2019). *ETFGI reports assets invested in the Global ETF and ETP industry reached a record US$5.64 trillion at the end of June 2019.* Retrieved from: https://etfgi.com/news/press-releases/2019/07/etfgi-reports-assets-invested-global-etf-and-etp-industry-reached

ETFStream (2020). *Morningstar: European ETF assets to hit €2trn by 2024.* Retrieved from: https://www.etfstream.com/features/morningstar-european-etf-assets-to-hit-2trn-by-2024/

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, *25*(2), 383–417. JSTOR.

Fama, E. F. (1998). Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics*, *49*(3), 283–306.

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

Fama, E. F., & French, K. R. (2010). Luck versus Skill in the Cross-Section of Mutual Fund Returns. *The Journal of Finance*, *65*(5), 1915–1947. JSTOR.

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

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

Ferri, R. A. & Phillips, D. (2009). *The ETF book: All you need to know about exchange-traded funds* (All-new, updated ed.). Hoboken, N.J.: John Wiley & Sons.

Fidelity (2011). *ETFs vs. mutual funds: cost comparison*. Retrieved from: https://www.fidelity.com/learning-center/investment-products/etf/etfs-cost-comparison

Financial Times (2020). *Index funds break through $10tn-in-assets mark amid active exodus*. Retrieved from https://www.ft.com/content/a7e20d96-318c-11ea-9703-eea0cae3f0de

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

French, K. R. (2008). Presidential Address: The Cost of Active Investing. *The Journal of Finance*, *63*(4), 1537–1573.

Froot, K. A., & Dabora, E. M. (1999). How are stock prices affected by the location of trade? *Journal of Financial Economics*, *53*(2), 189–216.

Gastineau, G. L. (2010). *The Exchange-Traded Funds Manual*.

Glosten, L. R., Nallareddy, S., & Zou, Y. (2016). *ETF Activity and Informational Efficiency of Underlying Securities* (SSRN Scholarly Paper ID 2846157). Social Science Research Network.

Greene, J. T., Hodges, C. W., & Rakowski, D. A. (2007). Daily mutual fund flows and redemption policies. *Journal of Banking & Finance*, *31*(12), 3822–3842.

Greenwood, R. M., & Sosner, N. (2007). Trading Patterns and Excess Comovement of Stock Returns. *Financial Analysts Journal*, *63*(5), 69–81.

Harris, L., & Gurel, E. (1986). Price and Volume Effects Associated with Changes in the S&P 500 List: New Evidence for the Existence of Price Pressures. *The Journal of Finance*, *41*(4), 815–829. JSTOR.

Heje Pedersen, L. (2018). Sharpening the Arithmetic of Active Management. *Financial Analysts Journal*, *74*(1), 21–36.

Hong, H., & Sraer, D. A. (2016). Speculative Betas. *The Journal of Finance*, *71*(5), 2095–2144.

Hou, K., Xue, C., & Zhang, L. (2020). Replicating Anomalies. *The Review of Financial Studies*, *33*(5), 2019–2133.

Index Industry Association (2019). *Index Industry Association's Third Annual Survey Finds 2.96 Million Indexes Globally*. Retrieved from: https://www.businesswire.com/news/home/20191015006005/en/Index-Industry-Association%E2%80%99s-Annual-Survey-Finds-2.96

Investment Company Institute (2020). *Investment Company Fact Book*. Retrieved from: https://www.icifactbook.org/ch4/20_fb_ch4

iShares (2020a). *WHAT IS SMART BETA?* Retrieved from: https://www.ishares.com/us/education/smart-beta

iShares (2020b). *iShares ETFs*. Retrieved from: https://www.justetf.com/en/etf-provider/ishares-etfs.html

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. JSTOR.

Kim, Y., Li, H., & Li, S. (2014). Corporate social responsibility and stock price crash risk. *Journal of Banking & Finance*, *43*, 1–13.

Lakonishok, J., Shleifer, A., & Vishny, R. W. (1994). Contrarian Investment, Extrapolation, and Risk. *The Journal of Finance*, *49*(5), 1541–1578. JSTOR.

Loviscek, A., Tang, H., & Xu, X. E. (2014). Do leveraged exchange-traded products deliver their stated multiples? *Journal of Banking & Finance*, *43*, 29–47.

Lynch, H., Page, S., Panariello, R. A., Jr, J. A. T., & Giroux, D. (2019). The Revenge of the Stock Pickers. *Financial Analysts Journal*, *75*(2), 34–43.

Lyon, J. D., Barber, B. M., & Tsai, C.-L. (1999). Improved Methods for Tests of Long-Run Abnormal Stock Returns. *The Journal of Finance*, *54*(1), 165–201. JSTOR.

Lyxor (2019). *Guide to pragmatic replication May 2019.* Retrieved from: https://www.lyxoretf.co.uk/pdfDocuments/Lyxor_Pragmatic%20replication%20guide.pdf

Markets Media (2020). *ESG ETFs Reach Record Assets.* Retrieved from: https://www.marketsmedia.com/esg-etfs-reach-record-assets/

Morningstar (2019a). *Morningstar's Annual Fee Study Finds That in 2018 Investors Paid Less to Own Funds Than Ever Before.* https://newsroom.morningstar.com/newsroom/news-archive/press-release-details/2019/Morningstars-Annual-Fee-Study-Finds-That-in-2018-Investors-Paid-Less-to-Own-Funds-Than-Ever-Before/default.aspx

Morningstar (2019b). *A Look at the Road to Asset Parity Between Passive and Active U.S. Funds.* Retrieved from: https://www.morningstar.com/insights/2019/06/12/asset-parity

Nasdaq (2019). *Assets Invested in ESG ETFs/ETPs Globally Reaches Record $52.35 Billion At End of November 2019.* Retrieved from: https://www.nasdaq.com/articles/assets-invested-in-esg-etfs-etps-globally-reaches-record-%2452.35-billion-at-end-of-november

Parsley, D., & Popper, H. (2020). Return comovement. *Journal of Banking & Finance*, *112*, 105223.

Pedersen, L. H. (2015). *Efficiently Inefficient: How Smart Money Invests and Market Prices Are Determined*. Princeton University Press.

Petajisto, A. (2011). The index premium and its hidden cost for index funds. *Journal of Empirical Finance*, *18*(2), 271–288.

Petajisto, A. (2017). Inefficiencies in the Pricing of Exchange-Traded Funds. *Financial Analysts Journal*, *73*(1), 24–54.

Piccotti, L. R. (2018). ETF Premiums and Liquidity Segmentation. *Financial Review*, *53*(1), 117–152.

Pionline (2019). *Investors continue to pay lower fees on mutual funds, ETFs – Morningstar.* Retrieved from: https://www.pionline.com/article/20190430/ONLINE/190439986/investors-continue-to-pay-lower-fees-on-mutual-funds-etfs-morningstar

Piotroski, J. D. (2000). Value Investing: The Use of Historical Financial Statement Information to Separate Winners from Losers. *Journal of Accounting Research*, *38*, 1.

Reuters (2017). *Index funds to surpass active fund assets in U.S. by 2024 -Moody's.* Retrieved from: https://de.reuters.com/article/funds-passive/index-funds-to-surpass-active-fund-assets-in-u-s-by-2024-moodys-idUKL1N1FN01O

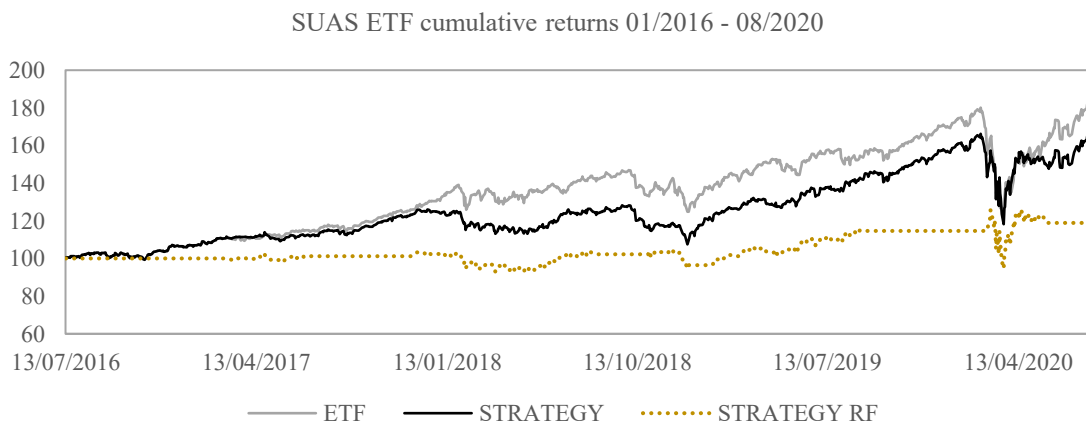Schneider, P., Wagner, C., & Zechner, J. (2020). Low-Risk Anomalies? *The Journal of Finance*, *75*(5), 2673–2718.

Sharpe, W. F. (1991). The Arithmetic of Active Management. *Financial Analysts Journal*, *47*(1), 7–9.

Shim, J. J. (2019). *Arbitrage Comovement* (SSRN Scholarly Paper ID 3287912). Social Science Research Network.

Shleifer, A., & Vishny, R. W. (1997). The Limits of Arbitrage. *The Journal of Finance*, *52*(1), 35–55. JSTOR.

Solnik, B., Boucrelle, C., & Yann Le Fur. (1996). International Market Correlation and Volatility. *Financial Analysts Journal*, *52*(5), 17–34.

Staer, A., & Sottile, P. (2018). Equivalent volume and comovement. *The Quarterly Review of Economics and Finance*, *68*, 143–157.

Sushko, V., & Turner, G. (2018). The implications of passive investing for securities markets. *BIS Quarterly Review: International Banking and Financial Market Developments*, 113–131.

Todorov, K. (2019). *Passive Funds Actively Affect Prices: Evidence from the Largest ETF Markets* (SSRN Scholarly Paper ID 3504016). Social Science Research Network.

UBS Research (2019). *Fund Managers Are More Moral Than You'd Think* Retrieved from: https://www.bloomberg.com/opinion/articles/2019-06-17/esg-adoption-shows-fund-managers-are-more-moral-than-you-d-think?sref=jwSJsa9K

Vanguard (2013). *Understanding synthetic ETFs*. Retrieved from: https://www.vanguard.com.hk/documents/understand-synthetic-etfs-tlrv.pdf

Vijh, A. M. (1994). S&P 500 Trading Strategies and Stock Betas. *The Review of Financial Studies*, *7*(1), 215–251. JSTOR.

Wurgler, J. (2010). *On the Economic Consequences of Index-Linked Investing*. 15.

Yu, F. (Frank). (2008). Analyst coverage and earnings management. *Journal of Financial Economics*, *88*(2), 245–271.
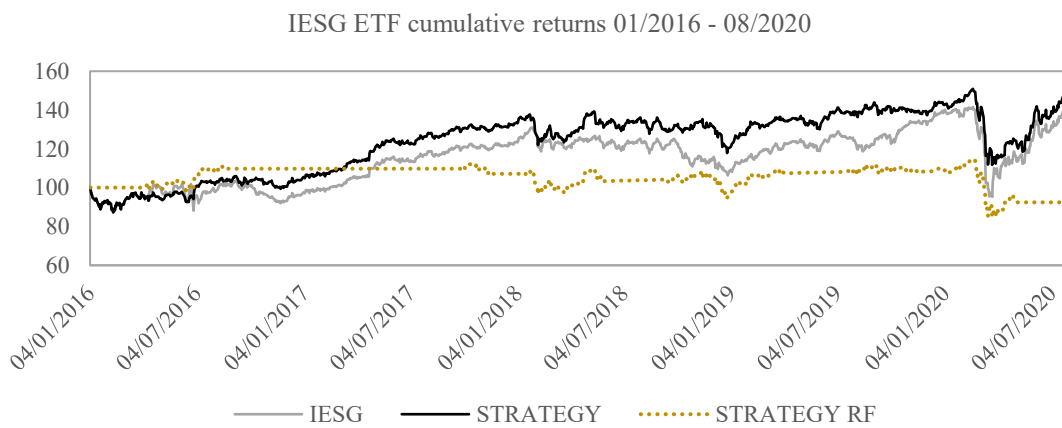
APPENDICES

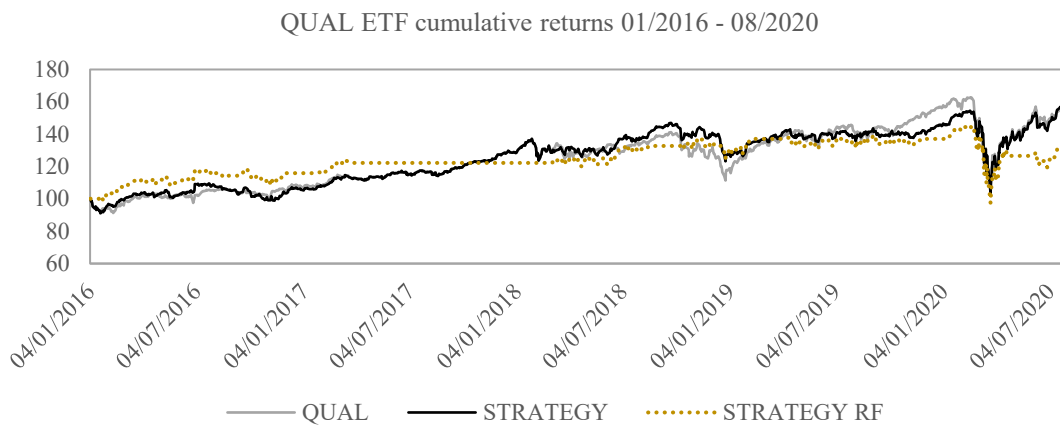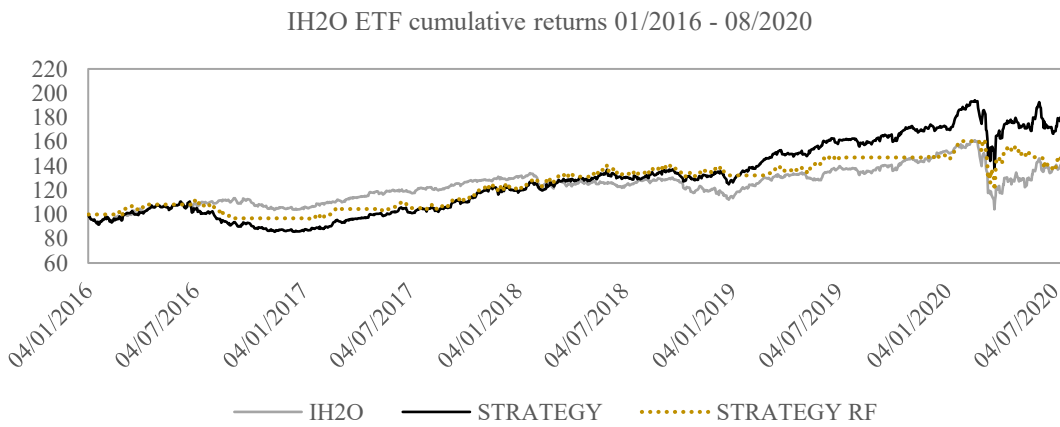**Appendix 1.** IVV ETF Cumulative returns

IVV ETF cumulative returns 01/2016 - 08/2020

ETF — STRATEGY ······ STRATEGY RF
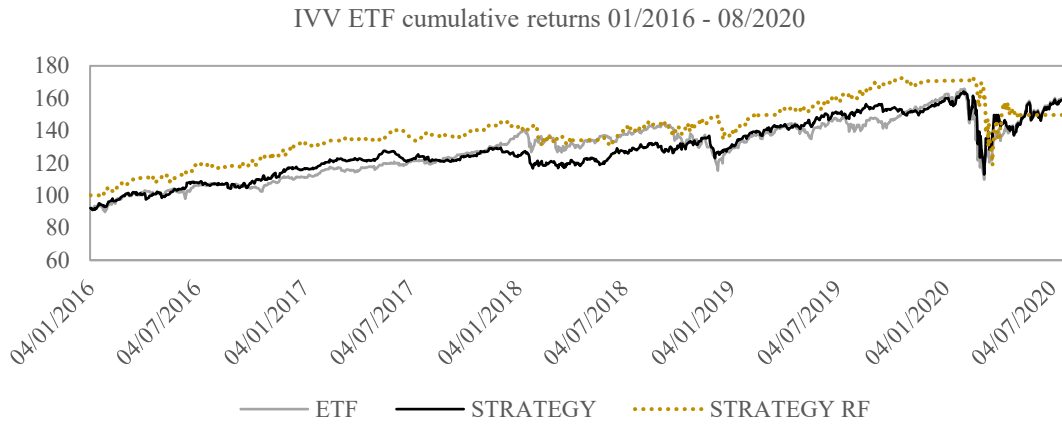
Appendix 2. SUAS ETF Cumulative return

SUAS ETF cumulative returns 01/2016 - 08/2020

ETF — STRATEGY ······ STRATEGY RF

Appendix 3. IESG ETF Cumulative returns

IESG ETF cumulative returns 01/2016 - 08/2020

IESG — STRATEGY ······ STRATEGY RF

Appendix 4. QUAL ETF Cumulative returns

QUAL ETF cumulative returns 01/2016 - 08/2020

Appendix 5. IH2O ETF Cumulative return

IH2O ETF cumulative returns 01/2016 - 08/2020

Appendix 6. SXRT ETF Cumulative returns

SXRT ETF cumulative returns 01/2016 - 08/2020
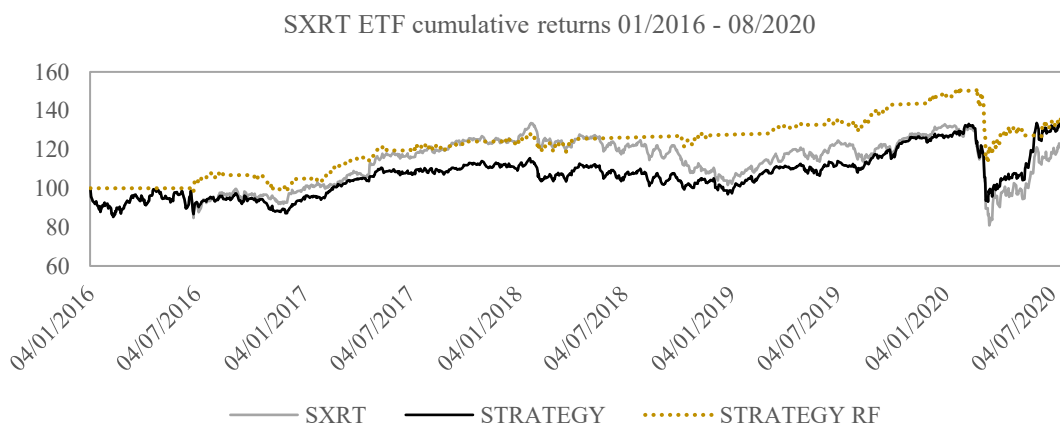
Appendix 7. COVID-19 Adjusted Fundamental Weighting CAARs

Average cumulative abnormal returns for outsider portfolios after abnormal ETF volume spike

	COVID-19 ADJUSTED FUNDAMENTAL WEIGHTING			
	Days after volume spike t			
	t+5	t+10	t+20	t+40
IVV	-0.001	-0.002	-0.003	0.007
	(0.99)	(0.50)	(0.59)	(0.40)
QUAL	0.002	0.008	0.008	0.015
	(0.71)	(0.07)	(0.10)	(0.07)
ISF	0.001	0.006	0.011	0.020**
	(0.79)	(0.71)	(0.11)	(0.03)
EXS1	0.002	0.006	0.005	0.009
	(0.79)	(0.48)	(0.46)	(0.35)
SXRT	0.001	0.004	0.007	0.011
	(0.76)	(0.38)	(0.31)	(0.16)
SUAS	0.008	0.012**	0.015**	0.020
	(0.14)	(0.04)	(0.02)	(0.06)
IESG	-0.001	-0.002	0.004	0.018
	(0.56)	(0.62)	(0.81)	(0.17)
IH2O	-0.004	-0.006	0.000	0.000
	(0.92)	(0.98)	(0.39)	(0.55)

Notes: This table presents the CAARs for each selected ETF by the periods of t+5, t+10, t+20 and t+40, t being the abnormal ETF volume event. The ETF-beta estimation period starts from 01/2015 and the full sample consists of 01/2016 - 08/2020. The Wilcoxon signed-rank test indicates significance levels of 0.05 (**) and 0.01 (***).

Appendix 8. Orbis F-Score variable codes

```
PROFITABILITY EU = (IF(ROA>#0,#1,#0))+(IF(15514>#0,#1,#0))+(IF(ROA[N]>ROA[N-
1],#1,#0))+(IF((15514/TOAS)*#100>ROA,#1,#0))


PROFITABILITY US = (IF(ROA>#0,#1,#0))+(IF(OTLO>#0,#1,#0))+(IF(ROA[N]>ROA[N-
1],#1,#0))+(IF((OTLO/TOAS)*#100>ROA,#1,#0))


LEVERAGE = IF((LTDB[N])/(TOAS[N])<(LTDB[N-1])/(TOAS[N-
1]),#1,#0)+IF((CURR[N]>CURR[N-
1]),#1,#0)+IF((ASTK_SHARES_OUT[N]<=ASTK_SHARES_OUT[N-1]),#1,#0)


EFFICIENCY = IF((GRMA[N]>GRMA[N-1]),#1,#0)+IF((NAT[N]>NAT[N-1]),#1,#0)
```