SSE Riga Student Research Papers
2020 : 2 (224)

THE GROWTH OF SMART BETA ETFS: IMPLICATIONS FOR MARKET EFFICIENCY AND FACTOR PREMIA

Authors: Angelina Lučinskaja
Evelina Lučinskaja

ISSN 1691-4643
ISBN 978-9984-822-46-4

September 2020
Riga

The Growth of Smart Beta ETFs:
Implications for Market Efficiency and Factor Premia

Angelina Lučinskaja

and

Evelina Lučinskaja

Supervisor: Tālis J. Putniņš

September 2020
Riga

Table of Contents

Abstract

The rapid growth of Smart Beta (SB) Exchange Traded Funds (ETFs) – new investment vehicles that allow investors to gain exposure to various asset pricing factors – have raised substantial concerns about their impact on asset prices. In this thesis, we analyze the effect of SB ETFs on the quantity and efficiency of factor information in stock prices and on factor return premia. Our analysis uses SB ETFs from 2012 to 2019 with exposure to five well-known factors (value, size, momentum, low volatility, and high dividend yield). We find that SB ETFs tend to increase the informational efficiency of factors, making their returns more difficult to predict. We also find that SB ETFs tend to increase the extent to which stock returns are driven by factor returns, suggesting an increase in co-movement of stock returns. However, we find no evidence that SB ETFs attenuate factor premia. These effects are more significant for factors with higher traded dollar volumes in SB ETFs. Thus, this thesis concludes that further growth of SB ETFs is likely to strengthen their impact on asset prices, which is an area for future research.

Keywords: smart beta, exchange traded funds, factor investing, informational efficiency.

Acknowledgments

We would like to thank our supervisor Tālis J. Putniņš for the guidance, expert advice and invaluable input throughout the thesis writing process.

1. Introduction

Unconventional investment vehicles – Exchange-Traded Funds (ETFs) – recently received a considerable amount of attention due to the many advantages they present for investors. Some of the benefits are transparency, liquidity, low fees, and variety: an opportunity to trade different assets (commodities, stocks, etc.), use multiple strategies (passive, active, smart beta, etc.), or reach a broader range of regions (Lettau & Madhavan, 2018). Most commonly, like a passive investing vehicle (similar to a mutual fund), an ETF tracks a particular stock index. The manager of an ETF uses Authorized Participants (AP) (e.g., large financial institutions) instead of directly interacting with markets (i.e., investors). An ETF manager issues or redeems shares in exchange for a basket of securities or cash from AP. Therefore, using this mechanism, shares outstanding in the market are based on the demand and supply (Lettau & Madhavan, 2018).

The popularity of passive investing has increased significantly in the recent past. Consequently, investors' willingness to outperform the market gave rise to smart beta (SB) or factor investing strategies, which lie between passive and active investing. Thus, SB ETFs' assets under management (AUM) have risen from approximately $350bn in 2015 to more than $1tn in 2019 (Rabener, 2019). In the US alone, AUM of all US-traded ETFs is around $800bn in 2020, 25% of which are SB ETFs (Rabener, 2020). SB ETFs use different portfolio asset weighting strategies than conventional ETFs – putting more weights on stocks with higher exposure to specific characteristics (called asset pricing factors). Widespread asset pricing factors include value (stocks with higher book-to-market ratio give higher returns), size (stocks of small firms provide higher returns than stocks of big firms), etc. (Lettau & Madhavan, 2018). We analyze value, size, momentum, high dividend yield, and low volatility factors.

Despite phenomenal growth as an investment product, SB ETFs are still under-researched in the academic literature. It has never been so easy to trade factors (i.e., pick stocks or stock indexes that try to capture factor premiums), which provides the ground for researching SB ETFs and their implications. Thus, we test SB ETFs implications for the factor efficiency, stock returns co-movement, and premia attenuation, which are three widely discussed topics in finance.

Our first topic is about factor efficiency (or the efficiency of factor information, or informational efficiency), which we define as the ability to predict future factor returns using current and past information. The popularity of SB ETFs and, consequently, the decrease in the transaction costs makes it easier and cheaper to trade factors, which gives a ground for exploiting inefficiencies. Scholars do not agree about the effect of higher ETFs trading on informational efficiency. Thus, our research supplements the literature by answering the question of whether SB ETFs trading can improve the weak and semi-strong efficiency of factor returns. The efficiency of factor returns is important for portfolio creation – investors can profit from inefficiencies or arbitrageurs can exploit uninformed investors (Asness, 2016a; Chordia et al., 2006).

In our second analysis, we study to what extent factor returns explain stock returns, which has implications for the excess co-movement of stocks (i.e., co-movement not associated with fundamentals). With higher stocks' exposure to a particular factor, the co-movement of stock also tends to increase. That is why excess co-movement can be associated with ETFs. In essence, co-movement occurs when more investors are investing in the same strategies generating higher returns, which usually focus on similar stock characteristics. The abovementioned reasons give a ground to analyze to what extent factor returns explain stock prices and how the rise of SB ETFs impacts that. When diversifying their portfolios, investors have to take into account the correlations of stock returns because stock returns co-movement makes all stocks exposed to similar shocks (Ando, 2019).

The last topic is about factor premia attenuation. Many authors argue that the publication or increase in popularity of the factors causes a crowding effect, which decreases the factor premia and consequently, the profitability of the factor trading strategies (Novy-Marx & Velikov, 2015; Arnott et al., 2019). SB ETFs made it easier for investors to trade factors, raising a concern that these funds might lose their market-beating strategies (Riding, 2018). Therefore, our last analysis tests the effect of the increase in SB ETFs trading on the factor premia.

Our analysis seeks to decrease a mismatch between investors' expectations and information available about SB ETFs. The novelty of our research lies in certain aspects. In short, we link three widely discussed topics in finance with the growth of SB ETFs, which has not been done yet for any of the topics. Firstly, we connect the informational efficiency of factor returns and factor premia with the growth of SB ETFs to see the implications of these investment vehicles. Moreover, we are not just concerned about whether stocks move together but want to see the richness of the factor information in prices. Thus, we are testing to what extent factor returns drive stock returns and how SB ETFs have impacted it. It is a multivariate form of co-movement, which therefore provides a richer story about the nature of correlations between stocks.

The three research questions we attempt to answer are the following:

1. **How have SB ETFs impacted the informational efficiency of factor returns?**
2. **How have SB ETFs impacted the extent to which stock returns are driven by factor returns?**
3. **Have SB ETFs attenuated factor return premia?**

We find that SB ETFs tend to strengthen the informational efficiency of factor returns, indicating lesser factor returns predictability. Moreover, SB ETFs tend to increase the extent to which stock returns are driven by factor returns, which implies the increase in stock returns co-movement. We find no evidence that SB ETFs attenuate factor premia. The effects are more potent for factors with more trading volume of SB ETFs. Therefore, our findings emphasize the implications of SB ETFs, which are essential for portfolio formation.

The thesis is structured as follows[1]: Section 2 describes the relevant literature and derives the hypothesis. Section 3 describes the data and data sources used in the analysis. Sections 4, 5, and 6 describe the methods used to do the analysis, present and discuss the empirical results of factor efficiency, weights of factor information in stock prices, and factor premia, respectively. Section 7 is for robustness tests. Section 8 describes the limitations of our analysis. Section 9 concludes.

2. Literature review

The literature review is structured as follows: firstly, we describe how markets and constituent stocks are affected by ETFs in general. Then we summarize the existing literature about SB ETFs and describe factors used in our analysis. Afterwards, we briefly discuss the existing literature on the three topics, which are the focus of the thesis: informational efficiency, the extent to which factors drive stock returns, and factor premia. We derive the hypotheses of each topic at the end of the respective section. We encourage the reader to pay attention to whether we are referring to the literature about stock returns or factor returns, which, in our view, may create confusion.

2.1. ETFs

Empirical evidence shows that an increase in ETFs trading leads to the improvement in informational efficiency (Glosten et al., 2016; contrary to Israeli et al., 2017). Stocks with higher ETFs activity better and timelier incorporate earnings news in the current stock prices.

[1] We have choosen the unusual thesis structure to clearly separate three topics. A broad scope of the thesis may create confusion; thus, by changing the structure, we tried to make the thesis more reader-friendly.

On the other hand, the weights of stocks in an ETF are determined mechanically; thus, the incorporation of information may be distorted based on the respective weight (Glosten et al., 2016).

ETFs play a positive role in the price discovery process of underlying stocks during the times of scarce liquidity (Madhavan & Sobczyk, 2014). Ben-David et al. (2017) argue that ETFs may have twofold effects on the underlying securities. On the one hand, information is impounded better into stock prices due to liquidity effects caused by ETFs activity. On the other hand, non-fundamental activities (such as rebalancing) strengthen the mispricing.

Ben-David et al. (2018) argue that ETFs activity increases the volatility of constituent stocks and causes more noise in the market (similarly to Broman & Shum, 2016; Xu & Yin, 2017). Higher trading increases liquidity, which creates demand shocks and increases volatility. Therefore, excess volatility created by ETFs ownership exposes investors to a non-diversifiable systematic risk. Moreover, since with ETFs trading investors are switching to strategies that deliver higher excess returns, the stocks start co-moving more than they are fundamentally supposed to (Broman, 2016; Da & Shive, 2018). Therefore, ETFs trading has implications on stock markets' efficiency, liquidity, volatility, price discovery, co-movement and other variables.

2.2. Smart beta ETFs

SB ETFs' managers use a different portfolio weighting strategy - more weight is put on stocks with higher factor exposure. Investors can benefit by investing in actively picked stocks but no longer have to pay high fees for active funds' managers. In 2019, more than 77 ETFs came to the market, which is approximately one-third of all ETFs launches (Murphy, 2019). At the end of 2018, AUM of SB ETFs with the exposure to growth factor had the highest AUM followed by SB ETFs with the exposure to value, low volatility, momentum, and quality factors (212, 186, 29, 13, 7 $billions respectively) (Rabener, 2019).

The rising popularity of SB ETFs might have negative implications. Arnott, Beck, Kalesnik, et al. (2016) argue that investors face certain risks by chasing the performance of factor investing strategies generating alphas. Firstly, rising valuations of stocks, sectors, asset classes or strategies lead to the illusionary magnified past performance. Subsequently, high valuations reduce future returns, which increases the chance of mean reversion to the historical levels. In short, the strategies' success lies in the fact that factors are increasing in price, meaning that alphas are generated purely from rising valuations. Due to the reasons mentioned above, Arnott, Beck, Kalesnik, et al. (2016) forecast a chance of SB crash (similarly to Asness, 2016b). Given the fact that SB ETFs' AUM skyrocketed in the past few years, this is supposedly a red flag for these investment vehicles. Moreover, many authors find no evidence of SB strategies' outperformance (Burton, 2014; Glushkov, 2016, etc.).

Arnott, Beck, and Kalesnik (2016) explain that SB strategies can go from "horribly wrong" to "beautifully right" (p. 10). Investors can generate excess returns if they put more emphasis on strategies that are trading relatively cheap compared to their historical norms and less emphasis on relatively expensive strategies. The authors argue that a portfolio of a few worse-performing and cheap factors can outperform an equally weighted portfolio and generate stable future returns (similarly to Wes & Pickard, 2019).

2.3. Asset pricing factors

In our research, we focus on five different types of SB ETFs. Each type of SB ETF tries to capture a specific asset pricing factor to generate extra returns. In the following sections, we describe each of the factors, which are also called risk factors or anomalies.

***Size Factor*.** Banz (1981) finds that the correlation between common stock returns and firm size is negative, implying that bigger firms have lower risk-adjusted returns. This effect is outstanding for the smallest firms; the return difference of average size and large firms is marginal. The concept of size effect became wider recognized when Fama and French (1993) found that size and value factors in addition to the market risk premia capture stock returns variation. Investing in size factor means picking stocks with smaller market capitalization to achieve higher returns which these stocks deliver.

Value Factor. High book to market ratio firms are value firms, while firms with low book to market ratio are growth firms. Low book to market ratio shows that the company has a high market value, which means that investors have positive prospects about the company's future, as a consequence, the stock has a higher price, which corresponds to lower expected returns (Rosenberg et al., 1985; Fama & French, 1992). Thus, investors can earn excess returns by buying value stocks with high book to market ratio and selling growth stocks with low book to market ratio.

***Momentum factor*.** Jegadeesh and Titman propose a profitable strategy based on the continuity of the performance of stock price: buying top-performing stocks aka "winners" and selling low performing stocks aka "losers" (p. 65). According to Fama and French (2004), "stocks that do well relative to the market over the last three to twelve months tend to continue to do well for the next few months, and stocks that do poorly continue to do poorly" (p. 40).

***Low volatility*.** Low-volatility investing focuses on picking stocks with historically low returns fluctuations (Hsu & Li, 2013). Classical financial axioms indicate that higher risks correspond to higher returns, that is, expected excess stock returns are proportional to their betas (Black et al., 1972). However, empirical evidence contradicts that: stocks having low volatility tend to have higher returns than the stocks bearing higher risks (Hsu & Li, 2013).

***Dividend*.** Stocks that have higher expected dividend-yield give higher returns than non-dividend paying stocks (Blume, 1980; Litzenberger & Ramaswamy, 1979). The strategy suggests increasing stockholding in the portfolio if the stock is expected to pay higher dividends. Although it is contradictory to the fact that dividends are taxed, which reduces stock returns, investors believe that on risk-adjusted basis returns are still higher even with the dividend tax (Blume, 1980).

2.4. Efficient market hypothesis

Fama (1965) introduce the efficient market hypothesis (EMH) that states that asset prices incorporate all available information; consequently, a current stock price reflects the fair fundamental value of a stock. Three forms of market efficiency exist. Weak-form efficiency states that asset prices follow a random walk - prices incorporate only past information; semi-strong market efficiency states that all past and currently available public information is reflected in prices; and under strong market efficiency, public and non-public information is incorporated into prices. Non-public information is unobservable and can be tested only using proxies (Fama, 1970). Thus, in our research, we focus on weak and semi-strong factor efficiency.

The EMH received many pushbacks: joint-hypothesis problem (Fama, 2014; similarly to Brenner, 1979), certain conditions do not hold in real life (e.g., no transaction costs) (Fama, 1970), etc. Thus, Fama (1991) modified the EMH to make it closer to the real financial markets. For weak-form efficiency tests, instead of focusing on the ability to forecast stock returns, researchers or investors should focus on returns predictability and cross-sectional returns predictability (on which we focus in our efficiency analysis).

2.5. Weak-form efficiency

Arnott et al. (2018) argue that factor momentum is a feature of universal factors, which implies inefficiency. Existence of factor momentum is essential for investors: buying factors that are performing the best and selling the worst-performing factors generates more exceptional performance than stock momentum strategy (Gupta & Kelly, 2019). Thus, factor efficiency has implications on the profitability of factor investing strategies (previous studies conducted by Schwert (2003) and Mclean & Pontiff (2016) discuss the arbitrageurs' activities in exploiting profitable opportunities). By timing factors, investors can generate economically and statistically better excess performance (Asness, 2016a; Asness, 2016b; contrary to Ilmanen et al., 2019). Factor timing can be defined as owning a factor when its return is higher than its average return and not owning the factor when it is the opposite.

Empirical evidence shows a positive relation between the increase in ETFs trading volume and the incorporation of news into prices (Glosten et al., 2016, similarly to Ben-David et al., 2017). Similarly, an increase in the active investing of hedge funds increased price convergence and market efficiency (Stulz, 2007). On the other hand, Israeli et al. (2017) findings suggest that a rise in ETFs trading contributes to a decline in the pricing efficiency of underlying securities. A rise in ETFs holdings leads to a decrease in liquidity and, subsequently, an increase in trading costs. As a result, fewer investors trade the firm-specific information, which leads to the overall worsening of the informational efficiency. Contrary, Hamm (2014) finds that ETFs increase stocks' liquidity. Chordia et al. (2006) argue that markets are the most efficient during the periods of high liquidity (similarly to Chung & Hrazdil, 2010), implying that factor efficiency may increase with more SB ETFs trading.

Thus, all these findings imply that an increase in factor-investing could have implications for the efficiency of factor returns and consequently, for the profitability of factor investing strategies. By testing weak-form predictability of factor returns, we can see how the extent of predictability relates to or changes with the growth of SB ETFs. Growth of SB ETFs made it easier and cheaper to trade factors, making it more profitable to exploit inefficiencies in factor pricing. Consequently, more effort devoted to exploiting such inefficiencies potentially reduces the inefficiencies. Therefore, we arrive at our first hypothesis: **the emergence and growth of SB ETFs improves the weak-form efficiency of factor returns.**

2.6. Semi-strong efficiency

Understanding the predictability of factor returns using public information can help investors to improve factor investing strategies (Asness et al., 2000). For example, investors usually try to hedge against specific macroeconomic shocks and require higher compensation if assets are exposed to these shocks (Israeli et al., 2019). Moreover, different factors have different cash flows duration; consequently, they react differently to interest rates (Lettau & Wachter, 2007; Gormsen & Lazarus, 2019). For example, **value** and defensive factors may be more sensitive to inflation shocks because they have a shorter duration. Furthermore, value strategy tends to underperform in financial distress since it heavily relies on the capital structure of companies (Berk et al., 1999). Contrary, the properties of **low volatility** strategy allow it to outperform during bad times (Ang et al., 2006). **Momentum** strategy tends to underperform during market turmoil – the periods of market underperformance and high volatility (Daniel & Moskowitz, 2016; Cooper et al., 2004; Stivers & Sun, 2010). Moreover, after controlling for firm size and systematic risk, Gombola and Liu (1993) find that **dividend yield** and stock returns move in the same direction in bear markets and the opposite direction in bull markets. Similarly, Perez-Quiros and Timmerman (2000) argue that due to less collateralization, **small firms** have tighter access to external funds. As a consequence, they have a higher sensitivity to the liquidity and short-term interest rates; that is, they are more exposed to tighter credit market conditions.

Many authors test the predictability of **factor returns** (Ilmanen et al., 2019; Asness et al., 2013; Hodges et al., 2017; Baltussen et al., 2019). Ilmanen et al. (2019) find no evidence of predictability of economic macro market variables on **factor returns**. While any predictability implies market inefficiency, predictability could also be the result of time-varying risk premiums rather than mispricing. Griffin et al. (2003), Cooper et al. (2004) show that macro factors do not explain the returns of momentum factor, which contradicts the findings of Chordia and Shivakumar (2002). Additionally, Baltussen et al. (2019) find that factor premiums tend to remain the same in the recessionary and expansionary periods. Moreover, there is no statistically significant relation of global factor returns (i.e., pooled factors) and macroeconomic factors. Lastly, macroeconomic factors have marginal explanatory power on factor payoffs. Thus, academicians find contradicting evidence regarding the predictive power of market macro variables.

Therefore, similarly to weak-form predictability of factor returns, we want to test how the extent of macro market variables' predictability on factor returns changes with the growth of SB ETFs. With an easier factor trading, investors can profit from the predictability of factor returns. That catches the interest of investors; thus, more of them are trying to exploit inefficiencies (i.e., predictability of factor returns using public information). Consequently, more trading potentially leads to a decrease in inefficiencies. Thus, we arrive at our second hypothesis: **the emergence and growth in SB ETFs weakens the predictability of factor returns using public information, implying an improvement in semi-strong informational efficiency of factor returns.**

2.7. Factor information in stock prices and stock returns co-movement

Previous literature outlines how factor information is incorporated into stock prices. In essence, similar stock characteristics (such as high book to market ratio) explain the sensitivity of stock returns to common risk factors (such as size, value, etc.) (Fama 1970, 1991; Jegadeesh & Titman, 1993). Consequently, stocks with exposure to the same factor are likely to move together. Daniel and Titman (1996) find evidence that not the factor loadings (that is, not the covariance structure of returns) but the similar nature of stocks affects stock returns. It is hard to determine one common conclusion of how stock returns co-move due to the existence of numerous models in the literature. The logic of these models is that factors and returns' covariance explain excess stock returns and that the existence of factors itself contributes to the higher stock returns co-movement (Kozak et al., 2018).

Barberis et al. (2005) explain that stock returns co-movement is associated with co-movement in the fundamentals of common stocks, at least that should be the case in a frictionless market. However, investors' irrationality and frictions in markets cause co-movements associated with non-fundamentals (friction or sentiment-based views) (Vijh, 1994; Barberis & Shleifer, 2003). For example, due to correlated investors' sentiment, co-movement of stocks may increase if investors are trading stocks in a particular category (e.g., risk factors, junk bonds, etc.).

The liquidity increase associated with the growth of ETFs contributed to the excess co-movement of stock returns (Broman, 2016). Da and Shive (2018) find evidence that high volume of ETFs trading causes excessive movement of prices – overshooting and price reversals. Moreover, arbitrageurs can cause non-fundamental shocks (i.e., excess co-movement) of common stock returns by trading against ETFs strategies. Lastly, the excess co-movement occurs when investors are switching towards strategies delivering higher excess returns – the demand for investment styles is correlated (Broman, 2016).

Thus, with higher SB ETFs trading activity, investors are putting more capital to the same trading strategies, which target stocks with similar characteristics – higher factor exposure. In recent years, investors put more than ever focus on strategies with higher factor exposure, which potentially increases the extent to which factor returns drive stock returns. Consequently, that leads to higher co-movement of stocks. Moreover, ETFs trading, in general, have contributed to the excess stock co-movements. Due to the reasons mentioned above, we are interested in checking how the extent to which stock returns are driven by factor returns have been impacted by the growth of SB ETFs. Therefore, we hypothesize that **the emergence and growth in SB ETFs increase the extent to which stock returns are driven by factors**.

2.8. Factors premia attenuation

Arnott et al. (2019) outline several reasons why investors may face the underperformance of a factor: data mining or crowding, tail behavior, and correlations of factors. The publication of the factor informs investors about mispricing, which causes crowding and consequently decreases premia (McLean & Pontiff, 2016; Arnott, Beck, Kalesnik, et al., 2016; Arnott et al., 2019; Hanson & Sunderam, 2013). Arnott et al. (2019) find that factor returns of the best-known factors deteriorated the most in the last 15 years, which proves that the popularity of the factors causes crowding that leads to the decrease in factor premia. Moreover, transaction costs decrease the potential benefit of trading factors (Novy-Marx & Velikov, 2015).

Ilmanen et al. (2019) test a variation of factor premia and find that single factor premia and volatility are not constant throughout the time. However, the authors find no evidence supporting that arbitrageurs cause this variation and no evidence of the decrease in factor premia. Hanson and Sunderam (2013) argue that due to limits to arbitrage (specifically, short selling), the excess returns of factor trading strategies are unlikely to be entirely eliminated. Additionally, Asness (2015) argues that factor investing strategies are not as successful as in the past. However, they have not disappeared because investors are compensated for higher risk of factor investing, and due to errors made by investors (such as mispricing, excessive reaction, behavioral biases, etc.).

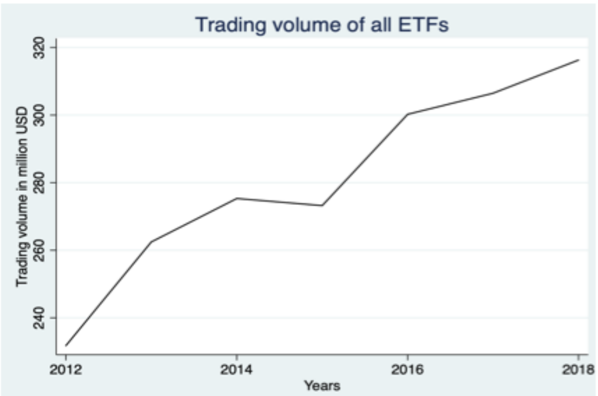
With less trading, factor premia may persist for a long time. However, SB ETFs made it cheaper and easier to capture the profits of factor premia, encouraging more people to harvest them, which potentially decreases the premia. Capturing a factor premium (e.g., size factor) involves buying the stocks that will have high future returns (e.g., small stocks) and shorting or just underweighting the stocks that will have relatively lower future returns (in this case the big stocks). That pushes up the price of the cheap stocks (small stocks) relative to the expensive stocks (big stocks), thereby making the difference in their future returns shrink, that is, attenuating the factor premia. Therefore, we hypothesize **that the emergence and growth in SB ETFs attenuates (reduces) factor return premia.**

3. Data selection

We download factor returns data on daily and monthly frequency for the period from 1963 January 3rd to 2019 March 28th for value, size, and momentum factors; from 2000/2001 January 3rd to 2019 March 28th for low volatility and high dividend yield factors respectively. We chose the more extended period to isolate what is a general background trend in the level of factor efficiency, factor premia and weights of factor information in stock prices and what is the impact of SB ETFs on that trend. We take daily and monthly returns of value, size and momentum factors for the US market from Kenneth R. French database. We proxy returns of low volatility and high dividend yield factors by MSCI indexes following the respective factors (MSCI USA Minimum Volatility and MSCI USA High Dividend Yield indexes respectively). MSCI Inc. is indexes' provider, one category of its indexes are factor indexes focusing on stocks with higher exposure to different factors (MSCI, 2020). However, the difference between Fama and French factors from Kenneth R. French database and MSCI indices should be acknowledged. Fama and French factors (value, size, and momentum) are theoretical, calculated based on long-short portfolios which exclude market portfolio (Fama & French, 1993; Carhart, 1997); thus, excess market returns are obtained. Contrary, the MSCI indices, which we use to proxy low volatility and high dividend yield factor returns, are actual investment vehicles that are formed by using long positions in the parent index but with different weighting. To evaluate the performance of indices, they should be benchmarked against market portfolio. Therefore, the results should be interpreted by taking into account the difference: the results obtained for low volatility and high dividend yield factor are based on the real investable indices, while for value, size, and momentum factors, results are based on the academic calculations and the actual effect might be more or less pronounced.

We download daily US-traded SB ETFs data from Factset dataset from 2012 January 3rd to 2019 March 28th. It is reasonable to have the year 2012 as a starting point, because SB ETFs are relatively new investment vehicles with significant growth in recent years; thus, we do not expect any significant effect of SB ETFs on the variables of our interest in years earlier than 2012. We assume that dollar trading volume was 0 for the years earlier than 2012, which is a reasonable assumption, because SB ETFs were just starting to grow in the US market. From all US-traded ETFs, we manually select SB ETFs following factors of our interest. We use dollar daily trading volume (in USD) as our SB ETFs growth proxy. Dollar trading volume of ETFs is an indicator of SB prevalence because not the passive holding of stocks drives price discovery, efficiency, etc. Instead, it is the active trading of stocks (in and out, factor timing, exploiting mispricing, etc.) that causes prices to adjust. Thus, if anything, we would expect that the asset pricing effects are linked to the trading volumes (not AUM). Using daily data, we calculate monthly averages of dollar trading volume for the month, which we use in our tests and regressions. For our analysis, we use levels (not changes) of variables; time variables are used as controls (described below) to control out the potential spurious effect. We did winsorization of variables using 1-4% and 99-98% as cut off points to remove outliers from our dataset. Therefore, our final dataset consists of 407 SB ETFs: 24 of which are betting on value, 135 on high dividend yield, 25 on low volatility, 25 on momentum, and 198 on size factor. We present the graphs visualizing the growth of SB ETFs dollar trading volume during our analyzed period for each factor in Appendix A. The graph for all factors jointly can be found below (Figure 1). We can see the significant increase in dollar trading volume throughout the years (from around 231 to 316 million USD). For individual factors, the curves are also upward sloping (with a drop in 2017 for almost all factors). High dividend yield SB ETFs have the highest dollar trading volume, followed by size, value, low volatility, and momentum SB ETFs.

Figure 1: Dollar trading volume of SB ETFs throughout 2012-2018 time period for all factors jointly



For semi-strong efficiency analysis, we download market macro variables from Thomson Reuters Datastream database for a period of 2004 January – 2019 April: quarterly US GDP growth, monthly 3-month Treasury Bill, S&P 500 dividend yield, US CPI inflation rate, VIX, and market returns (proxied by S&P 500 index returns).

For factor information into stock price analysis, we use daily data of all available individual stock returns of the three largest stock exchanges in the US: NYSE, AMEX, and NASDAQ. We take adjusted returns from the CRSP database for a period of 2000 January 1^{st} - 2019 March 28^{th}.

To avoid omitted variable bias, we use four control variables in our regressions:

$Trend_m$ – we introduce a linear time trend to capture non-measurable changes that occurr over time in the dependent variable.

$Dummy\_2016_m$ – a dummy variable that is equal to 1 if the year is 2016 and up to the recent period. This dummy divides our sample period into two parts accounting for non-linear time trends. An increase in the liquidity of the market, a decrease in the transaction costs and other improvements in the markets - these are the recent changes, which we expect to be captured by this dummy.

VIX_m – a volatility index. We introduce it to capture the differences in the dependent variable during the periods of higher volatility.

Mkt_volume_m – market trading volume of the S&P 500 index. We include this control variable because more and more investing activities are happening in the market, which itself may affect the dependent variable.

4. Factor efficiency

4.1. Measuring efficiency

For our analysis of **the efficiency of factor returns**, we use weak-form and semi-strong efficiency measures, which are well-known and commonly used in academic literature.

4.1.1. Weak form efficiency measure

To test weak form factor efficiency, we take factor returns and calculate different efficiency measures to understand what can explain factor returns. Autocorrelation, variance and delay measures are calculated for every month using month worth of daily data for each factor separately. We present the descriptive statistics in 4.1.3 section.

<u>First-order return autocorrelation.</u> Under the random walk, returns should be independent and identically distributed, and there should be no first-order log-return autocorrelation (random walk hypothesis). To calculate the autocorrelation measure, we take time-series daily data for a particular factor on a given month m. We calculate the measure every month for each factor separately:

$$Autocorrelation_m = |Corr(r_d\,;\, r_{d-1})| \qquad [1]$$

$Autocorrelation_m$ stands for autocorrelation measure of a certain factor calculated on monthly basis m using daily data; r_d stands for factor log-return on day d; r_{d-1} stands for lagged factor log-return on day $d-1$.

By calculating the absolute value of autocorrelation, we can capture both under- and overreaction of new information (positive and negative autocorrelation, respectively). Autocorrelation implies the ability to predict returns based on previous period returns, and the higher level of autocorrelation (both positive and negative) shows the higher inefficiency of a factor (Fama 1970, 1991; Worthington & Higgs, 2004).

Variance ratio. The variance (σ_d^2) of the d-period return should be d times bigger than the 1-period return's variance, that is – it should follow a linear function. This implies that factor returns are following a random walk and are weak-form efficient (Lo & MacKinlay, 1988):

$$\sigma^2_{d\text{-}period\ return} = d\ \sigma^2_{1\text{-}period\ return} \qquad [2]$$

Thus, we take daily data for a particular factor on a given month m. From the following relationship, we can construct the measure in monthly frequency for each factor separately:

$$Variance\_Ratio_m = \left|\frac{\sigma_d^2}{d\sigma_1^2} - 1\right| \qquad [3]$$

$Variance\_Ratio_m$ stands for variance measure of a certain factor calculated on monthly basis m using daily data; σ_d^2 stands for factor variance of d-period log-returns, d stands for the relationship coefficient of d- and 1-period returns, σ_1^2 stands for the variance of 1-period factor returns.

The higher absolute value of $Variance\_Ratio$ implies higher inefficiency of factor returns. Under the random walk hypothesis, this ratio should be equal to 0. We use 1 and 3-day log-returns to calculate the ratio.

Delay measure. Delay measure shows the magnitude of how lagged market returns can predict factor returns. We run the regression of daily log-factor returns on lagged daily log-market index returns (5 days lag) for each factor separately:

$$r_d = \alpha + \beta\, r_{mkt_d} + \sum_{k=1}^{5} \delta_k\, r_{mkt_{d-k}} + \varepsilon_d \qquad [4]$$

r_d stands for the return of a certain factor on day d; r_{mkt_d} stands for market return on day d; k stands for the number of lagged periods; δ_k stands for the coefficient before the k-period lagged market return r_{mkt} on day $d-k$; $r_{mkt_{d-k}}$ stands for k-period lagged market return r_{mkt}; ε_d stands for an error term of the regression.

We do not expect that lagged market returns of an earlier than five days will bring significant value. We choose S&P 500 index as a proxy for market returns since big stocks are likely to be the first ones to reflect market-wide information. Statistically significant β

coefficient imply the immediate response of the return to the market information. If any of δ_k coefficients are statistically significant, this imply the lagged response to the market information.

The second step is to save the unconstrained R^2, $R^2_{Unconstrained}$, from the regression [4] above and calculate constrained R^2, $R^2_{Constrained}$, by running regression [4] with a restriction that the coefficients before k-period lagged market return are equal to 0, $\delta_k = 0$. We take daily calculated $R^2_{Unconstrained}$ and $R^2_{Constrained}$ for a given month and calculate delay measure for each month and each factor separately using the following formula:

$$Delay_m = 1 - \frac{R^2_{Constrained}}{R^2_{Unconstrained}} \qquad [5]$$

The larger delay measure captures lower informational efficiency due to higher explanatory power of lagged market returns in explaining a factor returns' variation, which implies the slower process of incorporation of market-wide information into factor returns (Hou & Moskowitz, 2005).

<u>Combined efficiency measure</u>. Finally, after we calculate three different measures, to get one final measure of weak-form efficiency, we take the equally weighted average of all the measures. We do that to avoid inconsistencies among the measures.

$$WF\_Inefficiency_m = \frac{Autocorrelation_m + Variance\_Ratio_m + Delay_m}{3} \qquad [6]$$

$WF\_Iefficiency_m$ stands for equally weighted weak-form efficiency measure on month m; other variables as defined previously.

4.1.2. Semi-strong efficiency measure

Semi-strong factor efficiency tests show whether all available current and past information can predict future stock returns. We run the predictive regressions with factor returns as our dependent variable. We regress factor returns on one-period lagged market macro variables, our independent variables (listed below in this section).

We run predictive regressions for each factor separately taking monthly data to estimate one month ahead factor returns using lagged market macro variables. The general regression for factor returns predictability:

$$r_m = \alpha + \sum_{n=1}^{7} \beta_n F_{n,m-1} + \varepsilon_m \qquad [7]$$

r_m stands for the return of a certain factor on month m; n stands for the number of predictor variables; F_n stands for a predictor variable on month $m-1$; ε_m stands for an error term of the regression.

Adjusted R^2 and the significance of each predictive variable tell us whether market macro variables can predict future factor returns. Higher adjusted R^2 from the regression implies lower informational efficiency since more public information is incorporated into factor returns. Similarly, a higher and statistically significant beta coefficient of a market macro variable implies higher predictive power of that variable on factor returns, and, in turn, lower informational efficiency. We present these variables in the Descriptive statistics section 4.1.3.

Based on the previous literature and data availability, the predictor variables to test semi-strong factor efficiency are the following: lagged values of US GDP growth (GDP_{m-1}) (quarterly values are interpolated for each month), yield on the US 3-month Treasury Bill (T_Bill_{m-1}), S&P 500 dividend yield (Div_Yield_{m-1}), US CPI inflation rate ($Inflation_{m-1}$), VIX (VIX_{m-1}), factor return (r_{m-1}), and market returns (proxied by S&P500) ($r_{mkt,m-1}$). The chosen variables are widely used in the literature to predict stock or factor returns. The predictive power of the variables varies over time; thus, it is hard to determine one "best" model to predict the returns (Pesaran & Timmermann, 1995; Bossaerts & Hillion, 1999). The goal of our analysis is not to check the effect of each market macro variable on the future factor returns but to check the predictive power of the market macro variables jointly. Thus, we do not elaborate on the expected signs of the coefficients but focus on the possible predictive power of the variables mentioned above.

The second step of this analysis is to create a measure that could be linked to SB ETFs. Thus, after we run time-series predictive regressions for each factor separately using monthly data, we get time-series fitted (predicted) values of our predicted returns on a factor in the next month. Using predicted values, we calculate the prediction error by subtracting the predicted value from the actual factor return on that month (Stock & Watson, 2003; Dangl & Halling, 2012). By taking the absolute value of the prediction error, we get time-series of monthly measure for each factor. Lastly, we calculate the equally-weighted average prediction error of all factors ($|\varepsilon\_avg_m|$) and get the final measure which we can use in the second stage of the analysis – linking the predictive power of market macro variables with SB ETFs.

$$|\varepsilon_m| = r_m - \widehat{r_m} \qquad [8]$$

$|\varepsilon_m|$ stands for a prediction error of the month ahead predicted return of a certain factor; $\widehat{r_m}$ stands for the fitted value of a predicted factor return; other variables as defined previously.

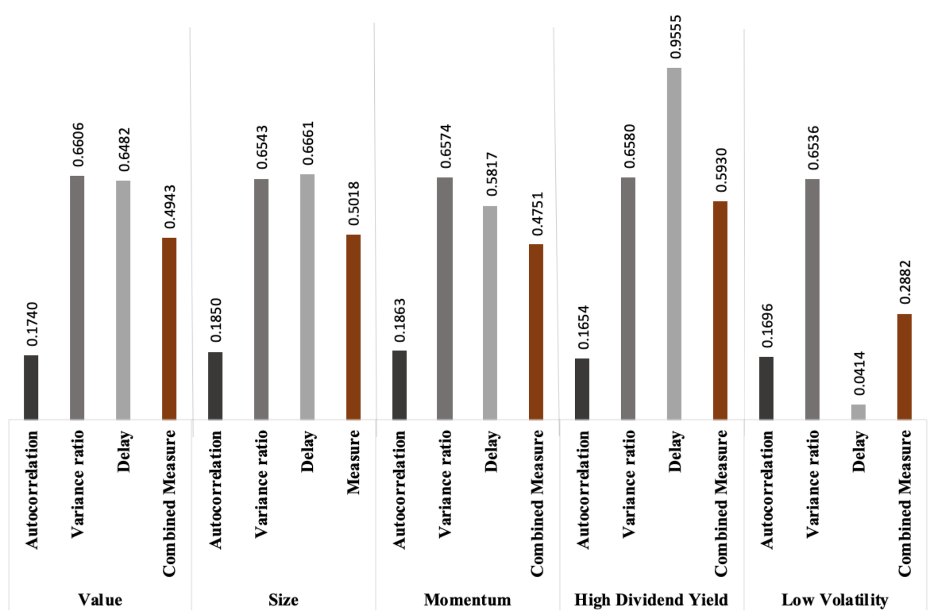
The higher absolute value of the prediction error implies that predictive regressions with lagged market macro variables are less accurate at predicting factor returns. Which, in turn, means higher semi-strong efficiency of factor returns. In section 4.1.3, we present descriptive statistics of prediction errors of all factors separately and the equally weighted measure.

4.1.3. Descriptive statistics of efficiency measures

In Figure 2 below, we can see the averages of all weak-form efficiency measures for each factor displayed in the bar chart. The combined efficiency measure ranges from 0.2882 to 0.5930 among all factors. It is the highest for high dividend yield factor and the lowest - for low volatility factor. All measures have a similar pattern: autocorrelation measure is the lowest among all measures for all factors while variance ratio and delay measures are somewhat similar across all factors. Delay and variance measures are higher than the autocorrelation measure. The only exception is the low volatility factor for which the delay measure is exceptionally low, 0.0414, which decreases the combined efficiency measure and makes it the lowest among all factors.

Figure 2: Weak-form effficiency measures

The bar chart displays the averages of all efficiency measures for each factor. Combined measure stands for combined weak-form efficiency measure.



Results of predictive regressions of semi-strong efficiency analysis are somewhat mixed (Table 1). That confirms the contradicting findings in the literature about factor returns predictability (for example, Ilmanen et al., 2019). Low adjusted R^2 measure for some of the factors (mainly size and momentum) gives the support for the semi-strong efficiency of factors returns. On the other hand, more than 70% of returns variance is explained by market macro variables for high dividend yield and value factors. The predictive power of macroeconomic variables for the momentum factor is low (adjusted R^2 around 1.5%), which coincides with Griffin et al. (2003) and Cooper et al. (2004) findings. However, any predictability of factor returns can be attributed to the time-varying risk premia rather than the inefficiency of factor returns.

Table 1: Semi-strong efficiency predictive regressions

Table shows beta coefficients and significance levels and adjusted R^2 measure for the following regression:

$$r_m = \alpha + \sum_{n=1}^{7} \beta_n F_{n,m-1} + \varepsilon_m \qquad [7]$$

r_m stands for the return of a certain factor on month m; n stands for number of predictor variables; F_n stands for a predictor variable on month $m-1$; ε_m stands for an error term of the regression.

Variables used in the regression are lagged by one period (i.e., month): GDP_{m-1} growth in the US; Div_Yield_{m-1} which stands for the dividend of yield the S&P 500 index; $Inflation_{m-1}$ which is the CPI US inflation rate; T_Bill_{m-1} which stands for the 3-month US treasury rate; VIX_{m-1} which is the volatility index; the market return $r_{mkt,m-1}$; and r_{m-1} lagged factor return.

The analysis is done for 2004-2019 period.

p<0.1; p<0.05**; p<0.01****

The coefficients before market macro variables for semi-strong efficiency tests

Factor	Adjusted R^2	GDP_{m-1}	Div_Yield_{m-1}	$Inflation_{m-1}$	T_Bill_{m-1}	VIX_{m-1}	r_{m-1}	$r_{mkt,m-1}$
Value	0.7415	0.5969 *	0.9360	-0.1314	0.0469	-0.0062	-0.1079 **	0.7449 ***
Size	0.0027	-0.0955	0.3657	-0.3713	-0.1916	-0.0076	-0.1709 **	0.0247
Momentum	0.0153	-27.5405	265.9228	38.2365	63.4340	2.1639	0.1510 *	13.3744
High Dividend Yield	0.7083	0.0578	1.1020	-0.1818	0.0816	0.0140	-0.1295 **	0.9819 ***
Low Volatility	0.1240	0.3293	1.1432	2.5951 ***	0.0449	-0.0172	0.2178 ***	-0.0517

The mean of prediction errors (reported in Table 2) ranges from 0.0121 to 2.5015 among different factors; momentum exhibiting the highest value and standard deviation. The average of the prediction errors of all factors jointly ranges from 0.0082 to 6.4550 throughout time. Time trend analysis shows that the prediction error has increased over time, which gives us a reason to analyze whether the increase in the prediction error, which implies the increase in semi-strong informational efficiency, is attributable to the growth of SB ETF.

Table 2: Descriptive statistics of prediction errors

Table shows the descriptive statistics of the prediction errors for all factors separately and the equally weighted average of the prediction errors for all factors jointly. The analysis covers the period of 2004-2019.

Factor	Mean	Standard Deviation	Min	Max
Value	0.0121	0.0084	0.0000	0.0937
Size	0.0189	0.0138	0.0001	0.0535
Momentum	2.5015	2.1222	0.0000	32.1058
High Dividend Yield	0.0121	0.0088	0.0000	0.0993
Low Volatility	0.0177	0.0143	0.0001	0.0766
Jointly	0.5125	0.4252	0.0082	6.4550

4.2. Efficiency measures and SB ETFs

4.2.1. Weak-form efficiency method

Calculating efficiency measures was the first step in answering our research question. Now we switch to the second stage to find out how the rise of SB ETFs affects factor efficiency. We use monthly panel data of SB ETFs and the combined efficiency measure for all factors jointly. To avoid the endogeneity problem, we run VAR regressions using monthly data of the combined efficiency measure and SB ETFs growth proxy (dollar trading volume) as variables of our interest. Afterwards, we supplement regressions with control variables which can affect the efficiency of factor returns. Control variables are described in the Data selection section.

$$WF\_Inefficiency_{i,m} = a_i + \sum_{k=0}^{l} \beta_k \, SB_{i,m-k} + \varepsilon_{i,m} \quad [9]$$

$$SB_{i,m} = a_i + \sum_{k=0}^{l} \beta_k \, WF\_Inefficiency_{i,m-k} + \varepsilon_{i,m} \quad [10]$$

$WF\_Inefficiency_{i,m}$ stands for equally weighted weak-form efficiency measure on month m for factor i; l stands for the number of lags; $SB_{i,m-k}$ stands for the monthly average of SB ETFs dollar trading volume on month $m-k$ for factor i; $\varepsilon_{i,m}$ stands for an error term of the regression.

We choose the Akaike Information Criterion (AIC) specify the number of lags for VAR regressions. In VAR regressions, all variables are endogenous, which makes it challenging to interpret the results. Thus, we use Wiener-Granger causality tests (hereafter Granger causality) to determine how each variable affects other variables. Using Granger causality tests, we are not testing the real relationship of cause and effect, but we test if a particular variable is followed by another: if X causes Y, then X can be used to predict Y (Geweke, 1984). If SB ETFs proxy Granger causes the combined weak-form efficiency measure, we can argue about the possible improvement in the efficiency of factor returns attributable to the growth of SB ETFs.

4.2.2. Weak-form efficiency results and discussion

Based on AIC, the number of lags for regressions was set to be 3. We present the results for Granger causality tests in Table 3, Panel A, left column. Granger causality test shows that SB ETFs proxy does not Granger cause the combined weak-form efficiency measure, but the measure does Granger cause SB ETFs proxy, which contradicts our hypothesis. We perform the same tests with the control variables (Table 3, Panel A, right columns). We find that SB ETFs proxy still does not Granger cause the combined weak-form efficiency measure. On the other hand, time trend, 2016-year dummy and VIX do Granger cause the combined weak-form efficiency measure. These findings indicate that efficiency changes can be attributable to the unmeasurable effects which appear over time.

Granger causality tests do not yield significant results; thus, to test the magnitude and direction of the effect, we repeat the analysis running 1 lag VAR regressions. Although VAR regressions' coefficients are not interpretable, 1 lag VAR's provide us at least some guidance about the direction and significance of the effect. We present the results in Table 3, Panel B. In both regressions, without and with control variables, coefficient before SB ETFs proxy is negative but statistically insignificant; thus, we cannot make any robust conclusions. However, the hypothesized negative sign gives motivation for further research around our first hypothesis. Additionally, statistical insignificance does not mean that there is no effect at all; there might be other reasons why the effect is not significant (e.g., the sample size is not sufficient enough, omitted variable bias, etc.) (Stock & Watson, 2003). Moreover, the lack of significance may be explained by contradicting findings in the literature. SB ETFs made factor investing easier than before, leading to an increase in the liquidity, which, in turn, positively affects informational efficiency (Glosten et al., 2016). On the other hand, liquidity might have decreased because stocks are locked up in ETFs, which decreases the number of stocks outstanding in the market (Israeli et al., 2017).

Table 3: Second stage of weak-form efficiency analysis

The table reports the Chi2 probability of the Granger causality for the following VAR regression [9]:

$$WF\_Inefficiency_{i,m} = a_i + \sum_{k=0}^{l} \beta_k \, SB_{i,m-k} + \varepsilon_{i,m} \qquad [9]$$

$WF\_Inefficiency_{i,m}$ stands for equally weighted weak-form efficiency measure on month m for factor i; l stands for the number of lags; $SB_{i,m-k}$ stands for the monthly average of SB ETFs dollar trading volume on month $m-k$ for factor i; $\varepsilon_{i,m}$ stands for an error term of the regression. We repeat the analysis with control variables described in the Data Selection section.

p<0.1; p<0.05**; p<0.01****

chi2 is reported in parentheses

Chi2 probability of the Granger causality of the SB ETFs proxy and controls on weak-form efficiency combined measure					
Without control variables	With control variables				
Panel A: Using 3 lags VAR					
$SB_{i,m}$	$SB_{i,m}$	$Dummy\_2016_m$	$Trend_m$	VIX_m	Mkt_volume_m
0.986	0.948	0.000 ***	0.025 **	0.000 ***	0.872
(0.15)	*(0.36)*	*(32.91)*	*(9.36)*	*(17.91)*	*(0.71)*
Panel B: Using 1 lag VAR					
-1.82E-10	-3.34E-10	4.10E-04 ***	-0.002 **	2.96E-04	0.000
(-0.29)	*(-0.53)*	*(3.48)*	*(-2.57)*	*(0.37)*	*(-0.54)*

4.2.3. Semi-strong efficiency method

For the semi strong-efficiency of factor returns analysis, we run time-series regressions of monthly data with the average prediction error as our dependent variable and dollar trading volume as an independent variable. We run the following regression [11] without and with control variables:

$$SS\_Inefficiency_m = a + \beta_1 \, SB_{i,m} + \varepsilon_{i,m} \qquad [11]$$

$SS\_Inefficiency_m$ stands for semi-strong inefficiency measure ($|\varepsilon\_avg_m|$ - average absolute prediction error across all factors) on month m; $SB_{i,m}$ stands for the monthly average of SB ETFs dollar trading volume on month m for factor i; other variables as defined previously.

The beta coefficient before the SB ETFs proxy tells us how much the prediction error increases/decreases with one unit increase in dollar trading volume. The positive beta coefficient implies that the prediction error increases with the rise of SB ETFs. It means that the predictive power of market macro variables decreases, which indicates the increase in semi-strong efficiency. Therefore, the expected sign to support our second hypothesis is a statistically significant positive beta coefficient before the SB ETFs proxy.

4.2.4. Semi-strong efficiency results and discussion

We present the results in Table 4. We find the support for our second hypothesis: the increase in SB ETFs trading decreases the inefficiency of factor returns (Table 4, left columns). One unit increase in dollar trading volume of SB ETFs leads to 2.49e-08 percentage points increase in the average absolute prediction error. The increase in the prediction error means that the predictive power of market macro variables decreases, which, in turn, implies higher semi-strong informational efficiency of factor returns.

Once we add controls to the regression, the significance of the coefficient before dollar trading volume slightly weakens but still gives strong support for our hypothesis (Table 4, right columns). All control variables are statistically significant. Linear trend increases the prediction error while VIX, the effect of the recent period (starting from 2016) and market trading volume decrease the prediction error.

According to Baltussen et al. (2019), returns of all factors jointly are not predicted by macro variables. Similarly, Ilmanen et al. (2019) did not find any factor returns predictability. On the other hand, due to specific factor properties, factors tend to outperform or underperform during certain macroeconomic conditions (Ang et al., 2006; Daniel & Moskowitz, 2016, etc.), which encourages investors to time factors to generate higher excess returns (Asness, 2016a; Asness, 2016b). Moreover, high adjusted R^2 from the predictive regressions for value, high dividend yield, and low volatility factors imply that these factor returns are not semi-strong efficient (Table 1). Therefore, investors can improve factor investing strategies by understanding factor returns predictability but should be cautious that the predictability decreased with the rise of SB ETFs.

Table 4: Semi-strong efficiency measure linked with SB ETFs proxy

Table shows beta coefficients, t statistics and significance levels for the following regression:

$$SS\_Inefficiency_m = a + \beta_1 SB_{i,m} + \varepsilon_{i,m} \qquad [11]$$

$SS\_Inefficiency_m$ stands for semi-strong inefficiency measure ($|\varepsilon\_avg_m|$ - average absolute prediction error across all factors) on month m; $SB_{i,m}$ stands for the monthly average of SB ETFs dollar trading volume on month m for factor i; $\varepsilon_{i,m}$ stands for an error term of the regression.
We repeat the analysis with control variables described in the Data Selection section.
The analysis is done for 2004-2019 time period.

p<0.1; p<0.05**; p<0.01****

t statistics is reported in parentheses

Beta coefficients for SB ETFs impact on the semi-strong efficiency of factor returns

Without control variables	With control variables				
$SB_{i,m}$	$SB_{i,m}$	$Dummy\_2016_m$	$Trend_m$	VIX_m	Mkt_volume_m
2.49E-08 ***	2.05E-08**	-0.2886	0.1464 ***	-0.1892 ***	-1.11E-08***
(2.60)	*(2.29)*	*(-25.74)*	*(47.54)*	*(-16.78)*	*(-34.51)*

4.3. One step approach

4.3.1. Method

Ehsani and Linnainmaa (2018) (hereafter E&L) test the predictability of factor returns using past returns and analyze the momentum in factor returns, which is an alternative way to check the factor returns efficiency and link it to the growth of SB ETFs. Firstly, we replicate E&L's findings for size, value and momentum factors using the same period, from 1963 until 2015. Afterwards, we use the same starting date (1963) but later end period (we use the latest available data up to 2019) to try capturing the longest time span. We proxy high dividend yield and low volatility factors using relatively new MSCI indexes; thus, we limit our analysis to the year 2001 and 2000 respectively. To check the efficiency of factor returns, we test the autocorrelation of factor returns by running a predictive regression [12]. The independent variable is an average factor return of the last 12 months, and the dependent variable is a factor return for a given month. We run a rolling windows time-series regression for each factor separately using monthly data.

$$r_m = a + \beta_1 \bar{r}_{(m-1,m-12)} + \varepsilon_m \qquad [12]$$

$\bar{r}_{(m-1,m-12)}$ is the average factor return of the last 12 months; other variables as defined previously.

To interpret the results, we use autocorrelation as an inefficiency measure (i.e., higher autocorrelation implies higher inefficiency). The statistically significant beta coefficient before the average past 12 months returns indicates that past returns can predict future returns; thus, it means that returns are autocorrelated.

E&L (2018) also check the momentum of factors returns. Since we already have four inefficiency measures, and a similar autocorrelation measure is used in the previous approach, we use factor momentum analysis only as a robustness test.

To find out how autocorrelation changes with the growth of SB ETFs, we supplement equation [12] with SB ETFs proxy and the interaction term of SB ETFs proxy and the average factor return of the last 12 months (assuming that the effect of average past year returns is the function of SB ETFs). We run time-series predictive regressions [13] for each factor separately using monthly data without and with control variables.

$$r_{i,m} = a + \beta_1 \bar{r}_{i,(m-1,m-12)} + \beta_2 SB_{i,m} + \beta_3 \bar{r}_{i,(m-1,m-12)} SB_{i,m} + \varepsilon_{i,m} \qquad [13]$$

$r_{i,m}$ stands for the return of factor i on month m; $\bar{r}_{i,(m-1,m-12)}$ is the average return of factor i of the last 12 months; other variables as defined previously.

The negative and statistically significant coefficient before the interaction term implies the decrease in the autocorrelation of factor return with the increase in SB ETFs dollar trading volume.

4.3.2 Results and discussion

We replicate E&L's analysis (sample period 1963-2015) only for size, value and momentum factors and can make the same conclusion: the returns of size and value factors are autocorrelated, while for the momentum factor, the negative coefficient before the average past 12 months returns implies no autocorrelation. That violates EMH and implies the inefficiency of factor returns. Furthermore, we do the same efficiency analysis but prolong our sample period to include recent years (until 2019) and low volatility and high dividend yield factors. We present the results in Appendix B. Results are similar to E&L: the beta coefficients of all factors except momentum are positive, indicating the existence of autocorrelation. However, the coefficients for size, momentum, and high dividend yield factors are not statistically significant. Our findings indicate that most factors are not efficient over a long period.

In the second stage of this analysis, we find that the increase in average dollar volume decreases the autocorrelation of factor returns, because the coefficient before the interaction term in the regression [13] is negative (Table 5, left column). Therefore, it gives support for our first hypothesis: the rise of SB ETFs increases the weak-form efficiency of factor returns. The coefficient before the interaction term and its significance changes once we added control variables to the regression (Table 5, right columns). The sign remains negative; however, it becomes statistically insignificant. Time trend, VIX, and market volume yield statistically significant negative coefficients, which means that these variables contribute to the increase in factor efficiency.

To summarize, the analysis using E&L's approach shows that factor returns are autocorrelated, which implies inefficiency (Fama, 1965). The change of the statistical significance of the beta coefficient before the interaction term and significance of control variables indicates that the decrease in autocorrelation is driven not purely by the rise of SB ETFs.

Table 5: Analysis of autocorrelation of factor returns using E&L (2018) approach linked with the growth of SB ETFs

Table shows beta coefficients, t statistics and significance levels for the following regression:

$$r_{i,m} = a + \beta_1 \bar{r}_{i,(m-1,m-12)} + \beta_2 SB_{i,m} + \beta_3 \bar{r}_{i,(m-1,m-12)} SB_{i,m} + \varepsilon_{i,m} \qquad [13]$$

$r_{i,m}$ stands for the return of factor i on month m; $\bar{r}_{i,(m-1,m-12)}$ is the average return of factor i of the last 12 months; $SB_{i,m}$ stands for the monthly average of SB ETFs dollar trading volume on month m for factor i; $\varepsilon_{i,m}$ stands for an error term of the regression. We repeat the analysis with control variables described in the Data Selection section.

p<0.1; p<0.05**; p<0.01****

t statistics is reported in parentheses

Beta coefficients for SB ETFs impact on the autocorrelation of factor returns					
Without control variables	With control variables				
$SB_{i,m}$	$SB_{i,m}$	$Trend_m$	$Dummy\_2016_m$	VIX_m	Mkt_volume_m
-3.77E-07*** *(-3.97)*	-1.16E-07 *(-1.20)*	-0.0084*** *(-3.67)*	0.1611 *(1.35)*	-0.6436*** *(-5.77)*	-8.17E-09** *(-2.43)*

To summarize the results of weak-form informational efficiency of factor returns from two approaches, we can emphasize that the results are sensitive to the choice of measures. Although some of the results do not support the first hypothesis, there are signs that SB ETFs increase the efficiency of factor returns. Furthermore, in the Literature Review section, we discussed other potential reasons which affect market efficiency: increase in active investing leads to the price convergence and market efficiency (Stulz, 2007); overshooting and price reversals (Da & Shive, 2018) or noise in the market provoked by the higher volatility of constitute stocks (Ben-David et al., 2018). These are the consequences of the high volume of ETFs trading, which lead to a decrease in informational efficiency. Thus, the above-mentioned factors may have affected our results, which gives an incentive for further research on this topic.

5. Weights of factor information in stock prices

5.1. Method

To analyze the extent to which factor returns drive stock returns, we take each individual stock and decompose it into factors- market, value, size, momentum, low volatility, high dividend yield, and the residual part - using months' worth of daily returns of factors and individual stocks. The logic is similar to Arbitrage Pricing Theory (APT) - we decompose the stock returns into factor loadings which might affect stock returns (Roll & Ross, 1980). However, we are not interested in checking the coefficients of our loadings, but we want to know what part of the variance of stock returns is explained by each factor. Thus, we regress daily individual stock returns on the factors mentioned above to calculate monthly partial R^2 for each factor and all factors jointly.

$$r_{s,d} = a + \beta_1 HML_d + \beta_2 SMB_d + \beta_3 MOM_d + \beta_4 DIV_d + \beta_5 LOW_d + \beta_6 r_{mkt,d} + \varepsilon_d \quad [14]$$

$r_{s,d}$ stands for a certain stock return on day d; HML_d stands for value factor return on day d; SMB_d stands for size factor return on day d; MOM_d stands for momentum factor return on day d; DIV_d stands for high dividend yield factor return on day d; LOW_d stands for low volatility factor return on day d; other variables as defined previously.

Partial R^2 tells us which proportion of the stock returns variance is explained by each factor. Thus, such a multifactor model allows us to check what proportion of individual stock returns can be explained by factor returns individually or jointly (Stock & Watson, 2003). Afterwards, we take the equal-weighted average of partial R^2's across stocks to get the value of partial R^2 in each month for each factor separately or jointly. In this way, we get a monthly time-series of partial R^2 associated with each of the factors.

Thus, using the monthly time-series of partial R^2 associated with each of the factors, we can check how the growth of SB ETFs impacts the richness of factor information in stock prices. We construct the panel dataset of monthly weights of factor information (R^2's) and the SB ETFs proxy. To account for the reverse causality problem, we run VAR regressions using monthly factor weights (i.e. partial R^2) and SB ETFs proxy of individual factors or all SB ETFs jointly, without or with the control variables.

$$PR_{i,m}^2 = a_i + \sum_{k=0}^{l} \beta_k \, SB_{i,m-k} + \varepsilon_{i,m} \qquad [15]$$

$$SB_{i,m} = a_i + \sum_{k=0}^{l} \beta_k \, PR_{i,m-k}^2 + \varepsilon_{i,m} \qquad [16]$$

$PR_{i,m}^2$ stands for partial R^2 of factor i on month m; other variables as defined previously.

We use AIC to specify the number of lags for VAR regressions. Afterwards, we run Granger causality tests to check whether dollar trading volume is a useful predictor of the variation of factor information in stock prices. We supplement the analysis with Impulse Response Function (IRF) graphs to visualize the response of a weight of factor information to one standard deviation shock in SB ETFs dollar trading volume. IRF is another way to check how a variable affects another variable. In essence, it shows how one variable responds to a one standard deviation shock imposed by another variable and the variation of the response throughout time (Stock & Watson, 2003). The positive response implies that the rise of SB ETFs increases the extent to which factors drive stock prices. Such analysis allows us to get information about the nature of correlations between stocks. If factor returns mainly explain stock returns variation, this implies the stock returns co-movements. Thus, by linking each factor's weights with SB ETFs, we can check whether ETFs trading has implications for the co-movement of stock returns.

5.2. Results and discussion

We present descriptive statistics of R^2 and partial of R^2 from regression [14] in Table 6. Up to 68.7% of stock returns variation is explained by market and five factors returns; the mean of R^2 is 45%, which implies that on average less than half of stock returns are explained by the factors used in our analysis. For each factor separately, partial R^2 is less than 10%, ranging from 7.32% for high dividend yield factor up to 9.45% for size factor. Thus, we can see that the variation of stock returns explained by any of the factors is similar across all factors.

Table 6: Descriptive statistics of weights of factor information in stock prices

Table shows the descriptive statistics of the partial R^2 for all factors separately and R^2 for all factors jointly, which were obtained from regression [14] for a period of 2000-2019. Regressions were run each month.

Factor	Mean	Standard Deviation	Min	Max
Jointly *R^2*	0.4502	0.0681	0.2454	0.6867
Value *Partial R^2*	0.0796	0.0133	0.0563	0.1323
Size *Partial R^2*	0.0945	0.0170	0.0586	0.1553
Momentum *Partial R^2*	0.0792	0.0129	0.0537	0.1280
High Dividend Yield *Partial R^2*	0.0732	0.0186	0.0037	0.2175
Low Volatility *Partial R^2*	0.0767	0.0107	0.0548	0.1179
Market *Partial R^2*	0.0863	0.0160	0.0558	0.1539

The more profound analysis of the weights of factor information in stock prices is to check how they have changed throughout time for all factors jointly and separately. We plot time-series graphs that can be found in Appendix C. We use moving averages to present a smoother pattern and eliminate short term deviations. The extent to which all factors jointly explain stock returns (i.e., R^2 from the regression [14]) has been increasing up to 2010, followed by sudden ups and downs afterwards. If we look at the extent to which stock returns are explained by factor returns separately (i.e., partial R^2 from the regression [14]), an upward trend is visible for the market, low volatility, size, momentum and value factors, meaning that these factors explain more significant proportion of stock returns in recent years. Partial R^2 for the high dividend yield factor has been fluctuating quite a lot; thus, we do not observe any clear pattern.

Therefore, most of the time plots indicate an upward trend of the factor information in stock prices, which raises a question of whether it can be attributable to the growth of SB ETFs. The left column of Table 7 below presents the results of VAR regression [15] without control variables. We find that dollar trading volume does not Granger cause the variation of stock returns explained by all factors returns jointly (R^2) but does Granger cause partial R^2 of size, low volatility factors and market return. Thus, we find the support for our third hypothesis – part of the upward trend in the richness of factor information in stock prices can be attributed to the growth of SB ETFs. Once we add control variables to the VAR's, the Granger causality results show that SB ETFs dollar trading volume is a useful predictor of the weight of factor information in stock prices only for low volatility factor (Table 7, right columns). Most of the control variables Granger cause the weights of factor information in stock prices.

Table 7: Granger causality tests for weights of factor information in stock prices analysis

Table shows the Chi2 probability of Granger causality tests for VAR regressions with weights of factor information in stock prices (R^2/partial R^2) and SB ETFs dollar trading volume. The analysis is done for 2000-2019 time period.

$$PR^2_{i,m} = a_i + \sum_{k=0}^{l} \beta_k SB_{i,m-k} + \varepsilon_{i,m} \qquad [15]$$

$PR^2_{i,m}$ stands for partial R^2 of a factor i on month m; $SB_{i,m-k}$ stands for the monthly average of SB ETFs dollar trading volume on month $m-k$ for factor i; $\varepsilon_{i,m}$ stands for an error term of the regression. We repeat the analysis with control variables described in the Data Selection section.

p<0.1; p<0.05**; p<0.01****
chi2 is reported in parentheses

Factor	Without control variables	With control variables				
	$SB_{i,m}$	$SB_{i,m}$	$Dummy\_2016_m$	$Trend_m$	VIX_m	Mkt_volume_m
Jointly	0.607	0.131	0.157	0.000 ***	0.000 ***	0.021 **
R^2	*(1.83)*	*(5.63)*	*(5.20)*	*(24.74)*	*(89.62)*	*(9.69)*
Value	0.103	0.789	0.000 ***	0.009 ***	0.000 ***	0.000 ***
Partial R^2	*(6.18)*	*(1.04)*	*(22.73)*	*(11.54)*	*(46.17)*	*(28.01)*
Size	0.013 **	0.101	0.002 ***	0.000 ***	0.0160 **	0.000 ***
Partial R^2	*(10.7)*	*(6.23)*	*(14.55)*	*(74.45)*	*(10.30)*	*(56.06)*
Momentum	0.482	0.631	0.000 ***	0.021 ***	0.000 ***	0.000 ***
Partial R^2	*(2.46)*	*(1.72)*	*(18.28)*	*(9.75)*	*(40.50)*	*(24.77)*
High Dividend Yield	0.33	0.481	0.542	0.000 ***	0.000 ***	0.073 *
Partial R^2	*(3.43)*	*(2.46)*	*(2.14)*	*(64.01)*	*(18.75)*	*(6.95)*
Low Volatility	0.005 ***	0.092 *	0.000 ***	0.000 ***	0.000 ***	0.000 ***
Partial R^2	*(12.91)*	*(6.45)*	*(61.21)*	*(89.07)*	*(30.56)*	*(47.94)*
Market	0.062 *	0.52	0.000 ***	0.000 ***	0.001 ***	0.0001 ***
Partial R^2	*(7.34)*	*(2.26)*	*(35.36)*	*(55.17)*	*(17.46)*	*(15.73)*

We present IRF graphs in Appendix D. IRF for all factors jointly (R^2) shows a positive response of R^2 to one standard deviation shock imposed by SB ETFs dollar trading volume, which flattens out to slightly above 0 in subsequent periods. The response of partial R^2 of size, value, momentum, low volatility factors and market return to shocks imposed by SB ETFs dollar trading volume stabilizes at the positive value in the long run.

Thus, the analysis gives support for our third hypothesis, that with the increase in SB ETFs trading, the extent to which factor returns explain stock returns increases. The increase of factor information into stock prices leads to the increase in co-movement of stock returns (similarly to Kozak et al., 2018; Barberis et al., 2005). Moreover, many authors argue that conventional ETFs trading leads to higher co-movement of stocks (Broman, 2016; Da & Shive, 2018); we find that the effect of SB ETFs on co-movement of stocks is evident as well. However, we cannot make conclusions about the magnitude of changes in stock returns co-movement, but we present evidence that SB ETFs trading has implications for co-movements of stock prices.

6. Factor premia attenuation

6.1. Method

We use our SB ETFs growth proxy to find whether factors return premia has changed with the growth of SB ETFs. We face the reverse causality problem: higher factor premia attract more investors to invest in successful SB ETFs strategies, which consequently, decreases factor premia. To account for that, we run VAR regressions using factor returns and dollar trading volume as variables of interest. We run the following VAR regressions for all factors jointly using monthly data without and with control variables:

$$r_{i,m} = a_i + \sum_{k=0}^{l} \beta_k \, SB_{i,m-k} + \varepsilon_{i,m} \quad [17]$$

$$SB_{i,m} = a_i + \sum_{k=0}^{l} \beta_k \, r_{i,m-k} + \varepsilon_{i,m} \quad [18]$$

All variables were defined previously.

As described in the Data Selection section, we use different factors measurements: value, size, and momentum factors are taken from Kenneth R. French database (they are calculated based on long-short portfolios and excess market returns are obtained), while low volatility and high dividend yields factors are proxied using respective MSCI indexes. Therefore, to obtain the market premia for the latter group of factors and have comparable results across all factors, market return should be subtracted from index return to get excess factor return. We do the analysis both using index returns and excess index returns.

We use AIC to specify the number of lags for VAR regressions. Afterwards, we run Granger causality tests to determine how each variable affects other variables. If dollar trading volume Granger causes factor returns, that implies that SB ETFs affect factor premia.

6.2. Results and discussion

We present the results for the Granger causality test without control variables in Table 8, left column. The results using index returns and excess index returns are just the same; thus, we present them without subtracting the market return since we are interested to check not how factor premia changed across different factors, but to check the overall effect for all factors jointly.

We find that dollar trading volume is not a useful predictor of factor returns (i.e., factor premia). When we add control variables, SB ETFs growth proxy and all controls yield insignificant results (Table 8, right columns).

The previous findings in the literature are diverse. Although many authors find premia attenuation throughout time (e.g., Arnott et al., 2019), others argue how certain factors are limiting the premia attenuation. Limits to arbitrage or behavioral biases lessen the efficiency of the market, leaving the premia persistent for extended periods (Hanson & Sunderam, 2013; McLean & Pontiff, 2016). Moreover, Asness (2015) argues that premia did not disappear because of the compensation for the higher risk and errors made by investors. Indeed, there is proof that higher ETFs trading leads to excessive price movement in the market (Da & Shive, 2018); consequently, investors require compensation for the risk involved in trading ETFs. Thus, since there are many reasons which affect the persistence of factor premia, it might explain why we do not get significant results.

Table 8: Factor premia attenutations linked with SB ETFs

The table reports the Chi2 probability of the Granger causality for the following VAR regression:

$$r_{i,m} = a_i + \sum_{k=0}^{l} \beta_k SB_{i,m-k} + \varepsilon_{i,m} \qquad [17]$$

$r_{i,m}$ stands for the return of factor i on month m; l stands for the number of lags; $SB_{i,m-k}$ stands for the monthly average of SB ETFs dollar trading volume on month $m-k$ for factor i; $\varepsilon_{i,m}$ stands for an error term of the regression. We repeat the analysis with control variables described in the Data Selection section.

p<0.1; p<0.05**; p<0.01****

chi2 is reported in parentheses

Chi2 probability of the Granger causality of the SB ETFs proxy and controls on factor returns					
Without control variables	With control variables				
$SB_{i,m}$	$SB_{i,m}$	$Trend_m$	$Dummy\_2016_m$	VIX_m	Mkt_volume_m
0.821 *(0.92)*	0.825 *(0.90)*	0.999 *(0.03)*	0.856 *(0.77)*	0.997 *(0.05)*	0.884 *(0.65)*

7. Robustness tests

We repeat all our analysis without winsorization of dollar trading volume, combined weak form efficiency measure, and controls. Since winsorization does not change the coefficients, we do not report the results. Therefore, our results are robust with and without winsorization.

7.1. Weak-form efficiency

Specification (1). We perform the analysis described in 4.2.1 section running VAR regressions [9], [10] for each factor separately (instead of using the panel data). For example, we link the combined weak-form efficiency measure of value factor with the dollar trading volume of value SB ETFs. We repeat the analysis with control variables.

We present results in Appendix E, Panel B. Dollar trading volume Granger causes the combined efficiency measure only for high dividend yield factor. This finding may explain why the Granger causality in the panel regressions is not statistically significant. The high dividend yield factor has the biggest dollar trading volume (followed by value and size factors), which indicates that the effect of SB ETFs on factors may depend on the size of the respective SB ETFs.

Specification (2). We repeat the analysis by taking the first difference of log variables on both sides of the equations to eliminate all possible non-stationarity.

The coefficients changed: contrary to the results using absolute values, dollar trading volume Granger causes the combined efficiency measure (Appendix E, Panel C). However, when we add the control variables to the regression, Granger causality disappears.

Specification (3). For the E&L (2018) approach, instead of regressing factor return on average factor return of the last 12 months, we use the alternative approach described in the paper and regress factor returns on a dummy variable. The dummy is equal to one if the average last year's return is positive and equal to zero if it is negative.

$$r_m = a + \beta_1 D_{\bar{r}\ (m-1,m-12)} + \varepsilon_m \qquad [19]$$

$D_{\bar{r}\ (m-1,m-12)}$ stands for a dummy variable which is equal to one if the average last year's return is positive and equal to zero if it is negative; other variables as defined previously

The coefficient before the dummy variable shows the distinction of average returns between the periods after a year of positive average returns and after a year of negative average returns. If the beta coefficient before the dummy variable is statistically different from zero, it implies momentum in factor returns, which indicates the inefficiency of factor returns.

Our findings of the analysis of the momentum in factor returns coincide with E&L's (2018). We find momentum in all factors except momentum (Appendix F), which implies inefficiency in factor returns.

Specification (4). To find out how autocorrelation changes with the growth of SB ETFs, we supplement equation [19] with SB ETFs dollar trading volume and the interaction term of SB ETFs proxy and the dummy variable. The interaction terms are interpreted the same as for regression [13].

$$r_{i,m} = a + \beta_1 D_{\bar{r}\ (m-1,m-12)} + \beta_2 SB_{i,m} + \beta_3 D_{\bar{r}\ (m-1,m-12)} SB_{i,m} + \varepsilon_{i,m} \qquad [20]$$

All other variables defined previously.

We get a negative but statistically insignificant beta coefficient before the interaction term in regression [20] (Appendix G), which means that dollar trading volume might lead to the decrease in factor momentum (i.e., the increase in efficiency). When we add control variables to the regression, the coefficient before the SB ETFs growth proxy remains negative and statistically insignificant. Thus, our results are robust across different E&L (2018) efficiency measures.

We get a negative but statistically insignificant beta coefficient before the interaction term in regression [20] (Appendix G), which means that dollar trading volume might lead to the decrease in factor momentum (i.e., the increase in efficiency). When we add control variables to the regression, the coefficient before the SB ETFs growth proxy remains negative and statistically insignificant. Thus, our results are robust across different E&L (2018) efficiency measures.

Specification (5). We repeat the analysis for both E&L (2018) approaches for each factor separately. For example, we regress value factor returns on the average value factor return of the last 12 months or the dummy variable of value factor; and in the second stage, we add interactions with dollar trading volume of value SB ETFs.

Both E&L's (2018) approaches show that SB ETFs growth proxy contributes to the increase in efficiency for size and high dividend yield factors (Appendix H). For other factors, results are either contradicting, insignificant, or the opposite sign than expected. However, similarly to the results of the previous tests, the significance of the beta coefficient before the interaction term decreases once we add control variables to the regression. Thus, the robustness test shows that results are robust for factors with the highest SB ETFs trading.

7.2. Semi-strong efficiency

Specification (6). For semi strong-efficiency analysis, we perform the robustness tests which are similar to weak-form efficiency analysis. Instead of running pooled regressions with all SB ETFs, we run the same regression [11] but regress the absolute prediction error of each factor separately on the respective SB ETFs. For example, we regress the absolute prediction error of the value factor on the dollar trading volume of value SB ETFs.

We present the results in Appendix I, Panel B. The increase in the respective SB ETFs trading leads to a statistically significant increase in the prediction error for high dividend yield and value factors. For other factors, coefficients are positive (except for momentum factor) but not statistically significant. Once we add control variables to the regression, the significance of the coefficient before SB ETFs trading volume weakens for value factor but remains the same for high dividend yield. To summarize, the results mostly do not hold for each factor separately. Similarly to other robustness tests, the results indicate that the effect of SB ETFs on factors may depend on the size of the respective SB ETFs.

<u>Specification (7).</u> We repeat the analysis by taking the first difference of log variables on both sides of the equations to eliminate all possible non-stationarity.

The signs of the coefficients do not change (except for VIX variable) if we take the first difference of log variables, however, the coefficient before dollar trading volume becomes insignificant (Appendix I, Panel C).

<u>Specification (8).</u> Lastly, we repeat the analysis by accounting for possible autocorrelation and heteroskedasticity in the error terms. We run all the regressions using Newey-West estimators.

The coefficients from the regressions using Newey-West estimators have changed only marginally indicating that we did not face an issue of the autocorrelation and heteroskedasticity in the error terms (Appendix I, Panel D)

7.3. Weights of factor information in stock prices

<u>Specification (9).</u> We use the same VAR regressions [15], [16] but instead of using panel data, we run them for each factor separately. For example, we take partial R^2 of value factor and run VAR with dollar trading volume of value SB ETFs.

The results indicate that none of the respective SB ETFs Granger cause the partial R^2 if we run VARs for each factor separately (Appendix J, Panel B).

<u>Specification (10).</u> We repeat the analysis by taking the first difference of log variables on both sides of the equations to eliminate all possible non-stationarity.

The significance of the Granger causality tests decreased. SB ETFs dollar trading volume Granger causes the partial R^2 only for low volatility factor in the regression with control variables (Appendix J, Panel C).

7.4. Factor premia attenuation

<u>Specification (11).</u> We run VAR regressions for each factor separately using factor returns and SB ETFs dollar trading volume as our variables of interest. For example, we take value factor returns and run a VAR with dollar trading volume of value SB ETFs.

Our results are statistically insignificant for all factors indicating that there are no factor premia attenuations due to the growth of SB ETFs (Appendix K, Panel B).

<u>Specification (12).</u> We repeat the analysis by taking the first difference of log variables on both sides of the equations to eliminate all possible non-stationarity.

The analysis using the first difference of log variables almost did not change the results – Granger causality tests yield insignificant results (Appendix K, Panel C).

8. Limitations of our study and further research areas

We limit our analysis to 5 well-known factors and the US market, which does not allow us to make generalizations whether the results hold for all discovered factors (including quality, profitability, investment, etc.) and all markets. That raises a question of what the results would be for other factors in the global market; thus, our work serves as a background for further research to see the effect of the respective SB ETFs on a wider variety of factors.

Although SB ETFs experienced an exponential growth in recent years and now constitute to around 20% of all ETFs, the size of each factor separately may not be sufficient enough to have a significant effect. This reason could explain why in our robustness tests, the effects were most significant for factors with higher dollar trading volume of respective SB ETFs (high dividend yield, size, and value). Therefore, with the expected further growth of factor investing, especially in SB ETFs (Mishra, 2020), our research could serve as a background research for further implications of SB ETFs for the efficiency and quantity of factor information into stock prices and factor premia.

For weak-form efficiency analysis, the results are sensitive to the choice of weak-form efficiency measures, which are calculated using different logic and assumptions. Thus, different measures may lead to different conclusions, which is apparent in our study. However, there is no universal agreement among researchers which measure is the most accurate.

For semi-strong efficiency analysis, the results should be interpreted cautiously due to the risk of data mining. Researchers have limited economic reasoning of why some of the factors are exposed to the macro market variables. Thus, predictive regressions and, consequently, the conclusion about the (in)efficiency are sensitive to the chosen market macro variables in predictive regressions.

9. Conclusions

In our thesis, we compile together three widely discussed topics in the finance world – market efficiency, returns co-movement and premia attenuation – to check how they are impacted by the growth of the US-traded SB ETFs. We focus on five asset pricing factors – value, size, momentum, high dividend yield, and low volatility. In the past years, the switch of focus from active to passive investing gave rise to the partly passive and partly active factor investing strategies and particularly SB ETFs, which are still under-researched in academic literature. Thus, since the market expects further growth of SB ETFs, the timing of our research is pivotal due to the need to identify what are the implications of these investment vehicles. In our research, we find support for the first three out of four hypotheses, which indicates that SB ETFs has implications for efficiency of factor returns and co-movement of stock returns.

The analysis of the weak-form efficiency of factor returns reveals the long-term inefficiency of most of the factor returns, which implies that investors can gain profits by timing factors. Moreover, we find support for our first hypothesis that the emergence and growth of SB ETFs improves the weak-form efficiency of factor returns. These findings are evidence that SB ETFs contribute to better incorporation of information in stock prices. Thus, with the rise of SB ETFs, factor trading became easier, which allows investors to exploit inefficiencies and gain from factor returns predictability. However, more trading towards exploiting inefficiencies potentially decreases these inefficiencies and consequently, the profitability of factor investing strategies.

Semi-strong efficiency analysis reveals that the predictability of market macro variables on factor returns decreases with the increase in SB ETFs dollar trading volume, implying an improvement in semi-strong informational efficiency of factor returns, which goes in line with our second hypothesis. Thus, these findings imply twofold implications for the profitability of factor trading strategies. Firstly, investors can predict factor returns and adjust their strategies accordingly. Secondly, investors can require higher compensation for risk if factors are exposed to certain macroeconomic conditions. However, with further forecasted growth of SB ETFs, these inefficiencies may considerably decrease.

The growth of SB ETFs has implications for the excess co-movement of stocks. In our research, we study not how stocks move together, but the nature of the correlation of stocks. As a result of the growth of SB ETFs, investors devote more capital to the same investing strategies. Consequently, the extent to which stock returns are explained by factor returns increases, indicating the increase in the co-movement of stocks due to their exposure to similar risk factors. Thus, we find the support for our third hypothesis that the emergence and growth in SB ETFs increase the extent to which stock returns are driven by factors. Our analysis notifies investors about the risks associated with the co-movements of stocks, which they should take into account in portfolio constructions.

Factor investing has never been so easy, meaning that in recent years, more capital is devoted to these strategies, which potentially causes crowding effect. However, we do not find support for our last hypothesis: factor premia have not attenuated with the rise of SB ETFs. Therefore, the factor premia persist due to multiple reasons; one of the main reasons - compensation for the risk of factor investing.

Our analysis gives an overview of the benefits and risks associated with investing in SB ETFs and of factor investing in general. We want to conclude by emphasizing the importance of our robustness tests: the effect of SB ETFs on the variables of our interest is most often significant for high dividend yield, value and size factors, which have the highest trading volume. Therefore, this indicates that the size of SB ETFs might not be sufficient enough at this stage, however, with the forecasted further growth of these investment vehicles, the results point out the need for further research around these topics, using our research as the ground.

10. References

Ando, S. (2019). *International financial connection and stock return comovement* (IMF Working Paper No. 19/181). Retrieved from International Monetary Fund website: https://www.imf.org/en/Publications/WP/Issues/2019/08/22/International-Financial-Connection-and-Stock-Return-Comovement-48552

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

Arnott, R., Beck, N., Kalesnik, V., & West, J. (2016). How can “smart beta” go horribly wrong? *Research Affiliates*. Retrieved from https://www.researchaffiliates.com/en_us/publications/articles/442_how_can_smart_beta_go_horribly_wrong.html

Arnott, R., Beck, N., & Kalesnik, V. (2016). Timing “smart beta” strategies? Of course! Buy low, sell high! *Research Affiliates.* Retrieved from https://www.researchaffiliates.com/en_us/publications/articles/541_timing_smart_beta_strategies_of_course_buy_low_sell_high.html

Arnott, R. D., Clements, M., Kalesnik, V., & Linnainmaa, J. (2018). Factor momentum. *SSRN*. Retrieved from http://dx.doi.org/10.2139/ssrn.3116974

Arnott, R., Harvey, C. R., Kalesnik, V., & Linnainmaa, J. (2019). Alice’s adventures in Factorland: Three blunders that plague factor investing. *Research Affiliates*. Retrieved from https://www.researchaffiliates.com/en_us/publications/articles/710-alices-adventures-in-factorland-three-blunders-that-plague-factor-investing.html

Asness, C. S., Friedman, J. A., Krail, R., & Liew, J. M. (2000). Style timing: Value versus growth. *Journal of Portfolio Management*, 26(3), 50-60.

Asness, C. S. (2015). How can a strategy still work if everyone knows about it? *AQR Capital Management*. Retrieved from https://www.aqr.com/Insights/Perspectives/How-Can-a-Strategy-Still-Work-If-Everyone-Knows-About-It

Asness, C. S. (2016a). My factor philippic. *AQR Capital Management*. Retrieved from https://www.aqr.com/Insights/Perspectives/My-Factor-Philippic

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

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

Baltussen, G., Swinkels, L., & Van Vliet, P. (2019). Global factors premiums. *SSRN*. Retrieved from http://dx.doi.org/10.2139/ssrn.3325720

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

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

Barberis, N., Shleifer A., & Wurgler J. (2005). Comovement. *Journal of Financial Economics*, 75, 283–317.

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

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

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

Black, F., Jensen, M., & Scholes, M. (1972). *The capital asset pricing model: Some empirical tests. Studies in the theory of capital markets.* New York: Praeger Publishing.

Blume, M. (1980). The relative efficiency of various portfolios: Some further evidence. *Journal of Finance*, 35(2), 269-281.

Bossaerts, P., & Hillion, P. (1999). Implementing statistical criteria to select return forecasting models: What do we learn? *Review of Financial Studies*, 12(2), 405-428.

Brenner, M. (1979). The sensitivity of the efficient market hypothesis to alternative specifications of the market model. *Journal of Finance*, 34(4), 915-929.

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

Bruce I. J., & Levy, J. (2015). Smart beta: Too good to be true? *Journal of Financial Perspectives*, 3(2), 1-5.

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

Chordia, T., & Shivakumar, L. (2002). Momentum, business cycle, and time-varying expected returns. *Journal of Finance*, 57(2), 985-1019.

Chordia, T., Roll, R., & Subrahmanyam, A. (2006). Liquidity and market efficiency. *Journal of Financial Economics*, 87(2), 249-268.

Chung, D., & Hrazdil, K. (2010). Liquidity and market efficiency: A large sample study. *Journal of Banking and Finance*, 34(10), 2346-2357.

Cooper, M. J., Gutierrez, R. C., & Hameed, A. (2004). Market states and momentum. *Journal of Finance*, 59(3), 1345-1365.

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

Dangl, T., & Halling, M. (2012). Predictive regressions with time-varying coefficients. *Journal of Financial Economics*, 106(1), 157-181.

Daniel, K., & Moskowitz, T. J. (2016). Momentum crashes. *Journal of Financial Economics,* 122, 221–247.

Daniel, K., & Titman, S. (1996). *Evidence on the characteristics of cross sectional variation in stock returns* (NBER Working Paper No. 5604). Retrieved from National Bureau of Economic Research: https://www.nber.org/papers/w5604

Ehsani, S., & Linnainmaa, J. (2018). Factor momentum and momentum factor. *SSRN*. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3014521

Fama, E. F. (1965). Random walks in stock market prices. *Financial Analysts Journal*, 51(1), 75-80.

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

Fama, E. F. (1991). Efficient capital markets: II. *Journal of Finance*, 46(5), 1575-1617.

Fama, E. F. (2014). Two pillars of asset pricing. *American Economic Review*, 104(6), 1467-1485.

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

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

Fama, E. F., & French, K. R. (1995). Size and book-to-market factors in earnings and returns. *Journal of Finance*, 501, 131-155.

Fama, E. F., & French, K. R. (2004). The capital asset pricing model: Theory and evidence. *Journal of Economic Perspectives*, 18(3), 25-46.

Geweke, J. (1984). Chapter 19 Inference and causality in economic time series models. *Handbook of Econometrics*, 1101–1144.

Glosten, L., Nallareddy, S., & Zou, Y. (2016). *ETF trading and informational efficiency of underlying securities* (Columbia University Research paper No. 16-71). Retrieved from SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2846157

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

Gombola, M., & Liu, F. L. (1993). Dividend yields and stock returns: Evidence of time variation between bull and bear markets. *Financial Review,* 28(3), 303-327.

Griffin, J. M., Ji, X., & Martin, J. S. (2003). Momentum investing and business cycle risk: Evidence from pole to pole. *Journal of Finance*, 58(6), 2515-2547.

Gupta, T., & Kelly, B. (2019). Factor momentum everywhere. *Journal of Portfolio Management*, 4, 1-24.

Hamm, S. (2014). The effect of ETFs on stock liquidity. *SSRN*. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1687914.

Hanson, S. G., & Sunderam, A. (2013). The growth and limits of arbitrage: Evidence from short interest. *Review of Financial Studies*. Retrieved from https://dash.harvard.edu/bitstream/handle/1/11148066/hanson%2Csunderam_growth-and-limits_RFS_final.pdf?sequence=1&isAllowed=y

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

Hou, K., & Moskowitz, T. J. (2005). Market frictions, price delay, and the cross-section of expected returns. *Review of Financial Studies*, 18(3), 981-1020.

Hsu, J., & Li, F. (2013). Low-volatility investing. *Journal of Index Investing*, 4(2), 67–72.

Ilmanen, A., Israel, R., Moskowitz, T., Thapar, A., & Wang, F. (2019). Do factor premia vary over time? A Century of Evidence. *SSRN*. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3400998

Israeli, D., Lee, C. M. C., & Sridharan, S. A. (2017). Is there a dark side to exchange traded funds? An information perspective. *Review of Accounting Studies*, 22(3), 1048–1083.

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

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

Lettau, M., & Madhavan, A. (2018). Exchange-traded funds 101 for economists. *Journal of Economic Perspective*s, 32(1), 135-154.

Litzenberger R. H., & Ramaswamy, K. (1979). The effects of dividends on common stock prices tax effects or information effects? *Journal of Finance,* 37(2), 429-443.

Lo, A. W., & MacKinlay, A. C. (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. *Review of Financial Studies*, 1(1), 41-66.

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

McLean, R. D., & Pontiff, J. (2016). Does academic research destroy stock return predictability? *Journal of Finance*, 71(1), 5-32.

Mishra, N. (2020). Should you invest in factor and smart beta ETFs? *Nasdaq*. Retrieved from https://www.nasdaq.com/articles/should-you-invest-in-factor-smart-beta-etfs-2020-01-09.

MSCI (2020). *Index solutions*. Retrieved from https://www.msci.com/index-solutions

Murphy, C. (2019). What's working in smart beta ETFs. *ETF.com*. Retrieved from https://www.etf.com/sections/features-and-news/whats-working-smart-beta-etfs?nopaging=1

Novy-Marx, R., & Velikov, M. (2015). A taxonomy of anomalies and their trading Costs. *Review of Financial Studies*, 29(1), 104-147.

Peng, L., & Xiong, W. (2006). Investor attention, overconfidence and category learning. *Journal of Financial Economics*, 80(3), 563-602.

Perez-Quiros, G., & Timmerman, A. (2000). Firm size and cyclical variations in stock returns. *Journal of Finance*, 3, 1229-1262.

Pesaran, M. H., & Timmermann A. (1995). Predictability of stock returns: Robustness and economic significance. *Journal of Finance*, 50(4), 1201-1228.

Rabener, N. (2019). Benchmarking smart beta ETFs. *Factor Research*. Retrieved from https://www.factorresearch.com/research-benchmarking-smart-beta-etfs

Rabener, N. (2020). Factor scoring smart beta ETFs. *Factor Research*. Retrieved from https://www.factorresearch.com/research-factor-scoring-smart-beta-etfs

Riding, S., (2018). Smart Beta moves into mainstream for large investors. *Financial Times*. Retrieved from https://www.ft.com/content/18497f00-c32b-11e8-84cd9e601db069b8

Roll R., & Ross, S. A. (1980). An empirical investigation of arbitrage pricing theory. *Journal of Finance*, 35(5), 1073-1103.

Rosenberg, B., Reid, K., & Lanstein R. (1985). Persuasive evidence of market inefficiency. *Journal of Portfolio Management*, 11, 9-17.

Schwert, G. W. (2003). Anomalies and market efficiency. *Handbook of the Economics of Finance*, 1(1), 939-974.

Stivers, C., & Sun, L. (2010). Cross-sectional return dispersion and time-variation in value and momentum premia. *Journal of Financial and Quantitative Analysis*, 45(4), 987-1014.

Stock, J. H., & Watson, M. W. (2003). *Introduction to Econometrics*. New York: Prentice Hall.

Stulz, R. M. (2007). Hedge funds: Past, present, and future. *Journal of Economic Perspectives*, 21(2), 175-194.

Vijh, A. (1994). S&P 500 trading strategies and stock betas. *Review of Financial Studies*, 7, 215–251.

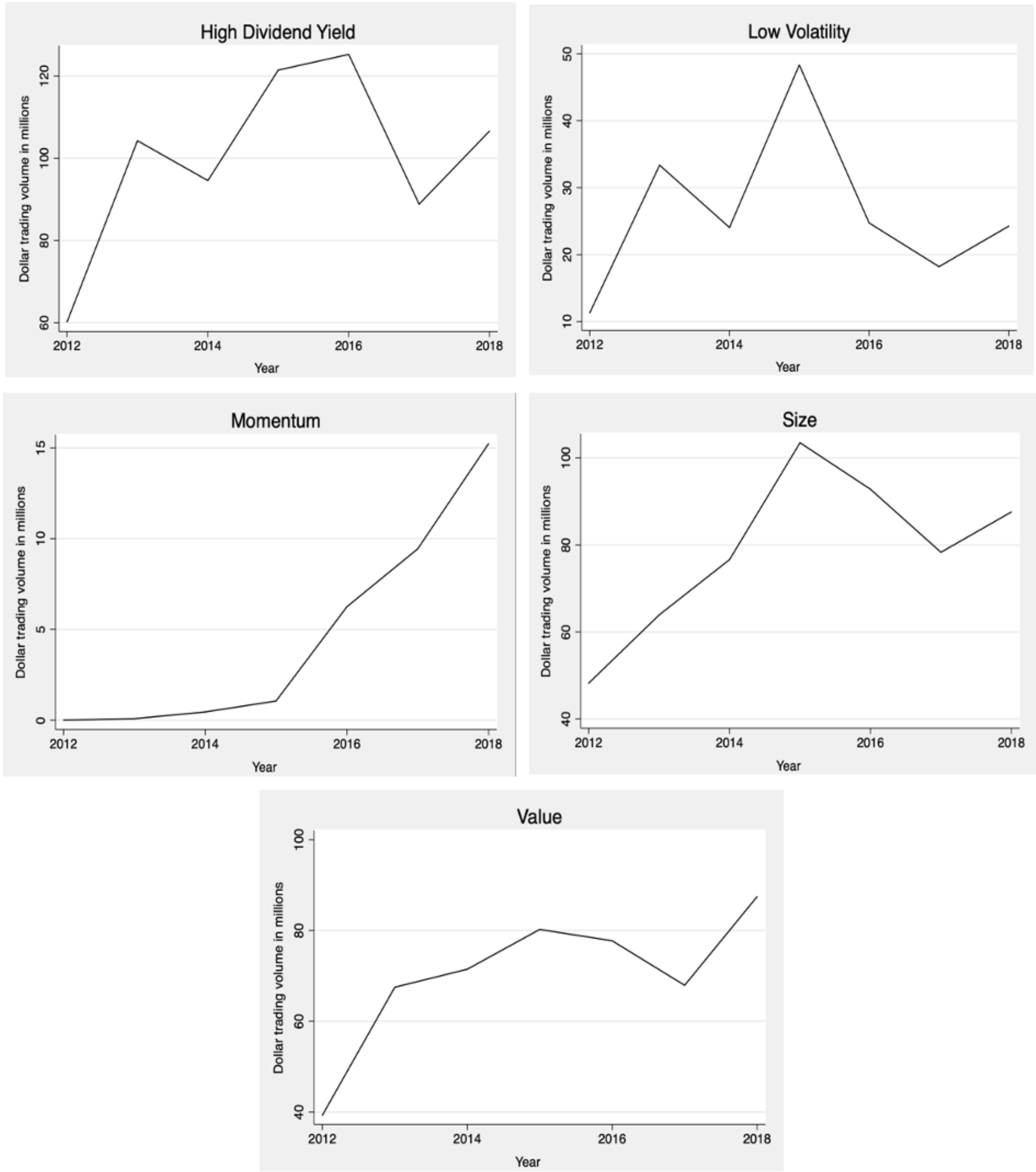
West, J., & Pickard, A. (2019). Plausible performance: Have smart beta return claims jumped the shark? *Research Affiliates*. Retrieved from https://www.researchaffiliates.com/en_us/publications/articles/767-plausible-performance.html

Worthington, A. C., & Higgs, H. (2004). Random walks and market efficiency in European equity markets. *Global Journal of Finance and Economics*, 1(1), 59-78.

Xu, L., & Yin, X. (2017). Exchange traded funds and stock market volatility. *International Review of Finance*, 17(4), 525–560.

11. Appendices

Appendix A. *Dollar trading volume of SB ETFs throughout 2012-2018 time period for each factor separately.*



Appendix B. *Analysis of autocorrelation of factor returns using Ehsani and Linnainmaa (2018) approach - long time span.*

Table shows beta coefficients, t statistics and significance levels for the following regression:

$$r_m = a + \beta_1 \bar{r}_{(m-1,m-12)} + \varepsilon_m \qquad [12]$$

r_m stands for a factor return on month *m*; $\bar{r}_{(m-1,m-12)}$ is the average factor return of the last year; ε_m is the error term of the regression.

The analysis for value, size and momentum factor is for 1963-2019 period; for high dividend yield: 2001-2019, for low volatility: 2000-2019. Periods were

p<0.1; p<0.05**; p<0.01****

t statistics is reported in parentheses

Beta coefficients for autocorrelation tests	
Factor	Before average factor return of the last year
Value	0.2558** *(2.29)*
Size	0.2082 *(1.64)*
Momentum	-0.0001 *(-0.00)*
High Dividend Yield	0.0019 *(0.01)*
Low Volatility	0.0516 *(0.23)*

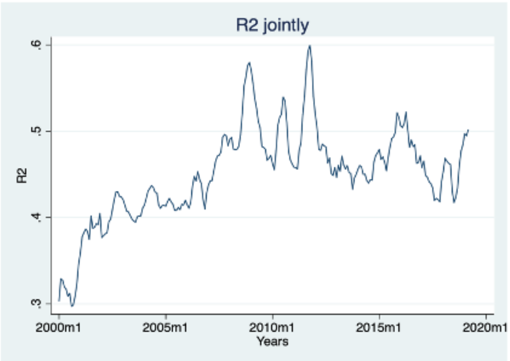
Appendix C. *The weights of factor information in stock prices - time trend graphs.*

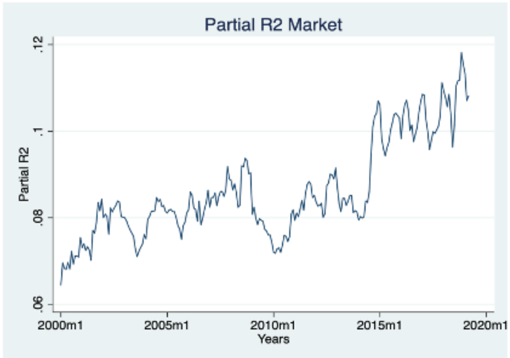
The weights of factor information in stock prices were calculated in the following way:

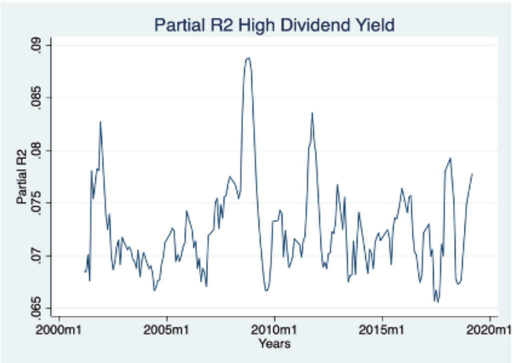
$$r_{s,d} = a + \beta_1 HML_d + \beta_2 SMB_d + \beta_3 MOM_d + \beta_4 DIV_d + \beta_5 LOW_d + \beta_5 r_{mkt,d} + \varepsilon_d \quad [14]$$

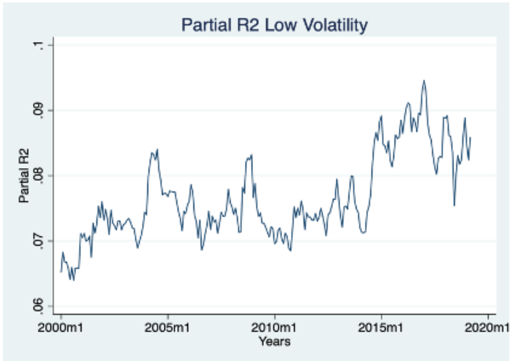
$r_{s,d}$ stands for a certain stock return on day d; HML_d stands for value factor return on day d; SMB_d stands for size factor return on day d; MOM_d stands for momentum factor return on day d; DIV_d stands for high dividend yield factor return on day d; LOW_d stands for low volatility factor return on day d; $r_{mkt,d}$ stands for market return on day d.

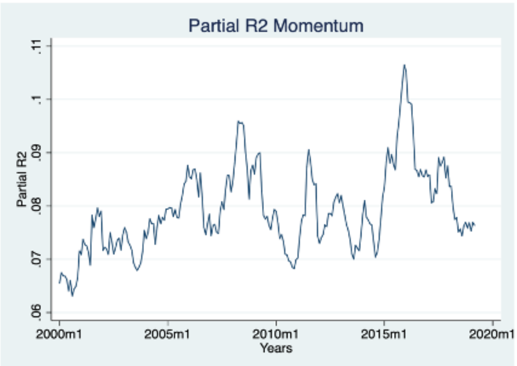
The graphs below show how the weights of factor information in stock prices have changed throughout time for all factors jointly and separately. Moving averages were used to present a smoother pattern and eliminate short term deviations. The graphs are for 2000-2019 period.

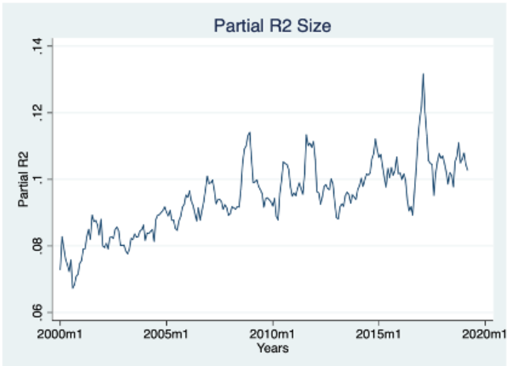


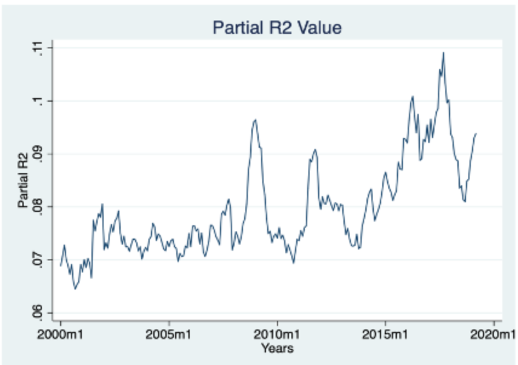






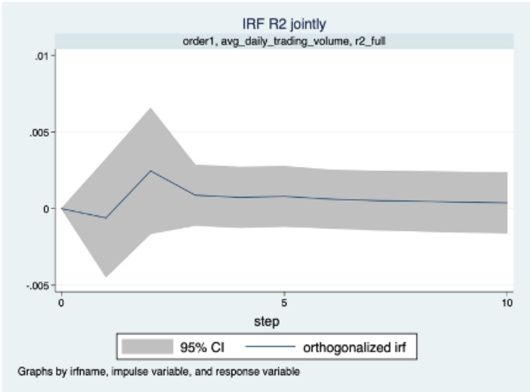


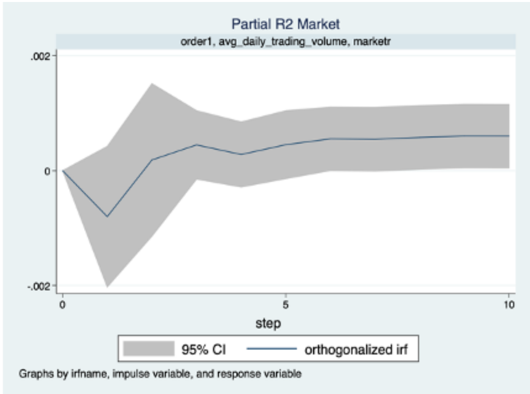


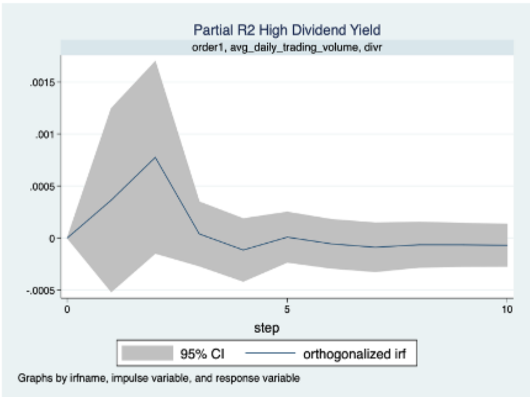


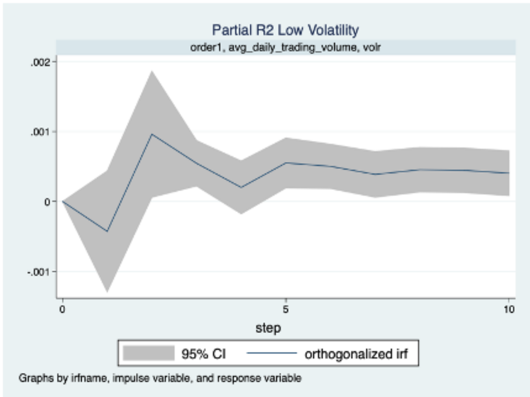
Appendix D. *IRF graphs of the response of the weights of factor information in stock prices to shocks imposed by SB ETFs dollar trading volume.*

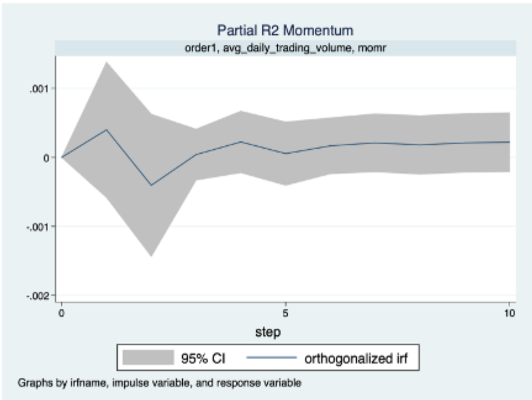
IRF graphs to visualize the response of a weight in factor information to one standard deviation shock in SB ETFs dollar trading volume. The positive response would imply the increase of weights of factor information in stock prices with the rise of SB ETFs. The graphs are for 2012-2019 period.

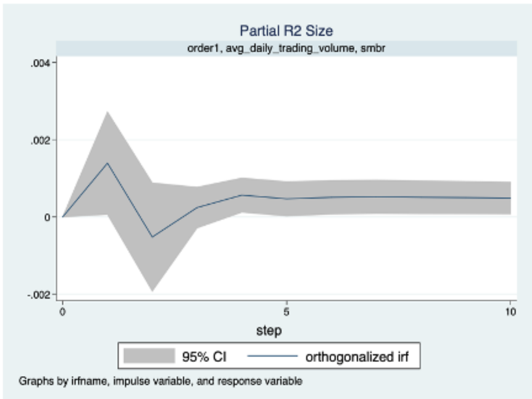


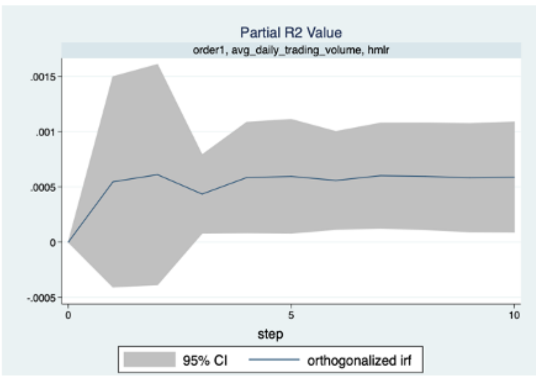












Appendix E. *Robustness analysis - second stage of weak-form efficiency.*

The table reports the Chi2 probability of the Granger causality of the SB ETFs proxy and controls on weak-form efficiency combined measure for VAR regression [9].

p<0.1; p<0.05**; p<0.01****

chi2 is reported in parentheses

Factor	Without control variables	With control variables				
	$SB_{i,m}$	$SB_{i,m}$	$Trend_m$	$Dummy\_2016_m$	VIX_m	Mkt_volume_m
Panel A: Base case						
Jointly	0.986 *(0.15)*	0.972 *(0.23)*	0.000 *** *(32.40)*	0.026 ** *(9.25)*	0.000 *** *(22.65)*	0.919 *(0.50)*
Panel B: Specification (1) - analysis for each factor separately						
Value	0.858 *(0.77)*	0.793 *(1.03)*	0.000 *** *(56.20)*	0.935 *(0.42)*	0.626 *(1.75)*	0.386 *(3.04)*
Size	0.980 *(0.19)*	0.999 *(0.03)*	0.000 *** *(24.02)*	0.001 *** *(17.12)*	0.000 *** *(32.16)*	0.077 * *(6.85)*
Momentum	0.960 *(0.30)*	0.969 *(0.25)*	0.240 *(4.21)*	0.937 *(0.42)*	0.584 *(1.95)*	0.896 *(0.60)*
High Dividend Yield	0.034 ** *(8.68)*	0.022 ** *(9.61)*	0.000 *** *(33.33)*	0.604 *(1.85)*	0.000 *** *(102.91)*	0.533 *(2.20)*
Low Volatility	0.962 *(0.29)*	0.790 *(1.05)*	0.412 *(2.87)*	0.785 *(1.07)*	0.000 *** *(23.55)*	0.243 *(4.18)*
Panel C: Specification (2) - analysis taking the first difference of log variables on both sides of the equations						
Jointly	0.041 ** *(4.19)*	0.137 *(2.21)*	0.832 *(0.04)*	0.673 *(0.17)*	0.838 *(0.04)*	0.222 *(1.49)*

Appendix F. *Robustness analysis - autocorrelation of factor returns using Ehsani and Linnainmaa (2018) approach - long time span.*

Table shows beta coefficients, t statistics and significance levels for the regression [19] of autocorrelation tests for all factors jointly and for each factor separately.
The analysis for value, size and momentum factor is for 1963-2019 period; for high dividend yield: 2001-2019, for low volatility: 2000-2019. Periods were chosen based on
p<0.1; p<0.05**; p<0.01****
t statistics is reported in parentheses

Specification (3) - autocorrelation of factor returns alternative analysis	
Factor	$D_{r\ (m-1,m-12)}$
Value	0.3985* *(1.78)*
Size	0.3574 *(1.48)*
Momentum	-0.0881 *(-0.22)*
High Dividend Yield	0.0074 *(1.03)*
Low Volatility	0.0113* *(1.92)*

Appendix G. *Robustness analysis - autocorrelation of factor returns analysis using Ehsani and Linnainmaa (2018) alternative approach linked with the growth of SB ETFs.*

Table shows beta coefficients, t statistics and significance levels for the regression [20] of SB ETFs impact on the autocorrelation of factor returns.
p<0.1; p<0.05**; p<0.01****
t statistics is reported in parentheses

Specification (4) - SB ETFs impact on the autocorrelation of factor returns					
Without control variables	With control variables				
$SB_{i,m}$	$SB_{i,m}$	$Trend_m$	$Dummy\_2016_m$	VIX_m	Mkt_volume_m
-9.99E-08 *(-1.10)*	-1.62E-08 *(-0.18)*	-0.0175 *** *(-4.30)*	0.1203* *(1.66)*	-0.1842 *(-1.51)*	-1.68E-08*** *(-5.81)*

Appendix H. *Robustness analysis - autocorrelation of factor returns analysis using two Ehsani and Linnainmaa (2018) approaches linked with the growth of SB ETFs.*

Table shows beta coefficients, t statistics and significance levels for regressions [13] and [20] (alternative approach) of SB ETFs impact on the autocorrelation of factor returns.

p<0.1; p<0.05**; p<0.01****

t statistics is reported in parentheses

Factor	Without control variables	With control variables				
	$SB_{i,m}$	$SB_{i,m}$	$Trend_m$	$Dummy\_2016_m$	VIX_m	Mkt_volume_m
Panel A: Base case						
Jointly	-3.77E-07 *** *(-3.97)*	-1.16E-07 *(-1.20)*	-0.0084*** *(-3.67)*	0.1611 *(1.35)*	-0.6436*** *(-5.77)*	-8.17E-09 ** *(-2.43)*
Panel B: Specification (5) - using regression [13]						
Value	0.1914 *** *(4.17)*	6.29E-08 *(0.45)*	0.0414*** *(8.25)*	2.7756 *** *(10.62)*	-1.3393*** *(-6.56)*	-1.02E-07 *** *(-13.13)*
Size	-0.7664 *** *(-12.70)*	-8.37E-08 *(-0.43)*	-0.0282 *** *(-6.12)*	1.5519 *** *(5.83)*	1.4845*** *(4.07)*	-6.85E-08 *** *(-7.87)*
Momentum	-0.0030 *(-0.03)*	-8.40E-08 *(-0.07)*	0.0060 *(0.71)*	0.5860 *(1.24)*	-1.2671** *(-2.12)*	-2.49E-08 ** *(-2.15)*
High Dividend Yield	-1.2006*** *(-18.52)*	1.55E-07 *(0.59)*	-0.3987 *** *(-12.63)*	0.6040 *(1.85)*	3.8721*** *(16.79)*	1.18E-08* *(1.77)*
Low Volatility	-1.0211*** *(-7.70)*	-4.35E-07 *(-0.99)*	-0.2614 *** *(-5.99)*	0.5683 *(1.04)*	2.7625 *** *(4.07)*	-3.88E-08 ** *(-2.44)*
Panel C: Specification (5) - using regression [20] (alternative approach)						
Value	0.6212 *** *(9.51)*	9.09E-08 *(0.47)*	0.0478 *** *(6.35)*	0.5421 *** *(4.08)*	-2.4453 *** *(-7.63)*	-1.10E-07 *** *(-15.83)*
Size	-1.2604 *** *(-21.37)*	7.22E-08 *(0.45)*	-0.0522 *** *(-6.78)*	-0.0059 *(-0.05)*	2.4693 *** *(9.92)*	-7.48E-08 *** *(-10.76)*
Momentum	-0.1784 *(-0.84)*	4.10E-07 *(0.20)*	0.0080 *(0.35)*	-3.4896 *** *(-4.72)*	-2.6956 * *(-1.66)*	-7.83E-08 *** *(-3.65)*
High Dividend Yield	-0.0432 *** *(-24.75)*	7.75E-12 *(0.00)*	-0.0074 *** *(-4.97)*	-0.1473 *** *(-5.17)*	-0.1162 *** *(-10.41)*	-4.49E-09 *** *(-3.86)*
Low Volatility	0.0130 *** *(2.92)*		0.0007 *(0.54)*		0.0098 *(0.42)*	

Appendix I. *Robustness analysis - semi-strong efficiency measure linked with SB ETFs growth proxy.*

Table shows beta coefficients, t statistics and significance levels for the regressions [11] of SB ETFs impact on the semi-strong efficiency of factor returns.

p<0.1; p<0.05**; p<0.01****

t statistics is reported in parentheses

Factor	Without control variables	With control variables				
	$SB_{i,m}$	$SB_{i,m}$	$Dummy\_2016_m$	$Trend_m$	VIX_m	Mkt_volume_m
			Panel A: Base Case			
Jointly	2.53E-08 ***	2.05E-08 **	-0.2886 ***	0.1464 ***	-0.1892 ***	-1.11E-08 ***
	(2.63)	*(2.29)*	*(-25.74)*	*(47.54)*	*(-16.78)*	*(-34.51)*
			Panel B: Specification (6) - analysis for each factor separately			
Value	6.31E-10 *	5.56E-10	0.0062 ***	-0.0006 ***	-0.0029 ***	-1.05E-10 ***
	(1.76)	*(1.59)*	*(12.30)*	*(-4.19)*	*(-6.10)*	*(-7.39)*
Size	3.30E-10	6.66E-11	-0.0075 ***	0.0034 ***	0.0025 ***	-1.42E-11
	(0.69)	*(0.15)*	*(-10.81)*	*(17.63)*	*(3.67)*	*(-0.71)*
Momentum	-4.29E-07	-3.55E-07	-1.7736 ***	0.9372 ***	-0.3275	-7.08E-08 ***
	(-0.85)	*(-0.75)*	*(-6.26)*	*(10.05)*	*(-1.16)*	*(-9.58)*
High Dividend Yield	1.15E-09 ***	1.20E-09 ***	-0.0072 ***	0.0016 ***	-0.0046 ***	1.13E-11 ***
	(2.61)	*(2.79)*	*(-17.49)*	*(12.74)*	*(-11.36)*	*(0.98)*
Low Volatility	4.17E-10	4.27E-10	0.01197 ***	0.00115 **	0.00983 ***	-3.76E-10 ***
	(0.40)	*(0.43)*	*(8.01)*	*(2.51)*	*(6.65)*	*(-8.81)*
			Panel C: Specification (7) - first difference of log variables on both sides of the equations			
Jointly	0.0007	0.0004	-0.0239 ***	0.0038 ***	0.1147 ***	-0.2741 ***
	(1.25)	*(0.96)*	*(-5.32)*	*(3.43)*	*(19.97)*	*(-17.53)*
			Panel D: Specification (8) - using Newey-West estimators			
Jointly	2.49E-08 **	2.04E-08 **	-0.2877 ***	0.1460 ***	-0.1894 ***	-1.11E-08 ***
	(2.28)	*(2.00)*	*(-14.80)*	*(22.45)*	*(-10.76)*	*(-24.25)*

Appendix J. *Robustness analysis - weights of factor information in stock prices analysis.*

Table reports the Chi2 probability of the Granger causality tests of dollar trading volume on weights of factor information in stock prices (R^2/partial R^2) from the regression [16].

p<0.1; p<0.05**; p<0.01****

chi2 is reported in parentheses

Factor	Without control variables	With control variables				
	$SB_{i,m}$	$SB_{i,m}$	$Dummy\_2016_m$	$Trend_m$	VIX_m	Mkt_volume_m
Panel A: Base Case						
Jointly	0.607 *(1.83)*	0.131 *(5.63)*	0.157 *(5.20)*	0.000 *(24.74)*	0.000 *(89.62)*	0.021 *(9.69)*
Value	0.103 *(6.18)*	0.789 *(1.04)*	0.000 *** *(22.73)*	0.009 *** *(11.54)*	0.000 *** *(46.17)*	0.000 *** *(28.01)*
Size	0.013 ** *(10.7)*	0.101 *(6.23)*	0.002 *** *(14.55)*	0.000 *** *(74.45)*	0.0160 ** *(10.30)*	0.000 *** *(56.06)*
Momentum	0.482 *(2.46)*	0.631 *(1.72)*	0.000 *** *(18.28)*	0.021 *** *(9.75)*	0.000 *** *(40.50)*	0.000 *** *(24.77)*
High Dividend Yield	0.33 *(3.43)*	0.481 *(2.46)*	0.542 *(2.14)*	0.000 *** *(64.01)*	0.000 *** *(18.75)*	0.073 * *(6.95)*
Low Volatility	0.005 *** *(12.91)*	0.092 * *(6.45)*	0.000 *** *(61.21)*	0.000 *** *(89.07)*	0.000 *** *(30.56)*	0.000 *** *(47.94)*
Market	0.062 * *(7.34)*	0.52 *(2.26)*	0.000 *** *(35.36)*	0.000 *** *(55.17)*	0.001 *** *(17.46)*	0.0001*** *(15.73)*
Panel B: Specification (9) - analysis for each factor separately linking it with the respective SB ETFs						
Value	0.593 *(1.90)*	0.107 *(4.47)*	0.289 *(2.48)*	0.637 *(0.90)*	0.028 ** *(7.14)*	0.232 *(2.92)*
Size	0.313 *(2.32)*	0.892 *(0.23)*	0.722 *(0.65)*	0.036 ** *(6.67)*	0.121 *(4.23)*	0.005 *** *(10.62)*
Momentum	0.857 *(0.77)*	0.609 *(1.83)*	0.147 *(5.36)*	0.679 *(1.51)*	0.026 ** *(9.26)*	0.122 *(5.80)*
High Dividend Yield	0.189 *(4.78)*	0.657 *(0.84)*	0.656 *(0.84)*	0.000 *** *(19.79)*	0.113 *(4.36)*	0.905 *** *(0.20)*
Low Volatility	0.724 *(1.32)*	0.676 *(0.17)*	0.337 *(0.92)*	0.009 *** *(6.77)*	0.414 *(0.67)*	0.024 ** *(5.12)*
Panel C: Specification (10) - first difference of log variables on both sides of the equations						
Jointly	0.630 *(1.73)*	0.705 *(1.40)*	0.108 *(6.07)*	0.002 *** *(15.41)*	0.000 *** *(49.60)*	0.002 *** *(14.3)*
Value	0.175 *(4.96)*	0.112 *(6.00)*	0.059 ** *(7.44)*	0.113 *(5.97)*	0.181 *(4.88)*	0.040 ** *(8.31)*
Size	0.691 *(1.46)*	0.510 *(2.31)*	0.017 ** *(10.17)*	0.000 *** *(50.94)*	0.040 ** *(8.31)*	0.000 *** *(25.95)*
Momentum	0.389 *(3.02)*	0.383 *(3.05)*	0.000 *** *(37.57)*	0.042 ** *(8.23)*	0.023 ** *(9.57)*	0.000 *** *(33.58)*
High Dividend Yield	0.699 *(1.43)*	0.375 *(3.11)*	0.792 *(1.04)*	0.031 ** *(8.88)*	0.149 *(5.34)*	0.319 *(3.51)*
Low Volatility	0.541 *(2.15)*	0.086 * *(6.58)*	0.128 *(5.69)*	0.000 *** *(26.16)*	0.000 *** *(22.99)*	0.009 *** *(11.57)*
Market	0.748 *(1.22)*	0.892 *(0.62)*	0.328 *(3.44)*	0.045 ** *(8.05)*	0.000 *** *(18.94)*	0.000 *** *(24.06)*

Appendix K. *Robustness analysis - factor premia attenuations linked with SB ETFs.*

The table reports the Chi2 probability of the Granger causality for the VAR regression [17] of factor premia attenuation analysis.
p<0.1; p<0.05**; p<0.01****
chi2 is reported in parentheses

Factor	Without control variables	With control variables				
	$Volume_m$	$Volume_m$	$Trend_m$	$Dummy\_2016_m$	VIX_m	Mkt_volume_m
Panel A: Base case						
Jointly	0.821 *(0.92)*	0.825 *(0.90)*	0.999 *(0.03)*	0.856 *(0.77)*	0.997 *(0.05)*	0.884 *(0.65)*
Panel B: Specification (11) - analysis for each factor separately						
Value	0.463 *(1.54)*	0.972 *(0.00)*	0.268 *(1.23)*	0.113 *(2.52)*	0.791 *(0.07)*	0.960 *(0.00)*
Size	0.998 *(0.04)*	0.945 *(0.00)*	0.717 *(0.13)*	0.671 *(0.18)*	0.819 *(0.05)*	0.947 *(0.00)*
Momentum	0.978 *(0.20)*	0.692 *(0.16)*	0.950 *(0.00)*	0.372 *(0.80)*	0.115 *(2.48)*	0.804 *(0.06)*
High Dividend Yield	0.964 *(0.00)*	0.870 *(0.71)*	0.921 *(0.16)*	0.913 *(0.53)*	1.000 *(0.01)*	1.000 *(0.01)*
Low Volatility	0.965 *(0.00)*	0.969 *(0.00)*	0.744 *(0.11)*	0.657 *(0.20)*	0.582 *(0.30)*	0.985 *(0.00)*
Panel C: Specification (12) - analysis taking the first difference of log variables on both sides of the equations						
Jointly	0.797 *(1.02)*	0.503 *(0.45)*	0.005*** *(7.88)*	0.017** *(5.71)*	0.956 *(0.00)*	0.993 *(0.00)*