TALLINN UNIVERSITY OF TECHNOLOGY
School of Business and Governance
Department of Economics and Finance

Veronika Matrossova

VOLATILITY EFFECT OF ETFs ON THE UNDERLYING ASSETS ON THE EXAMPLE OF S&P 500

Master's thesis

Programme Accounting and Finance, specialisation Finance

Supervisor: Kalle Ahi, MA

Tallinn 2020

I hereby declare that I have compiled the thesis independently
and all works, important standpoints and data by other authors
have been properly referenced and the same paper
has not been previously presented for grading.
The document length is 14464 words from the introduction to the end of conclusion.

Veronika Matrossova ……………………………
(signature, date)
Student code: 183010TARM
Student e-mail address: veronica.matrosova@gmail.com

Supervisor: Kalle Ahi, MA:
The paper conforms to requirements in force

……………………………………………
(signature, date)

Chairman of the Defence Committee:
Permitted to the defence
…………………………………
(name, signature, date)

TABLE OF CONTENTS

ABSTRACT

This thesis primarily focuses on excess volatility potentially caused by trading with ETFs. The aim is to examine the relationship between trading volumes of ETFs and volatility of the underlying index, S&P 500. An examined hypothesis states that trading volumes of ETFs have no influence on the volatility of the S&P 500. Thus, they do not cause any excess volatility. However, a thorough analysis with an application of empirical methods indicates that the hypothesis can be rejected. Thesis tests the hypothesis by applying OLS model for linear relationship testing and VAR for simultaneous relationships testing between multiple variables. Also, it uses GARCH for conditional volatility modelling and Granger-causality test for studying cause and effect between variables. Tests results indicate that trading volumes may affect conditional volatility when tests are performed on daily data. As for unconditional volatility, there is partial evidence on that ETFs trading volumes affect the underlying index. However, Granger causality tests find no strong causality between tested variables, but in very few cases, evidence indicates that causality exists. Thus, the results are mixed. The conclusion is that in some cases trading volumes of ETFs may affect the S&P 500 volatility.

Keywords: ETF, volatility, VAR, Granger causality, passive investing

INTRODUCTION

Over the years, financial markets have introduced many new products and instruments with their specific features. One such recent product is exchange-traded funds or ETFs. They were created in the previous century by John Bogle but got popular over the past decade. In the US ETF industry has reached 5 trillion dollars in AUM in 2020, which is a tremendous amount, but in comparison, mutual funds have 21 trillion dollars in AUM (Pisani, 2020). ETFs represent a range of passive investment products since they track already existing indexes and other financial instruments. Generally, there is a trend of a switch from active to passive investment style. There are various reasons and causes why this may happen.

ETFs have become very popular among investors. However, similarly to the Subprime mortgage crisis, investors think neither about product complexity nor about causal links influencing the price, and they keep pouring funds into ETFs. That is why ETFs trading serves as an object of this thesis. However, an increase in the volatility of underlying assets is not the only negative impact that ETFs can make. For example, a fire sale of ETFs may lead to a loss of liquidity. If concerns about excess volatility are correct, then better-informed parties may execute arbitrage strategies and benefit from them. Moreover, such issues like excess volatility may undermine market efficiency. That is why it is still very relevant to study this issue as it might help to understand possible weak spots and dangers to financial markets stability.

The main objective of this thesis is to study the relationship between trading volumes of Exchange-Traded Funds and excess volatility of the S&P 500 Index. It is a subject of great importance. It is so because ETFs are not well studied compared to other products, for instance, options or futures. Furthermore, they are rising in popularity, and they have not taken part in a severe financial crisis yet. Moreover, past events like the Flash Crash have also triggered further concerns about whether passive investments and ETFs, in particular, may cause issues or risks that result in a financial crisis. These are the reasons why ETFs and its components should be studied. Knowing the relationships and product’s behaviour might help to prepare the guidelines and restrictions to prevent severe consequences.

Research tasks of the thesis include a review of relevant theoretical literature and an overview of empirical studies. Research tasks also include data sample collection for testing purposes. Then, it is needed to perform regression analysis and estimation of econometric models for hypothesis testing. The hypothesis to be tested is whether dollar trading volumes of ETFs have any impact on volatility, and so cause excess volatility of the S&P 500 Index. Moreover, this analysis will shed light on whether changes in trading volumes of ETFs adversely affect underlying assets. Thus, ETFs intensify volatility and so potentially create arbitrage opportunities.

For hypothesis testing, a regression analysis is used to tests the linear relationship between ETFs trading volume and volatility of an underlying S&P 500 Index. Then vector autoregressive models are used for analysis. These models enable simultaneous relationship testing between the variables, and so it makes possible usage of the Granger causality test. We estimate the models constructed for both unconditional and conditional volatility, and the author has used the GARCH model for the latter.

Key building blocks of the first chapter of the thesis are a theoretical overview, comparing and contrasting active and passive investments, and a discussion of the current issues for passive investing products which include ETFs. ETFs themselves are explained, looking at their product specifics and the extent to which they are similar to or different from open-end mutual funds and closed-end mutual funds. Lastly, the role of ETFs in a Flash Crash in 2010 is briefly covered. The second chapter focuses on the methodology and data used, with an overview of the key empirical studies based on which the hypothesis of this thesis is constructed. An explanation is given of how the variables are determined and of the methods used for hypothesis testing. Finally, the third chapter includes all the econometric model outputs for daily and monthly data, as well as analysis of the overall results. The thesis ends with the summary.

1. THEORETICAL OVERVIEW OF ETF RELATED INVESTMENT ASPECTS

Historically there were two main investment approaches. Therefore, it is crucial to differentiate between two of them, namely active investing and passive investing. However, it should be noticed right away that having two polar opposite methods has brought a mixture of two that is a semi-active approach (Cremers and Petajisto, 2009). This approach combines features of both active and passive. In this chapter, the author will elaborate more on two different investing approaches, on its specifics and a recent rise of popularity of one approach over another. Also, the author will shed light on a semi-active investing style.

1.1 Active vs passive investing

Active and passive investing have many features that will be discussed below. To provide an understanding of what these two investment styles represent, a brief explanation of them both is as follows. Active investing assumes an active tracking of what is going on in the market. It involves a sophisticated analysis of the investments, prices, market developments, roots and causes of different events that affect security prices. Based on this analysis, active investors actively make changes in their portfolios. They buy and sell assets, change proportions in which they hold certain securities. Such adjustments are made with the purpose to earn greater risk-adjusted returns. On the other hand, passive investing assumes index tracking of indexes that consist of numerous stocks or any other financial assets. When following this approach investor may get positive returns as a result of a by-side effect. It means when stocks in the chosen index gain profits, passive investors also earn money. However, as it was mentioned already there is also a blend type named semi-active investing approach. For example, a semi-active investing approach is known for "closet-indexing" (Cremers and Petajisto, 2009). Closet-indexing is a notion when investors or funds claim to be using active strategy but, in reality, they track an index benchmark, so they are using a passive investing style (Cremers and Petajisto, 2009). According to BlackRock (2017a), some strategies are focusing on specific factors, be it value or volatility. Before-mentioned strategies are also called "smart betas", they are initially created using active selection, but later they are implemented passively (BlackRock, 2017a). In their study, Cremers and Petajisto (2009), find evidence that funds claiming to be active are, in fact, in many cases just replicating a benchmark. They have used active share to find how volatile the return is, compared to a benchmark index and tracking error, and so to find out how different is the portfolio construction from the benchmark index (*Ibid.*). The result was that indicators were similar to the ones of a benchmark (*Ibid.*). And so it can be concluded, that at times it may be hard to differentiate between actually active and passive investing styles.

One of the key distinctions in assumptions behind active and passive approaches is their relation towards Efficient Market Hypothesis or EMH developed by Eugene Fama in 1970. Efficient Market Hypothesis represents the idea that asset prices fully reflect all available information to all parties, and so nothing is hidden from investors, prices follow a random walk, and thus no one can beat the market (Fama, 1970). There are three forms of market efficiency described by EMH, these are strong, semi-strong and weak forms (*Ibid.*). Strong form assumes that historical, publicly available, and private information are taken into account by the market and reflected in asset prices (*Ibid.*). In this case, asset price data is white noise, so arbitrage opportunities do not exist and future prices cannot be predicted (*Ibid.*). Semi-strong form in its turn reflects historical and public information in asset prices (Latif et al., 2011). It means that knowing private information investors may use arbitrage opportunities in their favour, but insider's trading is banned yet technically it is still possible (Latif et al., 2011). One more form is a weak form which states that the market only reflects historical data, and so there should be a great number of market anomalies and inefficiencies (Latif et al., 2011). Generally, it can be said that active investors believe that markets are inefficient and so there are possibilities for arbitrage. That is the main reason why they seek alpha. On the other hand, passive investors support the Efficient Market Hypothesis and instead of trying to outsmart the market, they merely follow its successes, but also downfalls. Certainly, it can be argued that a strong form of Efficient Market Hypothesis exists, when both private and public information is accounted for in stock prices, but often when any unanticipated crises occur, the EMH proves to be flawed, as it is hard to predict any unexpected future events (Thaler, 2009). Consequently, investors adopt an active investing approach either investing themselves or paying high fees to fund managers because there is a generally prevailing idea that one can outsmart the market (Thaler, 2009). In addition to this, there is a Grossman-Stiglitz paradox which favours active investing. This paradox states that if securities incorporate all information into future prices, then information cannot be costly, since there is no reason to pay extra for something that provides no arbitrage opportunities (Grossman et al., 1980). However, markets should be competitive to exist thus not all info is reflected in prices, and traders are ready to pay for information that benefits them in taking a position that earns profits, hence markets are not efficient at all times (Grossman et al., 1980).

Investment horizon is another key difference between active and passive investing. Active investors focus on a shorter investment horizon and try to outperform the market, whereas passive investors have a long-term focus. However, as it was noticed by Benjamin Graham in his book "The Intelligent Investor" investing is a long-term activity based on a proper fundamental analysis, and so frequent adjustment to a portfolio of financial securities should be treated as trading, which is more of speculation and similar to gambling (Graham and Zweig, 2003). Moreover, Graham and Zweig (2003) have noted that trading cannot be a profitable endeavour over a long period, but many proprietary traders apparently disagree with this statement as they keep making profits and chasing returns by applying active strategy. Yet, it is worth remembering words of wisdom that an individual may lose its solvency a way before markets get back to a rational state.

Moreover, both investing styles have highly ranging costs for implementation. The active strategy involves frequent purchases and sales of assets, this entails payment of different fees like brokerage commission, transaction fee or management fee in case of a fund. Besides, Barber and Odean (2001) have found that frequent manipulations with individual stocks by individual investors result in lower returns. As for passive strategy, here one spends once and may hold a stake in an index. Perhaps, in this case, commission fees can be smaller. However, it may not be the case when an index is not tracked directly but rather constructed. In this case, investors still should pay a management fee if such a portfolio is constructed for them or they should pay brokerage fee and transaction fee when they update the portfolio in accordance with an index composition.

Furthermore, there is such a notion as portfolio rebalancing that is rather relevant for passive investing style and might be costly (Bohn and Tesar, 1996). With a passage of time weights of an asset held in an index change, therefore investors should adjust their portfolio weights in accordance with an index, be it on an annual basis or more frequently (*Ibid.*). They should either buy or sell some proportions of financial securities to align weights. That is a costly process for individual investors, consequently, direct tracking and pure passive strategy should be less costly, but those costs can be distributed and so minimised when index tracking is performed by a fund. However, Bohn and Tesar (1996) have also found that greedy investors might step back from the initial strategy and they prefer to hold return generating a stock for longer, forgetting to maintain required balance in the weights.

A novelty known as socially responsible investing or interchangeable term ESG investing is one more aspect that may affect a strategy cost. ESG stands for environment, social and governance. This type of investment means that investors purchase stocks that belong to companies compliant with ESG framework, in other words, those companies care for the environment, on its production sites they minimise waste and pollution, they give back to communities where they operate, set support packages and benefits for its employees, financially support foundations that help people in need or invest for education purposes, and the ways of corporate governance in such companies are very transparent as well (Matos, 2020). Therefore, ESG investing sets certain limits, and investors will not buy stocks of oil companies or coal mining companies, they will rather prefer those using solar or wind energy thus they should be very selective about what they invest in. Hence, it is costly, likewise a refusal from a great number of potentially profitable investment opportunities. One may argue that ESG investing brings higher profits as previous academic studies have revealed that there is a “link between ESG scores and financial returns” (Halbritter and Dorfleitner, 2015, p.30). However, research on recent data questions these findings and claims that ESG is no longer statistically significant, and so it does not generate superior returns (Halbritter and Dorfleitner, 2015).

Active and passive investing styles have coexisted for almost half a century now. Until recently active investment style was noticeably dominating in the market. Moreover, investors expected higher returns from active strategies and so were ready to spend time and money seeking positive alpha. However, it can be observed that interest in passive investing has risen significantly over time and that a shift from active to passive investing has already happened. To add, ETFs refer to the passive investing style and popularity of ETFs is rising now. For example, “from 1995 to 2017, cumulative net flows to passive MFs and ETFs totalled $4.2 trillion, compared to $2.4 trillion for active funds” (Anadu et al., 2018, p.1). According to Anadu et al. (2018), several factors have contributed to a shift from active to passive investing style. The first factor is an introduction of the EMH hypothesis in the mid-20th century that questioned the superiority of an active investment style (Anadu et al., 2018). In addition to this, a study by Mehra and Prescott (1985) in which authors find the existence of Equity Premium Puzzle that states that on average well-diversified portfolios of stocks earn 6 per cent higher returns than portfolios consisting of governmental bonds, and this percentage is much higher than economic models predict, also risk aversion is not an explanation because levels of risk aversion should be extremely high for such premium. And this is a reason why investors are so keen to both actively invest but also simply passively follow the market because one of the justifications is that stock investors regardless of their investment style get compensated for risk-taking. Furthermore, the first stock index fund, now called Vanguard 500, was established by John Bogle in 1976. This new product enabled retail investors to invest in a basket of top picked stocks as it was tracking the performance of the S&P 500 (Anadu et al., 2018). In addition to this, passive investors when they buy or sell ETF units, they trade "the entire basket of index constituents in response to fund inflows and outflows", therefore effects like volatility may affect not just an underlying index but also individual securities that compose this index, and as a result cause price discrepancy and overall market inefficiency (Sushko and Turner, 2018, p.119). Also, passive investing was and remains more compelling due to its lower fees. (Anadu et al., 2018) However, some other insights on this matter are that passive investing is rising simply because of closet indexing, but also because some active positions are getting cancelled out when executed within the same mutual fund (Cremers and Petajisto, 2009).

To sum up, in pure form active and passive investment styles are very different. They differ by the way of investing, by the investor's justification for why a certain strategy is more advantageous. Nevertheless, lines between those two types of investing are getting blurred, and sometimes it may be difficult to say if investors follow an active or passive strategy of investing. Also, active and passive investing strategies may differ in cost, but yet, in general, with respect to investing in indexes, passive investing should be less cost-intensive and time-consuming compared to active investing style. However, there is a clear switch from an active investing to a passive, and it is only left to study how such changes might influence prices of financial securities and market as such, whether this will be in favour of investors or this will bring more issues and, perhaps, eventually will lead to a financial crisis.

1.2. Current issues related to passive investing dominance

In this paragraph, current issues that are related to passive investing will be discussed, the relevance of the discussion is that these issues may endanger passive investments and possibly well-being of a financial system as a whole. The most crucial problems to be discussed are, for example, lack of liquidity, redemption-related risks, the problem of a co-movement, pricing differences coming from inclusion effect and volatility, industry excess concentration due to passive investing dominance. In addition, other concerns by industry professionals like unpredicted behaviour of passive investments in case of market turmoil and the possibility of a passive bubble burst.

The first issue to be discussed is well covered by academic literature and that is a liquidity risk that may damage whole financial stability. The idea of passive investing on the first sight looks simple and safe enough, but when favourable conditions change, consequences may be dramatic, both for individual and institutional investors. In case when the market situation deteriorates stockholders may decide to sell off their passive holdings, meaning the supply of the shares on sell will increase, or volatility can increase as well, which will lead to the assets price drop (Financial Stability Board, 2017). It means assets will become illiquid and unwanted. Hence, those who are last to react to market changes will lose money in parallel with how assets lose value. Though for the worst-case scenario to materialise considerable changes should happen.

Another challenge with liquidity can be in the rules for the redemption of the shares. It is known that from open-end funds investors can redeem shares directly, whereas from closed-end funds it is allowed to redeem shares by selling them to other investors, such that fund itself is not involved in the redemption process. At the same time, in such passive investing funds like ETFs shares can be redeemed via authorised participants only, but they may refuse to do so in case of financial distress when assets become illiquid or partially lose their liquidity (Goldstein et al., 2017). On the other hand, it is argued that ETFs reduce liquidity transformation because they use in-kind redemptions, which means they do not return cash to investors but give back shares to those investors who decided to part ways with the fund (Anadu et al., 2018). Since less cash is involved, the likelihood that flight to liquidity may happen in case of economic downturn remains low.

One more issue is related to a redemption risk. It, paired up with liquidity mismatch, can further trigger asset sales and might lead to financial difficulties and even further faults in the financial system (Bank of England, 2019). It is said that liquidity mismatch may exist because, for instance, ETFs hold less liquid assets but provide redemption on a daily basis as it is required by the Investment Company Act of 1940 (Anadu et al., 2018). Such funds make pay-outs immediately, but it takes time for them to sell assets they hold. Some less liquid assets that are more difficult to sell are, for example, corporate bonds, real estate or emerging markets assets (Bank of England, 2019). The opposing argument can be that ETFs, unlike open-end funds, can trade on a secondary market, which, in turn, gives investors more chances to close positions since underlying assets will not be affected as much as fund units (Bank of England, 2019). The real life examples include several cases when funds suspended redemptions after several waves of increased redemption levels caused by the Brexit uncertainty (*Ibid*.). One more example is that in 2019 LF Woodford Equity Income fund was suspended after several years of funds outflows and a substantial request of one of its investors to redeem assets as such action would put at risk assets of this fund's other investors (Bank of England, 2019). It was later decided by the Financial Conduct Authority or FCA to liquidate this fund and return stakes to investors (FCA, 2019). One more issue related to redemption is a customised basket. Generally, ETFs replicating some index should hold assets in accordance with what is included in the index, and in corresponding weights. However, in reality, ETFs' managers were not one hundred per cent matching ETFs with indexes they track, and so ETFs were not holding pro-rata baskets but customised ones (Kaminska, 2020). It further enabled authorised participants to use this for their benefit.

There are authorised participants or APs who exchange assets with ETFs as they execute purchase and sell orders. These authorised participants may use their favourable position to secure arbitrage profits. For example, they may persuade ETF managers to buy certain assets from them, and in turn, ETFs would give some more liquid assets of their own, and this would be a dumping scenario for APs (SEC, 2019). Another scenario that can accompany dumping is cherry-picking. In the case of a cherry-picking authorised participant would urge ETF managers to accept unwanted assets in exchange for the assets that ETF itself want to redeem (SEC, 2019). The ultimate result could be that ETFs are left with customised basket comprising of several illiquid securities in this basket (SEC, 2019). Illiquidity may lead to numerous problems and may endanger the fund's existence as such. However, to remove such arbitrage opportunities SEC has issued a new rule for ETFs requiring to present accepted policies stating in what circumstances fund may resort to using custom baskets, and most importantly fund should prove that it acts in the best interests of its investors (SEC, 2019).

An increase in a co-movement between an index and underlying assets included in this index is the next possible issue that can occur due to a switch from active to passive investing. In this case, co-movement is defined by positive covariance that is a tendency of stocks to move together in the same direction. Two main explanations for this are the inclusion effect and arbitrage activities of authorised participants.According to a study of Claessens and Yafeh (2012), there is evidence that after asset inclusion into an index, Beta and R-squared estimates increase. The first, Beta, is used as a systematic risk measure, and an increase of another statistical measure R-squared or R^2

is used as a measure for an explanatory power of the model or in other words it also shows the relative importance or significance of a model. These findings are correct for both daily and weekly data. Furthermore, Da and Shive (2017) find that there is a positive relationship between ETF ownership and co-movement of index assets with the underlying assets in the market. Moreover, inclusion in the index entails "an increase in stock turnover and analyst coverage" (Claessens and Yafeh, 2012, p.212). Another empirical evidence suggests that authorised participants can also contribute to an increase in co-movement of an index fund and underlying. Authorised participants can take either long or short position in ETF and an opposite in the underlying assets. APs can, for example, buy ETF units at a discount when available and sell short underlying assets by borrowing them from institutional investors (Da and Shive, 2017). Another option is to buy underlying shares on a secondary market, but redeem ETF units at the end of a trading day directly from fund managers without entering OTC market, and as a result, such action could imply overly co-movement of assets (Pan and Zeng, 2017).

However, there is one more evidence suggesting that in addition to excess co-movement there is an effect of convergence. Convergence means that individual stock with Betas below or above one after inclusion into index tends to go up or down, respectively, and over time reach Beta of one (Claessens and Yafeh, 2012). In their study, Claessens and Yafeh (2012) conclude that in the past evidence suggested that there was an increase in Beta of individual stocks regardless of initial levels, they have also discovered a presence of convergence and tested that this is not due to measurements errors since parameters are statistically significant. However, it is unclear for how long this effect lasts and whether mean reversion in Beta value eventually takes place. Nonetheless, it is known that Beta of one indicates that stock is as volatile as the market, it moves together with the market and is exposed to the same risks. Of course, one should bear in mind idiosyncratic risks as well. Nevertheless, increased co-movement might, in turn, increase systemic risk and so volatility for some assets. Authorised participants can influence the magnitude of co-movements and execute arbitrage strategies, thus can eventually lead to inefficiency of diversification of portfolio and, perhaps, to financial distress (Sushko and Turner, 2018). And if researchers claim that excess comovement has a timely effect and so later initial correlation restores, this can also be used by arbitrageurs.

Pricing differences are another issue documented by academic researchers. For example, there is some evidence that the index inclusion effect leads to changes in stock prices. A positive change can contribute to the popularity of passive investing and so lead to a financial bubble of overinvesting into indexes setting active investing aside (Anadu et al., 2018). On the one hand, according to Sushko and Turner (2018) inclusion effect implies a narrowing of a bid-ask spread and an increase of trading volumes. In general, narrow bid-ask spread means more active trading and higher market competition might also indicate better liquidity. It is tested, and such evidence is found on the example of the S&P 500 Index (Sushko and Turner, 2018). In addition, Shleifer (1986) has conducted an event study to examine an inclusion effect, and he concluded that stocks experience an increase in prices of roughly 3 per cent as a result of inclusion into S&P 500 index. However, recent evidence shows that inclusion effect is much lower than in earlier studies. On the other hand, Shleifer (1986) documented that this effect has lasted from ten to twenty days, though back then data did not let test for a longer duration of the effect. Moreover, Patel and Welch (2016) made similar research using more recent data and concluded that after inclusion into index asset prices jump up, but today this holds for a much shorter period and reverses back to original levels. Such price distortions undermine the idea of passive investing. Due to this, active investors come into play and may enjoy profits from such price jumps and reversals. Several academic articles find evidence for inclusion effect, but the final effect on passive investing remains unclear. It can both induce index investing and so lead to a bubble or vice versa switch attention towards active investing and so make the index bubble burst.

Excess volatility that creates further problems and endangers the whole financial system is one more major issue caused by passive investments. And, as already mentioned, the volatility effect is the central problem of this Master thesis. It is crucial to note that other aspects discussed above could generate excess volatility, though often indirectly. There are certain criteria based on which stocks are first considered, and then might be included into an index. However, the vast majority of these stocks are issued by big companies and in large amounts are actively traded in the market. Moreover, Cumberland Advisors analysed that in 2018 securities were held by traders for an average of 20 seconds compared to two months in 2008 (Greene, 2018). Such frequent trading activity of baskets with various securities leads to an increase in volatility, also may amplify price discrepancies between index units and underlying assets (Anadu et al., 2018). According to Ben-David, Franzoni, and Moussawi (2018), ETFs affect the distribution of underlying asset returns, in the short-run, they amplify volatility and increase firm-specific risk, and, as known, for the higher level of risk, investors demand a risk premium. It was already mentioned about arbitrage mechanism available to authorized participants, but it is also worth mentioning that this is possible due to price deviations between the price of ETF and ETF's NAV or Net Asset Value at the end of a trading day (Pan and Zeng, 2017). Furthermore, ETF units can be relatively easily exchanged for the underlying assets with the help of redemption mechanism, so prices should be close and priced fairly enough, but this is not the case and in reality, pricing discrepancy may persist (Malamud, 2015). However, it is also argued that the introduction of new ETFs may offset issues created by old ETFs, and so improve liquidity and reduce co-movement and volatility (Malamud, 2015).

Increased industry concentration in passive investments is yet another problem. Anadu et al (2018) empirically tested this, such that they applied Herfindahl-Hirschman Indexes for testing industry concentration for both active and passive funds, particularly for mutual and exchange-traded funds. It is commonly accepted to treat industries with Herfindahl-Hirschman Indexes of less than 1500 as with low concentration, higher than 2500 as concentrated, and with numbers in between as with mild concentration (U.S. Department of Justice, 2018). The end result was that for the period from the year 2004 to the year 2018 on average passive funds had 2800 HH index points, whereas active funds had 450 HH index funds (Anadu et al., 2018). This means that passive investing has contributed to the rise of levels of industry concentration, meaning a great number of potentially great investments remain out of focus simply because their specification or features do not go along with the funds' criteria.

In addition to this, there is a concern of industry professionals who notice signs of a passive investing bubble. Michael Burry is known to a large audience as the one who has managed to foresee the subprime mortgage crisis and made a fortune by betting against the market. He warns that index and ETF investors are too narrowly focused on big companies and pour money into them, whereas small-cap companies are out of their sight (Reinicke, 2019). This has led to a situation when small-cap stock is left aside and because of this, they may lack liquidity. Even though this far passive investing was a profitable option, at some point, a market situation may reverse. Howard Marks stresses that ETFs were successful due to overall market growth, but it remains unclear how they will act in market turmoil (Reinicke, 2019). COVID-19 indeed can be a litmus paper for ETFs, such stressed scenario as pandemic can be an ideal condition for testing ETFs to reveal how they act during markets turmoil and for testing views mentioned above. Nobel winner Robert Shiller points out to a problem of free riders who enjoy profits by tracking portfolios of other people who made all analysis and choices for them (Reinicke, 2019). Lastly, father of passive investing John Bogle, who has created first fund tracking S&P500 performance, states that when passive investments outnumber active investments, such density will not result in anything positive (Bogle, 2018).

Overall, a switch from active to passive investing has brought new opportunities for investors but also risks and problems associated with these investments. Academic research has documented most of the issues mentioned above, and so these issues certainly come out with different frequency and magnitude. Until recently, economic conditions were quite favourable as markets and country economies were growing well. However, a reversal in this trend or appearance of a Black Swan like coronavirus may intensify issues faced by passive investments, such that investors will lose money when the market downturn happens. In the essence, it means that if an event with an ultra-low probability materialises, investors will lose an unexpectedly huge amount of money, much bigger than an expected shortfall or similar downfall risk measure expressed in money terms (Bali et al., 2009). A recent study of Bank of Canada on fixed income ETFs unveiled that in March when Corona crisis has started, financial markets have experienced a stress scenario during which volatility has increased significantly (Arora et al., 2020). As a result, APs have reduced its redemption activity for fixed income ETFs such that these funds acted as closed-end funds because increased trading volumes and so activity was mainly among secondary market participants (Arora et al., 2020). Hence, liquidity levels were high, and price discovery process continued working as usual, but what catches attention is a conclusion that APs have contributed to fixed income ETFs being sold at discount (Arora et al., 2020). However, more studies about the corona crisis and its implications for financial markets are to come since this crisis is still very new. Having said how different active and passive investment approaches are and how many issues can be potentially connected to passive investing and ETFs, in particular, it is then crucial to study different relationships between ETFs and underlying assets or related investment products. That is why it will be tested if the following hypothesis finds its approval or should be rejected. The hypothesis to be tested states that dollar trading volumes of ETFs have no impact on the volatility of the S&P 500 Index. Considering the scope of problems such as liquidity, volatility or positive variance faced by investors, the topic remains hot and should be studied well for the possible causalities and consequences to be able to diminish negative outcomes.

1.3. ETFs and its key characteristics

In previous chapters, the difference of two investing approaches, active and passive, is considered. Also, issues related to passive investing are covered in more detail. In this chapter, the author will take a closer look at exchange-traded funds or ETFs as these are currently one of the hot topics, and they are also bright representatives of the passive investing style. Thus, it is essential to define, more specifically, what are ETFs and what are the characteristics that define them. This chapter will also shed light on how ETFs function and what makes them different from other index funds.

Exchange-Traded Funds are investment vehicles that track the performance of a basket of underlying financial securities or indexes (Lettau and Madhavan, 2018). First index funds were introduced in 1970, but then "the first US-listed ETF, the SPDR, was launched by State Street in January 1993 and seeks to track the S&P 500 index" (Lettau and Madhavan, 2018, p.135). ETFs are listed on stock exchanges, and when buying its shares, investors in one instalment get access and exposure to a great number of different financial assets, at the same time, minimising risks from individual securities (BlackRock, 2017b). Additionally, the fact that they can be traded on a secondary market close to its Net Asset Value means they provide better liquidity (Hill et al., 2015). Generally, ETFs should be way cheaper than any active funds, because in case of active funds fund managers demand a management fee which is common to be paid on annual basis and expressed in percentage terms from the assets under management, but the result is not necessarily successful in terms of moneyness of positive returns (Anadu et al., 2018).

One of the specifics lies in ETFs' mechanics. Just like mutual funds, they refer to index investing, but the difference in the way they function and in the way trading happens. As for open-end mutual funds, the value of its shares is evaluated using Net Assets Value or NAV at the end of the day and so NAV reflects the value of underlying assets (Lettau and Madhavan, 2018). Trading, along with, value recalculation happens at the end of the day, the fund itself controls money flows and a number of shares (Lettau and Madhavan, 2018). There is then a different story with closed-end mutual funds, in such funds number of shares is fixed, so the investors can only trade shares with each other on the market, not from the fund, but fund remains very active in its interaction with the market (Lettau and Madhavan, 2018). As for the value of shares, it can be different from NAV, at a discount or a premium, because the price reflects an expectation of investors, how they evaluate market conditions, how confident they are in fund manager's actions (Lettau and Madhavan, 2018). Similarly, to open-end mutual funds, exchange-traded funds price their NAV value that should always equal the fair value of underlying assets (Hill et al., 2015). Pricing happens at the end of the trading day, but the intraday price is determined by the interaction of buyers and sellers on an exchange (Lettau and Madhavan, 2018). However, the difference is in a shares creation and redemption mechanism, exchange-traded funds work directly with authorised participants usually large financial institutions who in turn issue or redeem shares and so cooperate with the market and they also bear costs of doing so (Hill et al., 2015).

ETFs are normally treated as passive investments, but there are reasons to believe that they can be of a mixed style. The fact that ETFs refer to passive investments means that they buy and hold an index, tracking its price both up and down, and enjoying profits when the value rises or when dividends pay-outs happen. It also assumes that no active changes to a constructed portfolio consisting of underlying stocks or indexes happen, but as it was discussed before this is not always the case and ETFs may also be of a mixed investing style (Cremers and Petajisto, 2009). Specifically for ETFs, it means that an initial setup of a strategy might be very active, but then should follow a stage of passive tracking. The following example also shows that the distinction of ETFs style is not evident. Apart from big ETFs like Vanguard or iShares, there are also smaller ones that track so-called "hot sectors" (Stepek, 2020). This means that such ETFs choose companies that operate in sectors promising innovation like "robotics" or trendy topics like "medical cannabis" (*Ibid*.). Basically, ETFs increase in size by building up their success on things that are hot and trendy, in which many people currently invest in (*Ibid*.). And again, initially, there is a choice of what companies should be included and tracked by the ETF, and that is already an active choice, thus the strategy is no longer that passive as it is claimed to be (*Ibid*.). It can be seen that ETFs, despite its simplicity at a glance, are a complex product. Perhaps, understanding of how traditional equity ETFs function is rather straightforward, but there are more sophisticated products available. In the case of the latter, hardly every investor can explain what stands behind a fancy term. In the vast majority of ETFs, equity shares or bonds serve as underlying assets. For example, there are Exchange Traded Notes or ETNs that use debt securities as an underlying, Exchange Traded Commodities or ETCs that track commodities like oil, gas or wheat, and Exchange Traded Instruments or ETIs that add extra features upon return from underlying instruments like an inverse return or "knock-out" feature, and much more similar to specific features of options (U.S. Securities and Exchange Commission Fixed Income Market Structure Advisory Committee, 2018).

ETFs are rising in popularity, shift to passive has happened and academics have started to pay more attention to any issues that are related to those topics, many of those issues were discussed in the previous chapter. However, there is a particular event that happened in the market in 2010 and this has triggered all the questions and worries about where passive investments are heading to. This event has also raised concerns about whether ETFs are the cause of the next financial crisis. Let us consider what has happened in 2010 and what was the role of ETFs in that market crash. Flash Crash happened on 6 May 2010, that was a less than one-hour distortion in work of financial markets. Of course, there were some other Flash Crashes but in this particular ETFs were heavily involved. And how come crises can have such narrow timing, that is because algorithmic trading and high-frequency trading are widely used nowadays and they enable frequent trading and in huge volumes. A great number of US equities and biggest indices experienced price distortions, many securities were traded at roughly 40% of their opening prices at that day (U.S. Commodity Futures Trading Commission and the U.S. Securities and Exchange Commission, 2010). Moreover, many financial instruments have suffered from a flight to liquidity and increased volatility, so VIX index that represents the volatility of the S&P 500 increased by 31,7% from the starting levels of that day, "which was the fourth largest single day increase in VIX" (U.S. Commodity Futures Trading Commission and the U.S. Securities and Exchange Commission, 2010, p.9). What happened was later investigated by the U.S. Securities and Exchange Commission and U.S. Commodity Futures Trading Commission. They identified that the main cause was that a large trader had launched a sell algorithm that took into account trading volume, but some other important inputs like price were not taken into consideration (*Ibid.*). This algorithm was intended to sell 75000 E-Mini futures contracts that use S&P 500 as an underlying, and a great number of futures contracts were sold in a very short period of time (*Ibid.*). This event has triggered all the questions and concerns about ETFs. SEC has studied if there is a causality between changes in liquidity of the S&P 500 and its effect on prices of ETFs (U.S. Commodity Futures Trading Commission and the U.S. Securities and Exchange Commission, 2010). Ben-David et al. (2018) have studied the vice versa relation, namely how changes in ownership of ETFs affect the underlying S&P 500 Index. In the former case, there was no evidence for such causality, but the latter has confirmed that changes in ownership of ETFs do affect volatility in the index. In addition to this, investment professionals questioned by the SEC have noted that ETFs like from their practical experience changes in E-Mini lead to changes in SPY and as a result in underlying securities (U.S. Commodity Futures Trading Commission and the U.S. Securities and Exchange Commission, 2010). However, the SEC has not found statistically significant evidence for that Contrary to that view ETFs are believed to be a good alternative for pricing other financial instruments like fixed-income products or indices (Hill et al., 2015). As an example, Hill et al. (2015) mention Arab Spring during which there was also a financial turmoil in the Egyptian market and ETFs were used as an indicator for investors' expectations about the market and were used as a mean for financial securities pricing (Hill et al., 2015). BlackRock (2020) also states that ETFs do not negatively affect the pricing of stocks because ETFs are in the vast majority traded on exchanges directly and underlying instruments are not involved in this process.

To sum up, ETFs are complex products, but investors hardly understand what they buy and trade. ETFs share some similarities with mutual funds but they are stand-alone investment products with their specific features. Not all of the ETFs are passive, though in most cases they are. And ETFs are currently one of the concerns as they are sometimes believed to be the cause of the next financial crisis. Also, ETFs set a lot of controversies as some say they are beneficial for the markets and its participants, while others believe that they cause price distortions and cause market inefficiencies. These are the reasons why ETFs are also a hot topic for research.

2. DATA AND METHODOLOGY

In the second part of this Master's thesis author plans to review past studies, discover what they have tested and what evidence they have found, also what methods were used. Based on the collected knowledge the most appropriate methodology will be chosen with the purpose to perform hypothesis testing. One more objective is to describe the gathered data and provide descriptive statistics. Moreover, all the used variables and how they are constructed will be explained.

2.1. Review of empirical studies

Exchange-Traded Funds are relatively new products because they did not become immediately popular among investors. That is why they were not in the focus of researchers. However, rising popularity changes an attitude towards ETFs. It also raises concerns and so makes ETFs a hot topic for research. There are still not many articles about ETFs, its relationships and behaviour. Below are some of the most relevant for this thesis and once with results worth considering. The initial idea comes from the article written by Ben-David et al. (2018) and similarly to them the goal is to test how changes in ETFs affect underlying assets. Let me elaborate on what Ben-David et al. (2018) have studied in their research and what conclusions they have reached. They have studied the relationship between ETFs ownership and the volatility of the underlying S&P 500 Index. They have found that ETFs with higher ownership levels affect underlying indices and cause higher volatility of them. Higher ownership, in this case, means index membership or, in other words, it represents a proportion of an underlying index held by the ETF. Increase in volatility is paired up with an increase in a firm-specific risk, but it is equally true that higher risk is rewarded with higher premiums (Den-David et al., 2018). “Economically, a one-standard-deviation change in ownership is associated with 16.4% of a standard deviation change in daily volatility” (Ben-David et al., 2018, p.2493). They have tested such effect by running an ordinary least squares or OLS model. Daily volatility of an index was chosen as a dependent variable. There were many independent variables. As for statistically significant, these happened to be: ETF ownership, lag of logarithm of Market Cap, lag of unit over Price, lag of Amihud ratio. As for the remaining variables, they were giving mixed results depending on a chosen sample. For example, lag of bid-ask spread, lag of book to market ratio, lag of past twelve months return, and several lags of volatility were used as additional testing variables.

Another study run by Hasbrouck (2003) examines causes and effects between ETFs and its underlyings. It uses co-integration analysis and VECM to explore a relation between S&P 500 ETF, Nasdaq-100 Index, S&P 500 Index, S&P 400 MidCap Index and E-minis futures contracts (*Ibid*.). The purpose of this study is to determine the price discovery process, where it originates and how it affects other securities (*Ibid*.). Hasbrouck (2003) states that VECM might be easy to estimate but tricky to interpret, and so he advises to use impulse response in this case. His findings reveal that price origination starts in futures and then translates to S&P 500 and Nasdaq-100. As for the S&P 400 MidCap, its price responds to changes in ETFs (*Ibid*.). ETFs themselves are traded in vast volumes. Though, its investment activity does not help to predict future prices of underlying indices. (Hasbrouck, 2003).

A study by Switzer et al. (2000) also examines the pricing efficiency of the futures market. They have tested asset prices in periods of pre-inclusion and post-inclusion into SPDR S&P 500 exchange-traded fund. For the pricing efficiency identification, they have applied the Cost of carry model, namely compared spot and theoretical futures prices to identify mispricings. They have used daily and intraday data for the period of find and a half years. Switzer et al. (2000) applied the OLS model and dummy variables to capture pricing inefficiency. Their findings reveal that before the introduction of SPDRs, there was small by magnitude but statistically significant mispricing (*Ibid*.). However, post-inclusion into SPDR demonstrates that pricing error has diminished, so SPDRs have improved market efficiency (*Ibid*.).

Xu and Yin (2017) have written an article that tests the relationship between ETFs and its underlying assets. They have used the S&P 500 Index and modelled its realized variance and conditional variance or volatility using GARCH(1,1) model (*Ibid*.). Then, taking the biggest ETFs that track the S&P 500 Index, they have aggregated trading volume (*Ibid*.). They have tested a linear relationship between realized variance and trading volume of ETFs, also they have used conditional variance instead of realized variance (*Ibid*.). As a next step, they have applied Granger-causality test that is based on VAR model and tested cause and effect between the volatility of S&P 500, trading volume of ETFs, momentum and illiquidity ratio, the latter is based on Amihud (2002) formula (Xu and Yin, 2017). Similarly, they have used other indexes and ETFs tracking them and tested a linear relationship between them (*Ibid*.). Ordinary least squares model and GRACH approach indicate that changes in ETFs' trading volumes imply further changes in the volatility of an underlying index (*Ibid*.). The results are relevant for both daily and monthly data. VAR model results suggest that the trading volume of ETFs and volatility of corresponding indexes have a bi-directional relationship, and so do their lags (*Ibid.*). Additionally, the Granger-causality is used on other indices, namely S&P 200 Index, the CAC 40 Index and FTSE 100 Index (*Ibid.*). The result is similar to the S&P 500 Index, so it states that Granger-causality works in both directions (Xu and Yin, 2017).

Lin and Chiang (2005) studied the volatility effect of ETFs on constituents of the Taiwan Index. They assessed an impact of Taiwanese Exchange-Traded Fund on the volatility of Taiwan Top 50 Tracker Fund or TTT (*Ibid.*). They have used realized variance or ABDL volatility and unconditional variance modelled using GARCH (1,1) for the Taiwan Index Fund. In both cases, Lin and Chiang (2005) have used intraday data with five-minute interval. Therefore, they have applied a dummy variable that divided the test sample on index pre-inclusion and post-inclusion batches. ANOVA analysis outcome is that inclusion effect exists for both realized and unconditional variance because the volatility of index constituents has either increased or decreased, but the change was significant (*Ibid.*). Roughly 60% of index constituents had an increase in volatility levels (*Ibid.*). However, the volatility effect may differ among sectors (*Ibid.*). For example, there is evidence for a stronger volatility effect for financial and electronic sectors (*Ibid.*).

Sullivan and Xiong (2012) have conducted research on passive investments and their influence on markets stability. They have used daily data like returns and trading volumes on stocks with a market capitalisation over 100 million dollars, data about U.S. ETFs and mutual funds. The regression analysis has identified a rise in beta values in ETF constituents over time (*Ibid.*). The main conclusion from this research is that a rising trend of passive trading, especially ETFs trading, leads to an increase in market risk or systemic risk, that is non-diversifiable (*Ibid.*). Consequently, other tests also show that assets returns are stronger correlated and so it is harder to diversify appropriately and mitigate risks (*Ibid.*).

Along with the rise of passive investing popularity, researchers pay more attention to passive investing style and products, like Exchange-Traded Products. Articles focus its attention on various topics related to ETFs. Some study price identification process, others focus on ETFs relationship with underlying assets, and its constituents. For instance, they find that ETFs may significantly increase or decrease systematic risk and volatility. However, ETFs are not that well studied yet, hence more research on ETFs is in demand to better understand its relationships and influences.

2.2. Methodology

This research will closely follow the majority of the tests used by Xu and Yin (2017). To be precise, daily and monthly data for both trading volume and realized variance will be used to study linear relationship using OLS model, then volatility will be modelled using GARCH. Therefore, the time series will be tested for the presence of a relationship with trading volumes and other variables using both OLS and VAR models.

As it was mentioned, the OLS model, GARCH model, VAR model, and also ADF and Granger-causality tests will be used in this thesis. Below is a thorough description of the tests and models. First of all, before the very beginning of the testing phase, it is important to ensure that the data series is stationary and so can be used for testing purposes. For this, we run the Augmented Dickey-Fuller test or ADF test. ADF test is developed by Dickey and Fuller (1979) and is used to test for a presence of unit root and so for stationarity. The null hypothesis states that time series has a unit root and so it is non-stationary. Alternative hypothesis examines the absence of unit root, and so supports stationarity of time-series. The model that gets tested is derived from Dickey and Fuller (1979, p.428) and looks as follows when more than two lags added:

$$\Delta y_t = \beta_0 + \beta_1 t + \psi y_{t-1} + \sum_{i=1}^{p} \alpha_i \Delta y_{t-i} + e_t \qquad (1)$$

where,
Δy_t – difference of a dependent variable
β_0 – intercept
y_{t-1} – first lag of a dependent variable
Δy_{t-i} – differences of lagged dependent variables
β_1, ψ, α_i – coefficients
t – deterministic trend
e_t – residuals
p – lag order

This equation tests if $\psi = 1$ (Dickey and Fuller, 1979). If it is equal to one, then the null hypothesis is accepted in this case, implying presence of unit root in the time series. Additionally, there are critical values for the test, and so every time the test value is lower than the critical value alternative hypothesis is accepted, and time-series stated to be stationary (Brooks, 2008). One more important condition for test validity is that residuals must represent white noise or in other words not to be autocorrelated (Dickey and Fuller, 1979). If the result suggests that data rows are non-stationary, then taking first differences of the rows might help to make them stationary and so prepare for further tests (Dickey and Fuller, 1979). In this thesis, the author will apply ADF test with constant, and also with constant and trend

Once the ADF test is performed and if all variables are found to be stationary at its levels, one can apply OLS or Ordinary Least Squares model. OLS demonstrates a linear relationship between tested variables and it looks as follows in the formula presented in Brooks (2008, p.30):

$$Y_t = \alpha + \beta x_t + u_t \tag{2}$$

where,
Y_t – dependent variable
α – constant
β – vector of coefficients of independent variables
x_t – vector of independent variables
u_t – error term

OLS model will be applied to test the relationship between realized variance as a dependant variable and either ETFV or RETFV as an independent variable. Moreover, in this model, differenced absolute and relative trading volumes ratios will be used as independent variables. Also, each model is tested with and without control variables. An example of the tested model with control variables looks as follows:

$$RV_1 = \alpha + \beta ETFV_1 + Amihud + Mom + u_t \tag{3}$$

where,
RV –realized variance
ETFV – absolute trading volume of ETFs
Amihud – index liquidity ratio
Mom – momentum indicator

It will enable us to check the significance of each variable and how they affect the dependent variable. Furthermore, iterations of the same model are applied for conditional volatility as a dependant variable. It is expected that trading volumes appear significant, and so this will signify that it affects volatility.

As a first step of modelling conditional volatility, one can apply the Autoregressive Conditional Heteroscedastic model developed by Engle (1982). The expected value of u_t is said to be equal to zero which makes ARCH effect to be expressed as in Brooks (2008, p.387) and it goes as follows:

$$\sigma_t{}^2 = var(u_t | u_{t-1}, u_{t-2}, \ldots) = E(u_t{}^2 | u_{t-1}, u_{t-2}, \ldots) \tag{4.1}$$

where,
$\sigma_t{}^2$ – conditional variance of residuals
$u_t, u_{t-1} \ldots$ – residuals of the model and its lags

And this formula can also be written as in Engle (1982, p.994):

$$\sigma_t{}^2 = \alpha_0 + \alpha_t \varepsilon_{t-1}{}^2 + \cdots + \alpha_p \varepsilon_{t-p}{}^2 \tag{4.2}$$

where,
$\sigma_t{}^2$ – conditional variance
α_0 – constant
$\alpha_{t\ldots p}$ – coefficient of squared lagged residual
$\varepsilon^2{}_{t-1\ldots t-p}$ – lag of squared residual

Some other important conditions for the formula above are that constant should be bigger than zero and every coefficient explaining lagged values of squared residuals must be bigger or equal to zero (Brooks, 2008). In other words, the ARCH model uses lagged values of squared residuals to predict conditional variance of the model. It should be noted that variance σ_t^2 is sometimes denoted as h_t. To test for the ARCH presence, one should first run the OLS model for the tested variable, then it is necessary to test the ARCH effect. If ARCH effect is present, then one can apply Generalized Autoregressive Conditional Heteroscedastic or GARCH (p,q) model developed by Bollerslev (1986, p.309), and a general view of a GARCH (p, q) model goes as follows:

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i} = \alpha_0 + A(L)\varepsilon_t^2 + B(L)h_t \tag{5}$$

where,
$h_t = \sigma_t$ – fitted variance
α_0 – intercept
$\sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 = A(L)\varepsilon_t^2$ – the sum of q lags of lagged residuals or past volatility
$\sum_{i=1}^{p} \beta_i h_{t-i} = B(L)h_t$= – the sum of p lags of the conditional variance

It is important to note that the ARCH model only uses lags of residuals, whereas GARCH also takes previous values of variance, to model current conditional variance. So the plan is to model volatility using GARCH method and then use it in further stages of the analysis process. The author has to model conditional volatility of S&P 500 Index fund because the ultimate goal is to check whether trading volumes of ETFs affect it. Thus, index price returns data is used for GARCH modelling. The author might include Amihud and Momentum in the GARCH model if there is evidence that they are significant and can be used in volatility modelling. Once conditional volatility is found, it can be used for testing purposes, to study its the relationship with trading volumes of ETFs, and other control variables.

The plan is to use is the VAR model. According to Brooks (2008), VAR models test a simultaneous relationship between multiple variables. Below is the simplest formula example of a VAR model for two variables, for X and Y, and the rest are their lags (Brooks, 2008). For the robustness of model, it is important to use the correct lag order (Brooks, 2008). In addition to VAR, it is planned to use Granger causality test based on this model. As it can be seen, below are two VAR models that represent a system in which both X and Y are used as dependent variables in order to examine how one variable and its lags affect another, to test eventually whether causality in just one or also in the opposite direction exists. Granger causality test developed by Granger (1969) is used to test if variable X Granger-causes changes in a variable Y, or vice versa (Granger, 1969). According to Brooks (2008), if one variable is Granger-caused by another, this does not necessarily mean that changes in first imply direct changes in the second, and so it is rather that variables are related and so some changes get translated into the second variable as a result. The test focuses on short-term changes (*Ibid.*). The null hypothesis is that X does not Granger-causes Y, for this, lag coefficients must equal to zero or β=γ=δ=0 (*Ibid.*). An alternative hypothesis is that lags of X explain Y, in other words, X causes Granger-causes Y. For instance, a bivariate VAR model of lag order m developed by Granger (1969, p.431) goes as follows:

$$\begin{cases} X_t = \sum_{j=1}^{m} \alpha_j X_{t-j} + \sum_{j=1}^{m} b_j Y_{t-j} + \varepsilon_t \\ Y_t = \sum_{j=1}^{m} c_j Y_{t-j} + \sum_{j=1}^{m} d_j X_{t-j} + \eta_t \end{cases} \tag{6}$$

where,
X_t – first variable
Y_t – second variable
α_j, b_j, c_j, d_j – coefficients of lags
ε_t, η_t – residuals

Vector autoregressive models will be applied to both realized variance and conditional volatility. VAR model enables testing for how several independent variables and its lags affect dependent variables. VAR models will include lags of absolute and relative trading volumes of ETFs as independent variables. Also, it will use lags of dependent variables, either realized variance or conditional volatility. Moreover, control variables Momentum and Amihud will be used as exogenous variables to test their effect on the same models. For example, vector autoregressive model with three lags for conditional volatility and absolute trading volume presented as a system will be used for Granger-causality test, and looks as follows:

$$
\begin{cases}
h_t = \sum_{j=1}^{3} \alpha_j h_{t-j} + \sum_{j=1}^{3} b_j ETFV_{t-j} + Mom_t + Amihud_t + \varepsilon_t \\
ETFV_t = \sum_{j=1}^{3} c_j ETFV_{t-j} + \sum_{j=1}^{3} d_j h_{t-j} + Mom_t + Amihud_t + \eta_t
\end{cases}
\tag{6.1}
$$

where,
$\mathrm{h_t}$ – conditional volatility
$\mathrm{h_{t-1}}$ – lags of conditional volatility
$\mathrm{ETFV_t}$ – absolute trading volume of ETFs
$\mathrm{ETFV_{t-j}}$ – lags of absolute trading volume of ETFs
$\mathrm{Mom_t}$ – Momentum
$\mathrm{Amihud_t}$ – Liquidity measure
α_j, b_j, c_j, d_j – coefficients of lags
η_t, ε_t – residuals

Then, F-test for joint significance is applied (Brooks, 2008). Granger causality test states where is the cause and where is the effect, but it does not specify whether there is a true causal relationship. For this, impulse responses can be used as they visualize what happens, namely they provide a graphical representation of where shock originates and how it translates to other variables (*Ibid.*). Also, it shows how many steps it takes for a change to happen.

The author of this thesis uses a vector autoregressive model to answer the research question of whether changes in trading volumes in ETFs affect the volatility of the S&P 500 Index. If the model shows that lags of trading volumes are significant, influence indeed exists. It also implies that ETFs via their trading volumes may cause such critical issue to market stability as excess volatility. If OLS model estimates the influence of variables at a tested level, checks their importance, then VAR also examines the statistical significance of lags. Hence, in the latter case, more lags are involved in explaining the dependent variable, and so if significant, then VAR can be used for prediction purposes. OLS and VAR models will also study the relationship between index volatility and control variables. Amihud represents the liquidity of the S&P 500 Index. It indicates whether an increase in liquidity affects the volatility of the same index. If the change is substantial, then it creates excess volatility. As for momentum, there is a similar goal to check if momentum anomaly, or the strategy of buying winners and selling losers, affects the volatility of the S&P 500 Index fund. A Granger-causality test can show if this relationship works in one direction and goes from trading volumes to volatility or vice versa. It may unveil the presence of a bi-directional cause. It means that initial change happens in one variable and translates to another, but this impulse returns and equilibrium takes new levels accordingly. Presence of Granger causality implies that a causal relationship between tested variables exists. Thus, change in one variable can be used to predict another. If so, this might ultimately put the financial system at risk of collapse.

2.3. Data sample description

Data for the analysis is collected from publicly available sources and some calculation is performed by the author. Data, for the regression analysis and other tests, is gathered from Yahoo Finance and Kenneth R. French online database (French, 2020). Both tests and models are performed in a statistical package Gretl. The data sample uses monthly observations, and it is constructed for the period from August 2010 to December 2019 which comprises 113 observations for a length of almost 10 years. However, it also includes daily data. There was no global financial crisis over this period of time, but the data captures a lot of significant events and financial turmoil of a lesser magnitude that has happened over the past decade. This thesis will focus on studying if changes in variables based on the data of such ETFs like SPDR S&P 500 ETF, iShares S&P 500 ETF, and Vanguard S&P 500 ETF affect the volatility of the underlying S&P 500 Index. For the S&P 500 Index data reflection, S&P real time price data will be used as it is a proxy or non-tradable index tracking movements in the S&P 500 Index. To remind, mentioned indexes are known under ticker symbols SPY, IVV, VOO and ^GSPC, correspondingly. It is also important to note that all the data is denominated in USD currency. For the regression analysis information about monthly prices and trading volumes for SPY, IVV, VOO and SPX was collected. Variables carrying values of closing prices for SPY, IVV, VOO and SPX are called PSPY and PIVV, PVOO and PSPX, respectively. Then, the author has calculated the dollar trading volume. Dollar trading volume for an ETF is created by multiplying trading volume for an index by closing prices for this index. Variables representing dollar trading volumes were named in the following, intuitive way: VolSPY, VolIVV, VolVOO and VolSPX. Data about prices and trading volumes are taken from Yahoo Finance. Prices and trading volumes data can be used for construction for other variables. Created ETFV variable represents a sum of trading volumes of the three biggest ETFs tracking S&P 500 Index.

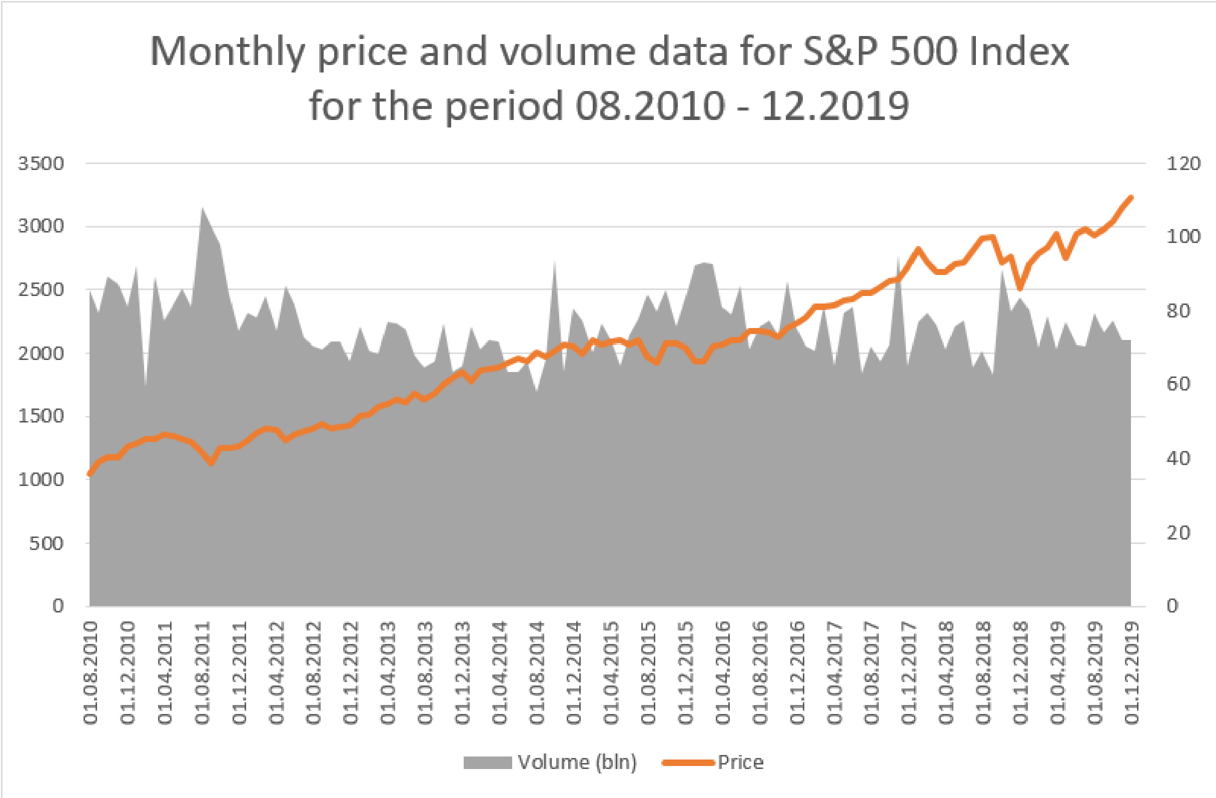

Figure 1. Monthly price and volume data for the S&P 500 Index for the period from 08.2010 to 12.2019
Source: author's calculations

$$ETFV = VolSPY + VolIVV + VolVOO \tag{7}$$

Another variable, RETFV, is the relation of ETFV to a trading volume of S&P 500 Index. It should be mentioned that since values of ETFV are huge they are scaled down or divided by a million, whereas values of RETFV were scaled up or multiplied by a million.

$$RETFV = \frac{ETFV}{VolSPX} \tag{8}$$

Along with main variables that represent a trading volume of ETFs, additional control variables will be used, namely liquidity and momentum. As for another variable, momentum, Wang and Xu (2015) find evidence that the market volatility can predict momentum. On the other hand, Xu and Yin (2017) examined a reverse relationship and have found that momentum has insignificant influence on market volatility. Yet it is planned to include momentum in empirical analysis for the testing purposes. Momentum comes from the research of Jegadeesh and Titman (1993). They have tested an idea of momentum investing which assumes that stocks that had high returns over the past month, usually three to twelve months are expected to keep doing go so it is worth buying such stocks, and at the same time stocks that did badly in the past are expected to be not that great investments so investors should sell them. The idea is very similar to hot hand fallacy, a cognitive bias studied by (Gilovich et al., 1985). In the study of Jegadeesh and Titman (1993), they have identified that momentum investing can, in fact, be the profitable strategy at least for a short period of time. Data for the momentum variable was taken from Kenneth R. French online database. For the momentum factor construction French (2020) uses the following formula:

$$Mom = \frac{1}{2}(Small\ High + Big\ High) - \frac{1}{2}(Small\ Low + Big\ Low) \tag{9}$$

What this formula (8) does, it takes "the average return on the two high prior return portfolios minus the average return on the two low prior return portfolios" (French. 2020). The data is gathered for 6 stocks (*Ibid.*). These stocks must be median by size, and this is identified by taking the median of NYSE stocks' market equity (*Ibid.*). Also, stock returns must be in the 30th percentiles of NYSE for badly performing or losers and in the 70th percentile of NYSE for well performing or winners (*Ibid.*). Table 1 and Table 2 demonstrate descriptive statistics for monthly and daily data, correspondingly.

Table 1: Descriptive statistics for monthly data

	Mean	Median	Minimum	Maximum	S.D.
ETFV	$4.37*10^5$	$3.99*10^5$	$2.66*10^5$	$9.05*10^5$	$1.26*10^5$
RETFV	3000.50	2963.1	1540.1	7029.8	884.12
dETFV	-313.05	9534.2	$-4.85*10^5$	$4.03*10^5$	$1.41*10^5$
dRETF	21.76	41.15	-3032.2	1938.8	706.24
RSPX	0.01	0.01	-0.10	0.10	0.03
Momentum	0.28	0.27	-8.65	10.29	3.18
Amihud	9.42	8.07	2.83	49.92	6.56
RV	16003	13840	5313.7	60777	8371.2

Source: author's calculations

Table 1 has ETFV values in its original representation, whereas RETFV are presented in adjusted form, namely they are scaled down by a million.

The liquidity parameter or Amihud ratio, developed by Amihud (2002), is another variable that is included in the model. Amihud ratio represents a measure of liquidity of the chosen stock. This ratio was also applied by Ben-David et al. (2018) and Xu and Yin (2017) in their articles. In order to create this variable, the following formula based on Amihud (2002) original work was applied using daily and monthly data depending on the testing needs:

$$Amihud\ ratio = ILLIQ = \frac{1}{N}\sum_{t=q}^{T} \frac{|rt_t\ |}{\$Vol_t} * 10^{16} \tag{10}$$

Absolute return for the formula above was computed by taking an absolute value from the price return for the S&P 500 Index. Equation (10) shows Amihud formula which takes the ratio of the sum of price return to the dollar trading volume and then multiplies the result by ten to the power of sixteen to make result visible and aligned with other variables. For the daily data, variables were summed day by day. As for monthly values, monthly return and the sum of all dollar trading volumes per month are used for variable construction. To clarify, price returns for indexes are calculated using such formula that uses adjusted closing prices, meaning dividends are reinvested and accounted for:

$$Total\ Return = r_t = \ln(\frac{P_t}{P_{(t-1)}}) \tag{11}$$

Table 2: Descriptive statistics for daily data

	Mean	Median	Minimum	Maximum	S.D.
ETFV	20814	18543	5854.9	97589	9000.9
RETFV	3022.5	2769.5	905.39	25707	1427.0
dETFV	0.95	-98.66	-51222	46111	7424.2
dRETF	-0.69	-22.69	-14747	12691	1051.6
RSPX	0.0005	0.001	-0.07	0.05	0.01
Momentum	0.02	0.05	-3.81	3.63	0.70
Amihud	9.47	6.32	0	235.55	12.07
h	82.97	52.50	19.45	1439.9	99.22
RV	15515	3289.2	0	$8.97*10^5$	40427

Source: author’s calculations

Since the plan is to test the volatility of the S&P 500 Index, the next logical step is to construct a variable representing realized variance, RV. For this the square root of the aggregate monthly return is used, latter constructed as a sum of squared differences of daily log-returns. This is how monthly RV is constructed. For daily realized variance squared difference of log-returns is used. Values of variance were scaled up by a million. And one more variable that is needed for modelling conditional volatility using the GARCH model is RSPX. RSPX is the daily log returns of the S&P 500 Index.

All in all, in this thesis hypothesis of whether changes in trading volumes of ETFs affect the underlying index, the S&P 500 Index, will be tested. A rationale behind is that an increase in trading volumes of ETFs represents a growing interest towards ETFs. Thus, more people join this trend by pouring more funds into Exchange-Traded Funds without a proper understanding of their purchases. Consequently, it intensifies trading volumes. If there is a piece of evidence that ETFs impact the volatility of an underlying index, then markets are inefficient. Hence, the whole financial system is endangered. Volatility effect further implies that some other discussed issues regarding ETFs might be true as well and should be examined more thoroughly, Since ETFs, in this case, are a potential threat to financial stability.

3. EMPIRICAL RESULTS

3.1. Monthly data models

Testing phase requires testing all the variables for stationarity. For this, the ADF test for stationarity is applied on series with constant and without a trend. ADF test results of absolute and relative trading volumes of ETFs are shown in Appendix 1. It appears that all variables have no unit root and are stationary at its levels with results significance of at least 5%. However, some series, like realized variance, are stationary because differencing is applied for their construction. In further testing, the first difference of the variables, ETFV and RETFV, will be used with the differenced variables to be shown as dETFV and dRETFV, correspondingly. The differencing will be used to account for the possible presence of multicollinearity between the trading volume of ETFs and the volatility of the S&P Index. As a result of the ADF test, the author has found that every variable is stationary and that the OLS model can be applied. Specifically, the OLS model for realized variance, RV, as a dependent variable will be used. Control variables, Amihud and Momentum, can be added to test a linear relationship of RV and such independent variables as trading volumes.

Table 3: OLS model estimates for monthly RV excluding control variables

	RV_t				dRV_t	
Constant	-8629.55 *** (1611.23)	3329.08 *** (796.67)	-6970.73 *** (1734.55)	2794.01 *** (875.69)	129.02 (738.58)	224.65 (736.94)
$ETFV_t$	0.05 *** (0.004)					
$dETFV_t$		0.05 *** (0.003)			-0.02 *** (0.005)	
$RETFV_t$			6.98 *** (0.61)			
$dRETFV_t$				10.07 *** (0.59)		-4.004 *** (1.05)
RV_{t-1}	0.17 *** (0.06)	0.80 *** (0.05)	0.13 * (0.06)	0.82 *** (0.05)		
dRV_{t-1}					-0.36 *** (0.08)	-0.36 *** (0.08)
R^2/Adj.R^2	0.70/ 0.69	0.82/ 0.81	0.64/ 0.63	0.78/ 0.78	0.25/ 0.24	0.26/ 0.24
R^2/Adj.R^2 lag excluded	0.70/ 0.67	0.28/ 0.29	0.63/ 0.62	0.24/ 0.24	0.13/ 0.12	0.13/ 0.13

Source: author's calculations

Table 3 demonstrates the output for linear regression model iterations. In Table 3 one can see tests for both realized variance and the first difference of realized variance. Independent variables include either ETFV or RETFV, and the second version includes first differences of mentioned variables. For both ETFV and RETFV, variables' first lag is included in the models testing relationship with a realized variance. It should be noted that tests for a relationship of the first difference of realized variance and both ETFV and RETFV are not reflected in the table because variables in these models were statistically insignificant. Table 4 shows the output for the same models except for it includes control variables, Amihud representing liquidity and Mom representing momentum. As can be seen, in Table 3 all models with levels data have significant trading volumes indicating that they affect realized variance. This result is consistent with the study of Xu and Yin (2017). It should also be noted that the first difference of realized variance entails different result, but R-squared for these models is low, so should not be trusted.

Table 4: OLS model estimates for monthly RV with control variables

	RV_t				dRV_t	
Constant	-6765.04 *** (715.62)	2924.76 *** (596.41)	4085.57 * (2457.45)	2812.71 *** (748.43)	-3343.35 ** (1321.43)	-3604.33 *** (1315.19)
$ETFV_t$	0.04 *** (0.002)					
$dETFV_t$		0.04*** (0.003)			-0.03 *** (0.005)	
$RETFV_t$			0.41 (1.26)			
$dRETFV_t$				6.68 *** (0.72)		-5.43 *** (1.09)
$Amihud_t$	758.15 *** (35.32)	530.78 *** (56.66)	975.16 *** (168.70)	509.35 *** (77.42)	364.49 *** (116.98)	403.75 *** (117.65)
Mom_t	-53.06 (61.37)	57.37 (81.53)	38.50 (136.01)	-7.71 (100.84)	-42.95 (225.72)	2.35 (222.71)
RV_{t-1}	-0.01 (0.03)	0.51 *** (0.05)	0.09 (0.06)	0.52 *** (0.06)		
dRV_{t-1}					-0.38 *** (0.08)	-0.38 *** (0.08)
R^2/Adj.R^2	0.94/ 0.94	0.90/ 0.90	0.72/ 0.71	0.85/ 0.84	0.32/ 0.29	0.33/ 0.31
R^2/Adj.R^2 lag excluded	0.70/ 0.69	0.78/ 0.78	0.72/ 0.71	0.75/ 0.74	0.18/ 0.16	0.19/ 0.17

Source: author's calculations

In Table 3 and Table 4, every column represents an output for a model with different specification and an estimated value of coefficients, with the values in the brackets representing estimated errors. Stars represent models' significance, namely *** means 1% significance of a variable, ** means 5% and * means 10% significance level.

The conclusions that can be made based on the regression output in Table 4 are that momentum is statistically insignificant, and thus has no impact on realized variance. Lags are only significant for the first differences of dependent variables. On the other hand, lags in Table 3 are statistically significant in every model reflected in the table. Liquidity, in its turn, is significant in every case shown in Table 4. Thus, it is true that liquidity affects the realized variance of the S&P 500 Index. Amihud represents liquidity, but it also reflects an interest of investors towards an asset. It is due to assets being more liquid when traded more often. Thus, a positive outcome for the Amihud ratio implies an increase in volatility of the index, similarly to trading volumes. This result is similar to the study of Xu and Yin (2017) and Ben-David et al. (2018). In both Table 3 and Table 4, an increase in either absolute or relative trading volumes leads to positive changes in realized variance. Consequently, it contradicts the hypothesis and thus it can be concluded that when testing models both with and without control variables, there is an indication that in some instances trading volumes of ETFs may affect the volatility of SPX. It is due to ETFV, dETFV and dRETFV demonstrating their significance in the regressions and because tested models have a good explanatory power due to high R-squared. Table 4 has low R-squared for differenced RV just as Table 3, so the result is not trustworthy. Based on the output of at level realized variance as a dependent variable, the author concludes that the tested hypothesis should be rejected.

3.2. Daily data models

Daily data includes all the same basic variables as monthly data. The logic of variables construction is the same for daily data as it was for monthly. And as a first step, all the variables were tested for the stationarity, and ADF test approved an absence of unit roots in the data. In this section, the OLS model will be used for daily volatility, and both realized variance and conditional volatility will be utilised for this. The return of the S&P 500 Index is checked for the presence of ARCH and GARCH effect. It is performed to model conditional volatility and then to use this variable further in the forthcoming tests. After those tests, the VAR model will be estimated for both conditional and unconditional volatilities. The reason why VAR models are not constructed for monthly data is that conditional volatility cannot be modelled because there is no evidence for ARCH and GARCH effects in returns data for the S&P 500 Index.

Firstly, the author has used daily index return data to model conditional volatility. For this, return data of the S&P 500 Index is regressed using the OLS model. This model appears to be significant and so it is tested for the presence of an ARCH effect. The test for an ARCH effect states that the effect does exist. Since there is evidence for an ARCH effect, it was then reasonable to test this variable for the presence of a GARCH effect. Moreover, an option to include control variables for the modelling of variance was considered as well, but it was decided to continue with the basic model and include control variables later to VAR model as exogenous variables. And so tests were continued only with the returns data for the S&P 500 Index. After performing several tests for GARCH with different specifications of p and q parameters, it was decided to continue with GARCH(1,1) model because it shows statistical significance at 5% level of all the parameters and it has lower Akaike criterion than other models with significant parameters. An output for GARCH(1,1) from Gretl can be found in Appendix 2. The same approach is applied by Xu and Yin (2017). Their study has also found evidence for the significance of GARCH effect.

As a next step, conditional volatility variable is created using saved values from GARCH(1,1) based on daily frequencies. Using a new variable, other models could be constructed to test how independent variables affect conditional volatility. Then, the OLS model utilised to test a linear relationship between the ETFs tracking S&P 500 Index, precisely ETFs trading volumes and the volatility of SPX. Table 5 and Table 6 present model outputs for mentioned dependent variable and for absolute and relative trading volumes both at levels and at first-differences, also Table6 includes control variables Momentum and Amihud, whereas Table 5 excludes them. As known, conditional volatility is positive by construction, but Table 5 and Table 6 demonstrate negative constant. However, from the descriptive statistics, we can see that conditional variance reaches a value of 19.45 at its minimum. Thus, adjusting for error, the negative constant should not be a problem since volatility never breaches limits and does not get into a negative domain.

The OLS model for conditional volatility demonstrates that trading volumes of ETFs both in absolute and relative terms do affect the volatility of the S&P 500 Index. The output of a model without control variables, shown in Table 5, demonstrates the significance of trading volume variables at 1% level which is reasonably high. In addition to this, all lags of GARCH(1,1) are significant as well and the model output is, in general, reliable because R-squared is considerable. As known, this parameter defines a model's ability to explain changes in a dependent variable. It is also crucial that the model is not overloaded with additional variables and has such a high R-squared. Table 6, shown below, demonstrates an output for the same dependent and independent variables and also it examines similar iterations of OLS models except for it includes control variables.

Table 5: OLS model estimates for monthly GARCH(1,1) without control variables

	h_t			
Constant	-19.16 *** (1.74)	5.81 *** (0.94)	-16.43 *** (1.62)	5.89 *** (0.94)
$ETFV_t$	0.001 *** ($8.30*10^{-5}$)			
$dETFV_t$		-0.001 *** ($9.77*10^{-5}$)		
$RETFV_t$			0.01 *** (0.001)	
$dRETFV_t$				-0.01 *** (0.001)
h_{t-1}	0.89 *** (0.08)	0.93 *** (0.01)	0.88 *** (0.01)	0.93 *** (0.01)
R^2/Adj.R^2	0.89/ 0.89	0.87/ 0.87	0.89/ 0.89	0.88/ 0.88

Source: author's calculations

When testing models that include control variables, Amihud and Momentum, one can find that control variable Amihud is insignificant only in one case. In the remaining cases, both Amihud and Momentum are statistically significant just as variables representing trading volumes. It indicates that onboarding momentum-based trading strategy may affect the volatility of SPX. In addition to this, models from Table 6 also have high values of R-squared. Nevertheless, both Table 5 and Table 6 demonstrate that differenced variables of trading volumes have negative values. Thus, changes in differenced data imply a decrease in volatility. However, absolute and relative trading volumes at levels have a positive result. Positive changes in trading volumes may entail positive investment activity. Hence, a decrease in volatility reflects rising confidence of investors in the current market conditions.

However, the model's explanatory power measured by R-squared remains on the same level when the model is tested with and without control variables. It means that such main independent variables as trading volumes of ETFs and lags of the dependent variable are the ones that affect the dependent variable. Here, again negative constant is not critical, because conditional volatility does not breach any limits.

Table 6: OLS model estimates for monthly GARCH(1,1) with control variables

	h_t			
Constant	-19.30 *** (1.73)	3.10 *** (0.996)	-20.58 *** (1.74)	2.59 *** (0.99)
$ETFV_t$	0.001 *** ($8.6*10^{-5}$)			
$dETFV_t$		-0.001 *** ($9.998*10^{-5}$)		
$RETFV_t$			0.01 *** (0.0007)	
$dRETFV_t$				-0.01 *** (0.001)
$Amihud_t$	0.10 (0.06)	0.50 *** (0.07)	-0.49 *** (0.08)	0.63 *** (0.07)
Mom_t	3.69 *** (0.98)	2.67 *** (2.02)	2.91 *** (0.98)	2.50 ** (1.01)
h_{t-1}	0.88 *** (0.01)	0.91 *** (0.01)	0.89 *** (0.01)	0.90 *** (0.01)
R^2/Adj.R^2	0.89/ 0.89	0.88/ 0.88	0.89/ 0.89	0.88/ 0.88

Source: author's calculations

As the following step, it is reasonable to consider vector autoregression models. Having volatility data as either conditional volatility or realized variance VAR model can be constructed. In this VAR, volatility can be used as a dependent variable, while the trading volume, either absolute or relative, and lags as independent variables. Additionally, control variables will be used to test whether they affect volatility. Also, the VAR model will be exercised for levels data and its first differences. So the next step is to build VAR models, but before this, it is crucial to scale up volatility by a million as this step should enable the better visual representation of the result. And the use of Gretl's functional helps to check the significance of different lags for the VAR model, the lowest statistical significant lag, in this case, is three. Hence, it is justified to construct a VAR(3) model or model with three lags for every independent variable. Below are Table 7 and Table 8 that represents an output for VAR models.

Vector autoregression model with three lags provides another evidence that trading volumes of ETFs affect the volatility of the S&P 500 Index. In Table 7 all but one lags of trading volume are significant. Therefore they help to explain the dependent variable, in our case, index volatility. As can be seen, momentum has no statistical significance in the VAR model for volatility and RETFV variable. However, the momentum variable is significant for a model with absolute trading volume.

Thus, there is a possibility that momentum effects may add up value in explaining the volatility of the S&P 500 Index. In vector autoregressive models, Liquidity or Amihud ratio or liquidity demonstrate higher significance levels compared to momentum. However, models without extra variables are meaningful as well. It means that zero hypotheses of my thesis can be rejected. It also means that some influence exists, since ETFs may cause excess volatility of the S&P 500 Index.

Table 7: VAR model for GARCH(1,1) with and without control variables

	Excluding Control Variables		Including Control Variables	
	X=ETFV	X=RETFV	X=ETFV	X=RETFV
	GARCH(1,1)			
Constant	-20.96 *** (2.02)	-18.13 *** (1.79)	-22.57 *** (2.04)	-18.55 *** (1.81)
h_{t-1}	0.80 *** (0.02)	0.80 *** (0.02)	0.78 *** (0.02)	0.81 *** (0.02)
h_{t-2}	0.22 *** (0.03)	0.23 *** (0.03)	0.22 *** (0.03)	0.22 *** (0.03)
h_{t-3}	-0.14 *** (0.02)	-0.16 *** (0.02)	-0.14 *** (0.02)	-0.16 *** (0.02)
X_{t-1}	0.002 *** ($9.58*10^5$)	0.02 *** (0.001)	0.002 *** ($9.56*10^5$)	0.02 *** (0.001)
X_{t-2}	-0.001 *** (0.0001)	-0.004 *** (0.001)	-0.001 *** (0.0001)	-0.005 *** (0.001)
X_{t-3}	-0.0003 *** (0.0001)	-0.003 *** (0.001)	-0.0003 *** (0.0001)	-0.003 (0.001)
Amihud			0.28 *** (0.06)	-0.14 ** (0.06)
Momentum			2.02 ** (0.90)	-1.47 (0.90)

Source: author's calculations

When testing the vector autoregressive model with three lags or VAR(3) for realized variance, presented in Table 8, it provides a mixed result. It does not strongly support the hypothesis, because it has the first lag of trading volume variables which is statistically significant in all cases both with and without control variables. There is consistency regarding Momentum variable as it remains insignificant at the 10% level. Perhaps conditional variance is superior for the vector autoregression model, yet in both cases, there is some indication that trading volume does affect the volatility of the S&P 500 Index.

Table 8: VAR model for RV with and without control variables

	Excluding Control Variables		Including Control Variables	
	X=ETFV	X=RETFV	X=ETFV	X=RETFV
	RV			
Constant	−9996.44 *** (2480.90)	-9423.11 *** (2207.93)	−28894.5 *** (1513.78)	2791.69 ** (1348.32)
RV_{t-1}	0.08 ** (0.02)	0.05 ** (0.02)	−0.02 (0.014)	0.09 *** (0.014)
RV_{t-2}	0.25 *** (0.02)	0.24 *** (0.02)	0.05 *** (0.01)	0.16 *** (0.01)
RV_{t-3}	0.08 *** (0.02)	0.10 *** (0.02)	−0.04 *** (0.01)	0.058 *** 0.014
X_{t-1}	0.56 *** (0.13)	5.90 *** (0.86)	0.49 *** (0.08)	-2.48 *** (0.54)
X_{t-2}	0.08 (0.14)	-0.01 (0.97)	0.19 ** (0.08)	-0.36 (0.58)
X_{t-3}	0.28 ** (0.12)	0.35 (0.9)	0.28 *** (0.07)	-3.26 *** (0.51)
Amihud			2599.88 *** (40.05)	2791.44 *** (43.57)
Momentum			−1088.02 * (635.94)	-451.36 (638.85)

Source: author's calculations

Lastly, a conditional variance is used for formal Granger causality test between volatility and trading volume. To be precise, both absolute and relative trading volumes are used, and the data is tested in all possible variations for levels and first differences. The only indication of potential causality is found between the first differences of volatility and absolute trading volume. For this, the outcome is shown in Table 9. It is the only statistically significant case that states that volatility may affect trading volume. All the remaining test result indicate that there is no Granger causality between tested variables.

Table 9. Granger causality test for conditional volatility and trading volume

Null hypothesis	Probability	
	Without control variable	With control variables
d_h does not GC ETFV	0.0867	0.0044

Source: author's calculations

All in all, the author has ensured that all models are stationary. For example, vector autoregressive models' inverse roots lay within the unit circle. The results are mixed. Monthly data demonstrates that roughly half of the models is significant. Daily data shows that the vast majority of the models constructed for conditional volatility as a dependent variable are meaningful and disprove the hypothesis. The realized variance, on the other hand, essential in OLS but insignificant for VAR models. Ultimately, a change in trading volumes of ETFs may imply further changes in the volatility of the underlying index. In this thesis, results are representative for the S&P 500 Index.

CONCLUSION

In this Master's thesis author has presented a broader overview of such investment products as Exchange Traded Funds, examined its mechanics in more detail. Also, considered the biggest issues that ETFs can potentially lead to. In addition to this thorough regression analysis with the use of different tests and models is performed to test the hypothesis that trading volumes of ETFs do affect the volatility of underlying assets.

It was then discussed in what environment ETFs exist, namely, an overview of two investment styles was presented, active and passive, then the author has compared and contrasted them to better understand to what type ETFs refer and why, but also why they are so popular among investors. It was mentioned that ETFs are cheaper, they suit for passive investors and generally for people who believe that markets are efficient and so no one can outperform them.

Thereafter, it was considered what issues ETFs may cause or at least increase a probability of their appearance. Among those issues, there are risks of a flight to liquidity that may lead to other critical problems like asset value drop, increase in volatility and as a result of all the changes normal operation flow of financial markets may collapse. Another considered issue is a redemption risk that mainly refers to a priority of some investors over the other, this may entail such risks like insolvency for funds. And it is also known that APs are in a position of power and so may dictate own conditions for assets sale and purchase to fund managers. There are mechanisms that should control fair conditions for everyone but the probability of those risks materialisation still exists.

Furthermore, there is a risk of increased comovement. In other words, assets related to ETFs may start moving in tandem in the same direction due to various reasons. Inclusion effect may add up to this due to an increase in analyst coverage and rising turnover. ETFs might also increase the volatility of underlying assets, which is again critical because it may lead to financial markets turmoil. According to industry practitioners, one more possible issues is an increased concentration of passive investments, mostly caused by ETFs popularity. It is so, because exchange-traded funds focus on stocks with large capitalisation and tremendous trading volumes, whereas small-cap stocks remain out of sight. Hence, this may create bubbles in financial markets. The SEC tries to study all the potential threats to financial stability and develop measures that prevent wrongdoing and minimise risks.

As for ETFs as a standalone product, they were also examined in this thesis. It is known that ETFs normally refer to passive investments but some mixed type ETFs may exist as well. They share some similarities with both open-end and closed-end mutual funds. However, the main difference lays within the shares creation and redemption mechanism. In addition to this, along with plain vanilla ETFs, there are more complex products like Exchange Traded Notes tracking debt instruments and many more. What is so tricky about ETFs though, is that there has already been financial markets disturbance, namely Flash Crash, which is said to be related to ETFs. That is why it is so important to study these investment products to prevent a reoccurrence of similar events and the ones with more dramatic consequences.

The second part of this thesis is devoted to empirical analysis. This Master's thesis is written based on two academic articles – Ben-David et al. (2018) and Xu and Yin (2017). The author has examined a relationship between either conditional or unconditional volatility on the one hand and either absolute or relative trading volume on the other hand. Data used at levels and first differences. Also, different models were built for the data with daily and monthly frequencies. All in all, the author has tested OLS and VAR models, and Granger causality test. The results have found evidence that trading volumes affect conditional volatility when tested on daily data. At the same time, both daily and monthly data for unconditional variance partially disprove the hypothesis. As for the Granger causality then in majority cases, no causality is found, and so it is unlikely that trading volumes can be used for volatility predictions.

Hence, the results are mixed. There is an indication that changes in trading volumes affect changes in volatility, in case of this thesis concerning S&P 500 Index, but used data also indicates that it is unlikely that trading volume Granger causes volatility. VAR models have statistically significant lags but formal Granger causality tests were only supporting one model for conditional volatility and absolute trading volumes of ETFs with control variables. And there is still a lot of space for research and too many open questions to say if this can be used for chasing abnormal returns. For example, intraday data could be examined as well. Then, it is important to understand how much time it takes for volatility to react to changes in trading volumes. Also, the relationship between volatility and other variables could be studied as well.

KOKKUVÕTE

BÖRSIL KAUBELDAVATE INDEKSFONDIDE MÕJU ALUSVARADE VOLATIILSUSELE INDEKSI S&P 500 NÄITEL

Veronika Matrossova

Aastate jooksul on finantsturud tutvustanud paljusid uusi instrumente ja tooteid, igaüks oma eripäraste omadustega. Üks selline hiljutine finantstoode on börsil kaubeldavad indeksfondid ehk ETF-id, mille loojaks on John Bogle. Uudne finantsinstrument loodi eelmisel sajandil, kuid eriti populaarseks sai viimase dekaadi jooksul. USAs on aastaks 2020 ETF-i tööstuse hallatavate varade maht jõudnud 5 triljoni dollarini, mis on tohutult palju, kui tuua võrdluseks, et investeerimisfondide hallatavate varade kogumaht on 21 triljonit dollarit (Pisani, 2020). ETF-id esindavad passiivsete investeerimistoodete valikut, kuna nad jälgivad juba olemasolevaid indekseid ja muid finantsinstrumente. Üldiselt on suundumus aktiivselt passiivsele investeerimise stiilile üleminek. Seliseks muutuseks võib olla mitu alust ja põhjust. Sarnaselt 2008. aastal toimunud finantskriisile ei pööra ka nüüd investorid tähelepanu finantstoodete keerukusele ning selle hinda mõjutavatele põhjuslikele seostele jätkates suures mahus ETF-idesse raha paigutamist. Seega, ETF-idega kauplemine on selle lõputöö uurimisobjekt, kuna võib väita, et see on üks üleliigse volatiilsuse peamisi põhjuseid ja mis võib kahjustada turu tõhusust. Seetõttu on selle teema uurimine endiselt väga asjakohane, sest võib aidata mõista finantsturgude stabiilsuse võimalikke nõrku kohti ja ohte.

Käesolevas magistritöös on autor esitanud laiema ülevaate börsil kaubeldavatest indeksfondidest, uurinud selle mehaanikat üksikasjalikumalt, käsitlenud ka suurimaid probleeme, mida ETF-id potentsiaalselt kaasa võivad tuua. Koos sellega on autor esitanud teoreetilise ülevaate, milles võrreldakse ja vastandatakse aktiivseid ja passiivseid investeeringuid, ning arutletakse passiivsete investeerimistoodete, sealhulgas ETF-ide, praeguste probleemide üle. Viimasena käsitletakse lühidalt ETFide rolli Flash Crash-is 2010. aastal. Käesolevas magistritöös viidi läbi põhjalik regressioonanalüüs, kasutades mitmeid teste ja mudeleid ning testiti püstitatud hüpoteesi, et kas

ETF-ide kauplemise mahud mõjutavad volatiilsust ja põhjustavad seega S&P 500 indeksi üleliigset volatiilsust. Seega keskendub see lõputöö peamiselt volatiilsusele, mis võib olla põhjustatud kauplemisest ETF-idega.

Magistritöö põhieesmärk oli uurida börsil kaubeldavate indeksfondide kauplemismahtude ja S&P 500 indeksi liigse volatiilsuse vahelist seost. See on suure tähtsusega teema. Seda seetõttu, et ETF-e pole teiste finantstoodetega, näiteks optsioonide või futuuridega võrreldes veel põjalikult uuritud. Pealegi on nende populaarsus kasvamas ja nad pole veel osalenud tõsises finantskriisis. Veelgi enam, varasemad sündmused, nagu näiteks Flash Crash, on põhjustanud täiendavat muret selle üle, kas passiivsed investeeringud ning eriti ETF-id võivad põhjustada probleeme või suurendada riske, mille tulemuseks on finantskriis. Need on põhjused, miks tuleks ETF-e ja selle komponente uurida. Seoste ja finantstoote käitumise tundmine võib aidata koostada juhiseid ja piiranguid finantsturgudel raskete tagajärgede ennetamiseks. Töö uurimisülesanded hõlmavad asjakohase teoreetilise kirjanduse ülevaadet ja ülevaadet empiirilistest uuringutest. Lisaks sellele ülesanded hõlmavad andmevalimi kogumist empiirilise hindamise eesmärgil. Seejärel hüpoteeside testimiseks teostatakse regressioonanalüüs ja hinnatakse ökonomeetrilisi mudeleid.

Käesoleva magistritöö empiirilise uuringu läbiviimisel on peamiselt tuginetud kahele akadeemilisele artiklile – Ben-David et al. (2018) ning Xu ja Yin (2017). Hüpoteeside testimiseks kasutatakse regressioonanalüüsi, et testida ETF-idega kauplemise mahu ja aluseks oleva S&P 500 indeksi volatiilsuse lineaarset suhet. Autor on uurinud suhet ühelt poolt kas tingimusliku või tingimusteta volatiilsuse ja teiselt poolt kas absoluutse või suhtelise kauplemismahu vahel. Andmed, mida kasutatakse on nii tasemete kui ka diferentside kujul. Mudelid on ehitatud igapäevaste ja igakuiste andmete sagedustega. Hüpoteesi testimisel rakendatakse OLS-mudelit, et testida lineaarset seost ning VAR-i, et testida samaaegselt mitme muutuja vahelist seost. Hinnatakse mudeleid, mis on konstrueeritud nii tingimusliku kui ka tingimusteta volatiilsuse jaoks. Samuti kasutatakse GARCH mudelit tingimusliku volatiilsuse modelleerimiseks ja Granger-põhjuslikkuse testi muutujate põhjuslikkuse uurimiseks.

Magistritöö tulemused on mitmeti tõlgendatavad. Näiteks leiti empiirilist kinnitust sellele, et börsil kaubeldavate indeksfondi kauplemismahu muutused mõjutavad volatiilsuse muutusi S&P 500 indeksis, kuid kasutatud andmed näitavad ka, et on ebatõenäoline, et kauplemismaht Granger-põhjustab volatiilsust. VAR-mudelitel on statistiliselt olulised viiteajad, kuid formaalsedGrangeri-põhjuslikkuse testid toetasid ainult üht mudelit, nimelt tingimusliku volatiilsuse ja ETF-ide absoluutse kauplemismahu mudeli koos kontrollmuutujatega. Edasiseks uurimistööks on aga veel palju ruumi ja liiga palju lahtisi küsimusi, sh kas töö tulemusi saaks kasutada finantsturgudel lisatootluse teenimiseks. Lisaks võiks uurida ka päevasiseseid andmeid. Seejärel on oluline mõista, kui palju aega vajab volatiilsus, et reageerida börsil kaubeldavate indeksfondi kauplemismahu muutustele. Samuti võiks uurida suhet volatiilsuse ja teiste muutujate vahel.

LIST OF REFERENCES

Anadu, K., Kruttli, M., McCabe, P., Osambela, E. and Shin, C., 2018. The Shift from Active to Passive Investing: Potential Risks to Financial Stability?. Finance and Economics Discussion Series, 2018-060. Washington: Board of Governors of the Federal Reserve System.

Arora, R., Betermier, S., Leblanc, G., Palumbo, A. and Shotlander, R., 2020. Concentration In The Market Of Authorized Participants Of US Fixed-Income Exchange-Traded Funds. [online] Bank of Canada. Available at: <https://www.bankofcanada.ca/2020/11/staff-analytical-note-2020-27/> [Accessed 23 December 2020].

Bali, T., Demirtas, K. and Levy, H., 2009. Is There an Intertemporal Relation between Downside Risk and Expected Returns?. The Journal of Financial and Quantitative Analysis, 44(4), pp.883-909.

Bank of England, 2019. Financial Stability Report July 2019. [pdf] Bank of England. Available at: <https://www.bankofengland.co.uk/-/media/boe/files/financial-stability-report/2019/july-2019.pdf> [Accessed 24 November 2020].

Barber, B. and Odean, T., 2001. Boys will be Boys: Gender, Overconfidence, and Common Stock Investment. The Quarterly Journal of Economics, 116(1), pp. 261–292.

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

BlackRock, 2017a. Viewpoint: Index Investing Supports Vibrant Capital Markets. [pdf] Available at: <https://www.blackrock.com/corporate/literature/whitepaper/viewpoint-index-investing-supports-vibrant-capital-markets-oct-2017.pdf> [Accessed 23 November 2020].

BlackRock, 2017b. Index Investing And Common Ownership Theories. [online] BlackRock. Available at: <https://www.blackrock.com/corporate/literature/whitepaper/viewpoint-index-investing-and-common-ownership-theories-eng-march.pdf> [Accessed 23 December 2020].

BlackRock, 2020. Ishares INVESTIGATES: ETF MARKET REALITIES. [pdf] BlackRock. Available at: <https://www.blackrock.com/corporate/literature/investor-education/ishares-investigates-ten-common-questions-about-etfs-en-us.pdf> [Accessed 25 November 2020].

Bogle, J., 2018. Bogle Sounds A Warning On Index Funds. [online] WSJ. Available at: <https://www.wsj.com/articles/bogle-sounds-a-warning-on-index-funds-1543504551> [Accessed 25 November 2020].

Bohn, H. and Tesar, L., 1996. U.S. Equity Investment in Foreign Markets: Portfolio Rebalancing or Return Chasing? The American Economic Review, 86(2), pp. 77-81.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31(3), pp.307-327.

Brooks, C., 2008. Introductory Econometrics For Finance. 2nd ed. New York: Cambridge University Press.

Claessens, S. and Yafeh, Y., 2012. Comovement of Newly Added Stocks with National Market Indices: Evidence from Around the World. Review of Finance, 17(1), pp. 203-227.

Cremers, K. and Petajisto, A., 2009. How Active Is Your Fund Manager? A New Measure That Predicts Performance. Review of Financial Studies, 22(9), pp.3329-3365.

Da, Z. and Shive, S.,2017. Exchange traded funds and asset return correlations. European Financial Management, 24 (1), pp. 1-33.

Dickey, D. and Fuller, W., 1979. Distribution of the Estimators for Autoregressive Time Series With a Unit Root. Journal of the American Statistical Association, 74(366), pp.427-431.

Engle, R., 1982. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. Econometrica, 50(4), p.987.

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

FCA, 2019. Update On The LF Woodford Equity Income Fund. [online] Available at: <https://www.fca.org.uk/news/news-stories/update-lf-woodford-equity-income-fund> [Accessed 24 November 2020].

Financial Stability Board, 2017. Policy Recommendations To Address Structural Vulnerabilities From Asset Management Activities. [pdf] Available at: <https://www.fsb.org/wp-content/uploads/FSB-Policy-Recommendations-on-Asset-Management-Structural-Vulnerabilities.pdf> [Accessed 24 November 2020].

French, K., 2020. Data Library. [online] Available at: <https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html> [Accessed 23 December 2020].

Gilovich, T., Vallone, R. and Tversky, A., 1985. The hot hand in basketball: On the misperception of random sequences. Cognitive Psychology, 17(3), pp.295-314.

Goldstein, I., Jiang, H. and Ng, D., 2017. Investor flows and fragility in corporate bond funds. Journal of Financial Economics, 126, pp. 592–613.

Graham, B. and Zweig, J., 2003. The Intelligent Investor - Revised Edition. New York: PerfectBound.

Granger, C., 1969. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. Econometrica, 37(3), pp.424-438.

Greene, M., 2018. Passive Investing Is Storing Up Trouble. [online] Available at: <https://www.ft.com/content/cdbdd01a-95b4-11e8-95f8-8640db9060a7> [Accessed 25 November 2020].

Grossman, S. and Stiglitz, J., 1980. On the Impossibility of Informationally Efficient Markets. The American Economic Review, 70(3), pp.393-408.

Halbritter, G. and Dorfleitner, G., 2015. The wages of social responsibility — where are they? A critical review of ESG investing. Review of Financial Economics, 26, pp. 25-35.

Hasbrouck, J., 2003. Intraday Price Formation in U.S. Equity Index Markets. The Journal of Finance, 58(6), pp.2375-2399.

Hill, J., Nadig, D. and Hougan, M., 2015. A COMPREHENSIVE GUIDE TO EXCHANGETRADED FUNDS (Etfs). [pdf] The CFA Institute Research Foundation. Available at: <https://www.cfainstitute.org/-/media/documents/book/rf-publication/2015/rf-v2015-n3-1-pdf.ashx> [Accessed 25 November 2020].

Kaminska, I., 2020. A Little ETF Rule Change That Could Make A Big Difference. [online] Financial Times. Available at: <https://ftalphaville.ft.com/2020/02/05/1580899809000/A-little-ETF-rule-change-that-could-make-a-big-difference/> [Accessed 24 November 2020].

Latif, M., Arshad, S., Fatima, M. and Farooq, S., 2011. Market Efficiency, Market Anomalies, Causes, Evidences, and Some Behavioral Aspects of Market Anomalies. Research Journal of Finance and Accounting, 2(9/10).

Lettau, M. and Madhavan, A., 2018. Exchange-Traded Funds 101 for Economists. Journal of Economic Perspectives, 32(1), pp.135-154.

Lin, C. and Chiang, M., 2005. Volatility effect of ETFs on the constituents of the underlying Taiwan 50 Index. Applied Financial Economics, 15(18), pp.1315-1322.

Malamud, S., 2015. A Dynamic Equilibrium Model of ETFs. SSRN Electronic Journal,.

Matos, P., 2020. ESG AND RESPONSIBLE INSTITUTIONAL INVESTING AROUND THE WORLD. [pdf] The CFA Institute Research Foundation. Available at: <https://www.cfainstitute.org/-/media/documents/book/rf-lit-review/2020/rflr-esg-and-responsible-institutional-investing.ashx> [Accessed 24 November 2020].

Mehra, R. and Prescott, E., 1985. The Equity Premium: A Puzzle. Journal of Monetary Economics, 15(2), pp. 145-161.

Pan, K. and Zeng, Y., 2017. ETF Arbitrage Under Liquidity Mismatch. [pdf] European Systemic Risk Board. Available at: <https://www.esrb.europa.eu/pub/pdf/wp/esrb.wp59.en.pdf?c32ad640616557344d67f09d62590646> [Accessed 24 November 2020].

Patel, N. and Welch, I., 2016. Extended Stock Returns In Response To S&P 500 Index Changes. [pdf] Available at: <https://ssrn.com/abstract=2638660> [Accessed 24 November 2020].

Pisani, B., 2020. U.S. ETF Market Tops $5 Trillion In Assets As Investors Stampede Into Stocks On Vaccine Hopes. [online] CNBC. Available at: <https://www.cnbc.com/2020/11/17/us-etf-market-tops-5-trillion-in-assets-as-investors-stampede-into-stocks-on-vaccine-hopes.html> [Accessed 30 December 2020].

Reinicke, C., 2019. Big Short' Investor Michael Burry Is Calling Passive Investment A 'Bubble.' He's Not The Only Finance Luminary Sounding The Alarm. [online] Available at: <https://markets.businessinsider.com/news/stocks/investors-that-have-spoken-out-against-passive-investing-2019-8-1028485512#jack-bogle5> [Accessed 25 November 2020].

SEC, 2019. Exchange-Traded Funds. [online] Securities and Exchange Commission. Available at: <https://www.sec.gov/rules/final/2019/33-10695.pdf> [Accessed 24 November 2020].

Shleifer, A., 1986. Do Demand Curves for Stocks Slope Down? The Journal of Finance, 41(3), pp. 579-590.

Stepek, J., 2020. The Boom In Passive Investing Won’T Cause The Next Crash. [online] Available at: <https://moneyweek.com/520292/passive-investing-etfs-wont-cause-the-next-market-crash> [Accessed 25 November 2020].

Sullivan, R. and Xiong, J., 2012. How Index Trading Increases Market Vulnerability. Financial Analysts Journal, 68(2), pp.70-84.

Sushko, V. and Turner, G., 2018. The Implications Of Passive Investing For Securities Markets. [pdf] BIS Quarterly Review. Available at: <https://www.bis.org/publ/qtrpdf/r_qt1803j.pdf> [Accessed 24 November 2020].

Thaler, R., 2009. Markets Can Be Wrong And The Price Is Not Always Right. [online] FINANCIAL TIMES. Available at: <https://www.ft.com/content/efc0e92e-8121-11de-92e7-00144feabdc0> [Accessed 23 November 2020].

U.S. Commodity Futures Trading Commission and the U.S. Securities and Exchange Commission, 2010. FINDINGS REGARDING THE MARKET EVENTS OF MAY 6, 2010. [pdf] U.S. Securities and Exchange Commission. Available at: <https://www.sec.gov/news/studies/2010/marketevents-report.pdf> [Accessed 25 November 2020].

U.S. Department of Justice, 2018. Herfindahl-Hirschman Index. [online] Available at: <https://www.justice.gov/atr/herfindahl-hirschman-index> [Accessed 25 November 2020].

U.S. Securities and Exchange Commission Fixed Income Market Structure Advisory Committee, 2018. Recommendation For An Exchange-Traded Product Classification Scheme. [pdf] U.S. Securities and Exchange Commission. Available at:

<https://www.sec.gov/spotlight/fixed-income-advisory-committee/fimsac-etp-naming-convention-recommendation.pdf> [Accessed 25 November 2020].

Wang, K. and Xu, J., 2015. Market volatility and momentum. Journal of Empirical Finance, 30, pp.79-91.

APPENDICES

Appendix 1. ADF test for ETFV and RETFV

ADF for ETFV	ADF for RETFV
Augmented Dickey-Fuller test for ETFV testing down from 12 lags, criterion AIC sample size 109 test with constant: asymptotic p-value 4.119e-005 with constant and trend asymptotic p-value 7.949e-005	Augmented Dickey-Fuller test for RETFV testing down from 12 lags, criterion AIC sample size 111 test with constant: asymptotic p-value 0.01741 with constant and trend p-value 2.294e-007

Source: author's calculations

Appendix 2. GARCH(1,1) model output

Model: GARCH, using observations 2010-08-02:2019-12-31 (T = 2371)
Dependent variable: RSPX
Standard errors based on Hessian

	Coefficient	*Std. Error*	*z*	*p-value*	
const	0.000791197	0.000137521	5.753	<0.0001	***
alpha(0)	4.13650e-06	6.18205e-07	6.691	<0.0001	***
alpha(1)	0.179080	0.0201846	8.872	<0.0001	***
beta(1)	0.772533	0.0212844	36.30	<0.0001	***

Mean dependent var	0.000454	S.D. dependent var	0.009060
Log-likelihood	8143.746	Akaike criterion	−16277.49
Schwarz criterion	−16248.64	Hannan-Quinn	−16266.99

Unconditional error variance = 8.54881e-005

Appendix 3. Non-exclusive licence

A non-exclusive licence for reproduction and publication of a graduation thesis[11]

I ___Veronika Matrossova________________________________ (*author's name*)

1. Grant Tallinn University of Technology free licence (non-exclusive licence) for my thesis VOLATILITY EFFECT OF ETFs ON THE UNDERLYING ASSETS ON THE EXAMPLE OF S&P 500____
____
____,

(*title of the graduation thesis*)

supervised by________________Kalle Ahi____,

(*supervisor's name*)

1.1 to be reproduced for the purposes of preservation and electronic publication of the graduation thesis, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright;

1.2 to be published via the web of Tallinn University of Technology, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright.

2. I am aware that the author also retains the rights specified in clause 1 of the non-exclusive licence.

3. I confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights, the rights arising from the Personal Data Protection Act or rights arising from other legislation.

_04.01.2021______________ (date)

[1] *The non-exclusive licence is not valid during the validity of access restriction indicated in the student's application for restriction on access to the graduation thesis that has been signed by the school's dean, except in case of the university's right to reproduce the thesis for preservation purposes only. If a graduation thesis is based on the joint creative activity of two or more persons and the co-author(s) has/have not granted, by the set deadline, the student defending his/her graduation thesis consent to reproduce and publish the graduation thesis in compliance with clauses 1.1 and 1.2 of the non-exclusive licence, the non-exclusive license shall not be valid for the period.*