The Implications of Passive Investing on Market Volatility and Efficiency

Investigating the Consequences of Passive Investing on the S&P 500 Index and Stocks Added to the Index During 2015-2017

Lasse Mejlholm Larsen

Study No.: 20184557

Supervisor: Cesario Mateus

Aalborg University Business School

MSc in Finance - Master's Thesis

Date

01-06-2023

Abstract

The paper seeks to investigate the influence of growing volumes in passive index funds on market volatility and efficiency in the United States. Existing literature strongly asserts that the inclusion of stocks in indexes leads to increased volatility due to the trading patterns being influenced by passive investors following a specific set of rules. However, the findings regarding market efficiency have been inconclusive. Some studies suggest a decline in price discovery and an increase in return co-movement, which negatively affect market efficiency. On the other hand, increasing trading volumes, narrowing bid-ask spreads, and enhanced liquidity indicate a potential positive impact. To estimate the portion of the S&P 500 index held by ETFs, data on ETFs' assets under management were utilized. Valuation metrics were employed to assess price discovery, and data on S&P 500 index additions during the period of 2015-2017 were collected to analyze changes in volatility and market efficiency. The findings indicate an increase in volatility, an increase in return co-movement, a rise in trading volume, a widening bid-ask spread, and mixed results relating to the change in liquidity.

Contents

1. Introduction 1
2. Literature Review 2
3. Research Question 7
   3.1 S&P 500 Index Inclusion Criteria 8
4. Methodology 9
   4.1 Data Collection 9
      4.1.1 ETF Statistics 9
      4.1.2 Price Discovery 9
      4.1.3 Data on Individual Stocks 9
   4.2 Analytical Methods 13
      4.2.1 Notation 13
      4.2.2 Volatility Measure 13
      4.2.3 Return Co-Movement 14
      4.2.4 Trading Volume 15
      4.2.5 Bid-Ask Spread Estimator 15
5. Analysis 18
   5.1 Market Share of Passive Index Funds 18
   5.2 Relationship Between Passive Index Funds and Market Volatility 20
   5.3 Relationship Between Passive Index Funds and Market Efficiency 23
      5.3.1 Price Discovery 23
      5.3.2 Repercussions of Stocks Added to the S&P 500 Index 23
6. Discussion 28
   6.1 Passive Index Funds and Market Volatility 29
      6.1.2 Summary of Volatility Findings 36
   6.2 Passive index funds and market efficiency 37
      6.2.1 Price Discovery 37
      6.2.2 Return Co-Movement 37
      6.2.3 Trading Volume 39
      6.2.4 Bid-Ask Spread 43
      6.2.5 Liquidity 47

6.2.6 Summary of Market Efficiency Findings 48
7. Conclusion 49
8. Bibliography 51
Appendix 57
1) Stocks Added to the S&P 500 Index During 2015-2017 57

1. Introduction

Over the past thirty years, there has been a significant shift from active investing to passive investing, which has been gaining momentum. Passive ETFs operate based on predetermined rules rather than stock fundamentals, and this rise in passive investing has resulted in notable economic implications. In 1990, the world witnessed the launch of the first ETF in Canada, followed by the introduction of the SPDR S&P 500 ETF in the United States in 1993 (Vanguard, n.d.; Deville, 2008). Since then, ETFs have become an increasingly popular investment product (Bianco Research, 2018; Collins Advisors, n.d.; Vanguard, n.d.). By September 2008, net inflows into ETFs exceeded those into mutual funds, and this trend has continued, with ETFs experiencing increased inflows while mutual funds have seen a decline (Bianco Research, 2018). Notably, 97% of ETFs are classified as passive (Anadu, Kruttli, McCabe, & Osambela, 2020).

For instance, one notable ETF, the SPDR S&P 500 ETF Trust, has the objective of mirroring the performance of the S&P 500 index. In fact, the three largest ETFs in terms of assets under management (AUM) all focus on the S&P 500 index (VettaFi, n.d.). When a significant amount of investment flows into a collection of ETFs that track the same index, it begins to have meaningful consequences.

When enough flows go into a limited number of stocks, the valuation premium of these stocks rises. The S&P 500 constituents trades on premium compared to non-index stocks and the premium is growing (Wurgler, On the Economic Consequences of Index-Linked Investing, 2011). Worst case scenario is the creation of 'indexing bubble', which might bust (Wurgler, On the Economic Consequences of Index-Linked Investing, 2011).

Other consequences include index-inclusion effects on new additions to the index. Passive investing may boost the volatility of underlying asset as they are forced to rebalance their portfolios in order to replicate the market, which implies trading in the same direction as the same-day market moves (Anadu, Kruttli, McCabe, & Osambela, 2020). Additionally, stock returns are elevated around the index inclusion date (Beneish & Whaley, 2002; Wurgler & Zhuravskaya, 2002), and the return co-movement with fellow index constituents increases upon index inclusion (Claessens & Yafeh, 2011), which limits the diversification gain. The trading volume of the index additions is found to increase quite significantly (Edmister, Graham, & Pirie, 1996) together with a bid-ask spread that tends to narrow (Edmister, Graham, & Pirie, 1996), suggesting an increase in liquidity.

In this project, the trading pattern of stocks added to the S&P 500 index during the period January 1. 2015 to December 31. 2017, will be analyzed. Stocks added during the period, but that do not have 5 years of trading data before and after their inclusion date will be excluded. The project will analyze how the sample stocks' trading pattern regarding volatility, volume, return co-movement, and bid-ask spread changes from different periods before and after being added to the S&P 500 index.

Generally, the market share of passive investing continues to increase, potentially causing an elevated valuation of S&P 500 stocks. When analyzing the ±3- and ±5-year periods relative to the index inclusion date, the results seem to be heavily affected by the COVID-outbreak. Adjusting for the COVID-outbreak by focusing on shorter time periods there is found an increasing volatility, co-movement, trading volume, and a widening bid-ask spread.

Following this introduction, a literature review is presented followed by the methodology. Hereafter comes the analysis of market volatility and market efficiency, whereafter the results of the analysis are discussed. Subsequently, the main findings of this report are summed up in a conclusion.

2. Literature Review

The literature is structured as follows. First, there is an explanation of the rationale behind passive investing supporting the massive flows coming into passive investing products i.e., ETFs. Hereafter, the repercussions of this development on a broader and on a single stock level will be clarified.

The rationale of passive investing and the rising popularity hereof

A passive investment strategy may seem rational when having the efficient market hypothesis in mind. This theory states that all available information is instantly incorporated in the markets' pricing of securities, implying that future alpha is not predictable. Following this mindset then the possibility for active investment managers to outperform the market is very limited if not impossible (Sushko & Turner, 2018). This questions the rationale of the higher management fees active funds tends to have. 90% of passively and just 13% of actively managed funds have an annual fee below 0.5% (Graves & Pratt, 2022).

Moreover, the proposition is put forth that market outperformance operates as a zero-sum game as posited by (Sharpe, 1991). In essence, this implies that some investors' excess returns are offset by the losses of others. Consequently, all active managers returns pooled together before costs are equal to all passive managers' pooled returns before costs, whereby the average return for active managers after trading costs will be less than for passive managers (Sushko & Turner, 2018).

The above-mentioned factors have all boosted the trend towards more flows coming into passive investing strategies. In the period 2003-2022 the AUM in ETFs has increased more than 46x and reached 9,552 billion dollars globally (Statista, 2023). In this period, the AUM has been growing consistently year-over-year with 2008, 2018, and 2022 as the only exceptions with negative growth (Statista, 2023). In the same period, the number of ETFs worldwide has increased from 276 in 2003 to 8,754 in 2022 (Statista, 2023). This trend is a global phenomenon as index funds in most countries globally account for an increasing market share during the decade following the financial crisis in 2008 (Carneiro, Junior, & Yoshinaga, 2022), but of developed markets the trend is strongest in the United States, where the share was 14.7% in 2017, up from 6% in 2007 (Sushko & Turner, 2018). Many investors link ETFs to passive index investing, but there are multiple types of ETFs. In its essence an ETF, Exchange-Traded-fund, is a portfolio carrying multiple securities and the number of possible combinations of securities exceeds the actual number of securities multiple times. Today, there are ETFs available to investors who want exposure to almost any opportunity. Some of the more common ETFs are equity ETFs, bond/fixed income ETFs, commodity ETFs, currency ETFs, sustainable ETFs, factor ETFs, and specialty ETFs i.e., inverse funds and leveraged funds (BlackRock, u.d.). There are both active and passive ETFs within each category. Data from Morningstar suggest that 97% of ETF assets are in passive ETFs (Anadu, Kruttli, McCabe, & Osambela, 2020).

REPURCUSSIONS OF ETFS INCREASING MARKET SHARE

The current literature provides mixed results regarding the repercussions of the current shift from active passive investing. Repercussions of the rising share of passive investment are assessed by paying attention to either general market stability defined by flows, asset management industry concentration, valuation and volatility levels or the index-inclusion effects on single stocks.

Some academic papers find a positive relationship between correlation between fund flows and the performance of passive mutual funds and ETFs (Goetzmann & Massa, 2003; Clifford, Fulkerson, Jordan, & Waldman, 2014). Others argue that passive mutual funds serve to mitigate the risk associated with substantial and procyclical fund flows, as well as significant redemptions during periods of financial strain. Notably, even when both passive and active funds encounter comparable subpar returns, passive mutual funds continue to attract net inflows (Anadu, Kruttli, McCabe, & Osambela, 2020).

These net inflows are going into a relatively small number of ETF suppliers causing some management concentration risk. When comparing the active and passive asset-management industries it is shown that competition is far more concentrated regarding the passive, where the 10 largest passive fund managers combined accounts for about 90% of the total passive-fund AUM (Anadu, Kruttli, McCabe, & Osambela, 2020). This difference, is supported by the fact that greater AUM allows fixed costs to be spread across more assets, which reduces costs (Anadu, Kruttli, McCabe, & Osambela, 2020) and for actively managed funds the ability to outperform the market weakens as funds get bigger (Green & Berk, 2004). This trend raises some serious risks associated with a possible idiosyncratic event at any of the large asset-management firms leading to massive redemptions (Anadu, Kruttli, McCabe, & Osambela, 2020).

Index inclusion effects

The S&P Index Committee determines when there will be a shift in S&P 500 constituents. A stock is removed from the S&P 500 index if it gets below certain levels regarding liquidity, gets delisted or if its otherwise decided by the S&P Index Committee that a stock is less representative of the market than the next available candidate (Wurgler, 2011). These additions and removals have significant consequences for those stocks trading characteristics:

Passive index investing may amplify the volatility of underlying assets by forcing fund managers to trade in the same direction as the same-day market moves (Anadu, Kruttli, McCabe, & Osambela, 2020). (Ben-David, Franzoni, & Moussawi, 2018) found that stocks with higher ETF ownership tend be more volatile than peers with less ETF ownership. (Steliaros & Rossi, 2022), who analyze the US stock market, find that the effect of rising volatility is concentrated at the end of a trading day, because it might be when ETFs rebalance their portfolios. Additionally, (Bogusslavsky & Muravyev, 2022) finds that when stocks are added to or deleted from the S&P 500 index closing volumes permanently increase by 20% and decrease by 15% relatively to the intraday respectively. Furthermore, they show that institutional investors deploy inflows at the beginning of month resulting in a higher turnover and that the turnover seems to be higher on option expiry days (Bogusslavsky & Muravyev, 2022). This phenomenon is affecting stocks added to the S&P 500 index as institutional investors' investment universe is limited due to their size, which assumably means there are more institutional investors in larger stocks, i.e., S&P 500 constituents, than in smaller stocks.

More broadly, (Kamara, Lou, & Sadka, 2008; Kamara, Lou, & Sadka, 2010) shows that average return betas increased and decreased for large and small stocks respectively during the period 1968-2008, explained by passive investing having a higher impact on large stocks. (Bolla, Kohler, & Wittig, 2016) obtain similar trends regarding increasing betas in developed equity markets but found the trend to slow down during the financial crisis. In contrast, (Honghui, Singal, & Whitelaw, 2016) who continued the work from (Barberis, Schleifer, & Wurgler, 2005) on index inclusion effects, do not find an upward trend in betas during 2001-2012, which was not covered in the older paper. Several researchers have come to similar conclusions when arguing that stocks' systematic risk, beta or excess volatility increases after being added to an index or more specifically added to the S&P 500 index (Vijh, 1994; Honghui, Singal, & Whitelaw, 2016; Barberis, Schleifer, & Wurgler, 2005; Sullivan &

Xiong, 2012). (Honghui, Singal, & Whitelaw, 2016) found a smaller effect on stock betas when added to the S&P 500 index, even though index-investing strategies have become more common. An older study from (Vijh, 1994), who analyze the daily and weekly betas of stock added to the S&P 500 index during 1985-1989[1] and find an average increase of 0.211 and 0.130 after inclusion respectively. Thereby, inclusion-effects resulting in increasing volatility started prior to the rise of passive investing through ETFs. (Barberis, Schleifer, & Wurgler, 2005) builds on (Vijh, 1994) and found even stronger results on more recent data.

When individual stocks are added to an index, passive flows from ETFs come towards that stock within a relatively short time frame. Calculations approximate that about 8.7% of each stock added to the S&P 500 index must be bought by index fund managers (Wurgler, On the Economic Consequences of Index-Linked Investing, 2011). In an investigation of the American market from 1990 to 2005, (Petajisto, 2011) find these "inclusion effects" to start by the date when the change in index constituents is announced. From announcement to the effective day additions has averaged +8.8% and +4.7%, while deletions have averaged -15.1% and -4.6% for the S&P 500 and Russel 2000 respectively (Petajisto, 2011). Several other studies have shown comparable inclusion effects on stock prices after joining the S&P 500 index, (Beneish & Whaley, 2002) (Wurgler & Zhuravskaya, 2002). (Harris & Gurel, 1986) find an immediate price increase of 3 percent, when new additions to the S&P 500 is announced, (Lynch & Mendenhall, 1997) find this surge to be 3.807%, and for index deletions the pattern is similar, but inverted (Beneish & Whaley, 1996) and (Schleifer, 1986) confirm that new additions' prices increases more on the announcement until the effective day. Additionally, (Patel & Welch, 2017) found that the surge in stock prices after index inclusion tend to be rather short term, because price increases do not seem to be permanent. Oppositely, (Morck & Yang, 2001) find the surge to be rather permanent, when they document a large value premium of firms in the S&P 500 index relative to peers outside of the index and the value premium is growing over time. This premium is due to either some intangible asset connected with being a member of the S&P 500 index or a direct result of the growth of index investing (Morck & Yang, 2001).

The index inclusion-effects on stocks' liquidity can be twofold. The liquidity can increase as some assets listed on smaller exchanges might be easier to trade indirectly, when being part of an ETF listed on major exchanges. (Brogaard, Heath, & Huang, 2019) found mixed results, when concluding that passive equity ETFs make liquid equites more liquid and illiquid equities less liquid. Another study, find index-inclusion effect on liquidity to depend on whether or not the stock is optioned (Erwin & Miller, 2005). (Hegde & McDermott, 2003) also found

[1] Note: this is before the launch of the first ETF (find source)

a sustained increase in liquidity among additions to the S&P 500 index but tied it to the decrease in direct trading costs and a decline in the asymmetric information component. To verify the index-inclusion effect they additionally found deleted stocks' liquidity to decline over the following three months after exclusion (Hegde & McDermott, 2003).

Trading volume, the number of shares traded, can be considered as a measure of liquidity. A high trading volume means a stock is traded more frequently suggesting a high liquidity and vice versa (CFI, 2023). (Edmister, Graham, & Pirie, 1996) find a permanent increase in volume upon a stock's inclusion in the S&P 500 index indicating improved liquidity. (Harris & Gurel, 1986) who analyzed the short-term impact on stocks' trading volume when the index inclusion of a stock is announced and found significant increases on both a 1-day and a 1-week basis.

Another discussed index-inclusion is about how stocks return co-movement changes upon addition or deletion from an index. Fund managers of passive index-funds trade a set of rules, which largely is that the return of the ETF must be similar to the return of the index. This implies that the holdings must be equal to the holdings and the individual weightings must be aligned with those of the index, causing these managers to buy and sell the same stocks simultaneously. In this regard, (Da & Shive, 2016) found that a higher ETF ownership increase stocks return co-movement. (Claessens & Yafeh, 2011) who study stock data from additions to indices in both developed and undeveloped countries around the world, find an increasing return co-movement upon index inclusion to be a global phenomenon. Furthermore, they find this phenomenon to prevail more recently. Part of the reasoning behind this is that post-inclusion stocks in general experience increases in analyst coverage, liquidity and to investor behavior (Claessens & Yafeh, 2011). However, it can be discussed how co-movement has changed with the increasing popularity of passive investing strategies as more index-inclusion effects was there before the 1990s when the first ETF was launched (Grégoire, 2020; Vijh, 1994). (Grégoire, 2020) found an increasing trend in return co-movement in S&P 500 stocks prior to 1990, or prior to the rise of passive investing through ETFs. Therefore, he splits his analysis into two periods: before and after 1990. In the period 1990-2017, the increase in co-movement was even stronger, probably due to the rising of passive investing strategies (Grégoire, 2020). To verify the development, it is found that the co-movement with non-S&P 500 stocks decrease upon index inclusion (Grégoire, 2020).

The rise of passive ETFs may also cause an increase in the co-movement of stocks' liquidity. This also means that the risk of stocks becoming illiquid simultaneously increases. Both (Kamara, Lou, & Sadka, 2008), (Kamara, Lou, & Sadka, 2010), and (Bolla, Kohler, & Wittig, 2016), tie the increasing trend in systematic liquidity among American equities to the increasing market share of institutional and indexed investing.

When investigating the relationship between single stocks and index-inclusion effects, many researchers add the impact on bid-ask spreads to the research area. (Edmister, Graham, & Pirie, 1996) explore how trading costs are affected by S&P 500 index replacement stock announcements. They found the relative bid-ask spread on replacement stocks to narrow more for stocks, which prior to inclusion had a wider spread than stocks with a narrower spread (Edmister, Graham, & Pirie, 1996). (Erwin & Miller, 2005) find similar results regarding stocks' bid-ask spread that narrows after being included in the S&P 500 index, however only for stocks that were not trading listed options. It is argued that the narrowing bid-ask spread for non-optioned stocks is due to index arbitrage trading, while this effect is mitigated for optioned stocks (Erwin & Miller, 2005).

3. Research Question

This project aims to answer the following research question:

"How does the rising volumes in passive index funds impact the American market's volatility and efficiency?"

To answer this research question, the American S&P 500 index and its constituents is used as a proxy of the market. An American index is chosen as index funds and ETFs together have a ~12% share of the American equity market or 58% more than corresponding share of the global equity market, ~7% (BlackRock, 2017). Because the S&P 500 index is not directly tradable, the trading data of the SPDR S&P 500 ETF Trust, world biggest ETF in terms of AUM (VettaFi, u.d.), is used in the calculations. It will be estimated how big a percentage of the S&P 500 constituents market capitalization is placed in ETFs solely covering the index.

The relationship between passive index funds and the market's volatility will be discovered by estimating the volatility for various periods and holding it up against the sample stocks' index inclusion date. Thereby it will be possible figure out how volatility changes from periods prior to the index inclusion date to periods after.

The efficient market hypothesis states that security prices fully reflect all available information (Fama, 1970). In its purest form, both publicly and privately held information is reflected in the prices. This hypothesis cannot be tested with one specific analytical tool as information cannot be measured. This project aims to investigate how an increasing market share from passive index funds affects market efficiency. First, it will be tested if the spread between US small and large cap regarding key valuation metrics has changed over time. Secondly, it will be tested how sample stocks' trading statistics, like return co-movement with other index constituents, trading volumes, bid-ask spread, and liquidity change after being added to the S&P 500 index.

3.1 S&P 500 Index Inclusion Criteria

Before proceeding to the methodology, it is important to have some background knowledge regarding the importance of the index and the criteria determining when a stock is capable of being part of the S&P 500 index.

The S&P 500 index is often referred to as a prominent benchmark for the U.S. large cap or for general U.S. equity performance (S&P Global, 2021). Because the constituents of this single index has a combined market cap 33.78T or 83.4% of the total U.S. equity market value (Siblis Research, u.d.), it is of importance to be aware of key parameters of how this index is constructed.
The following illustration provides a simple overview of the S&P 500 index committee.

Figure 1: Index Committee Oversight

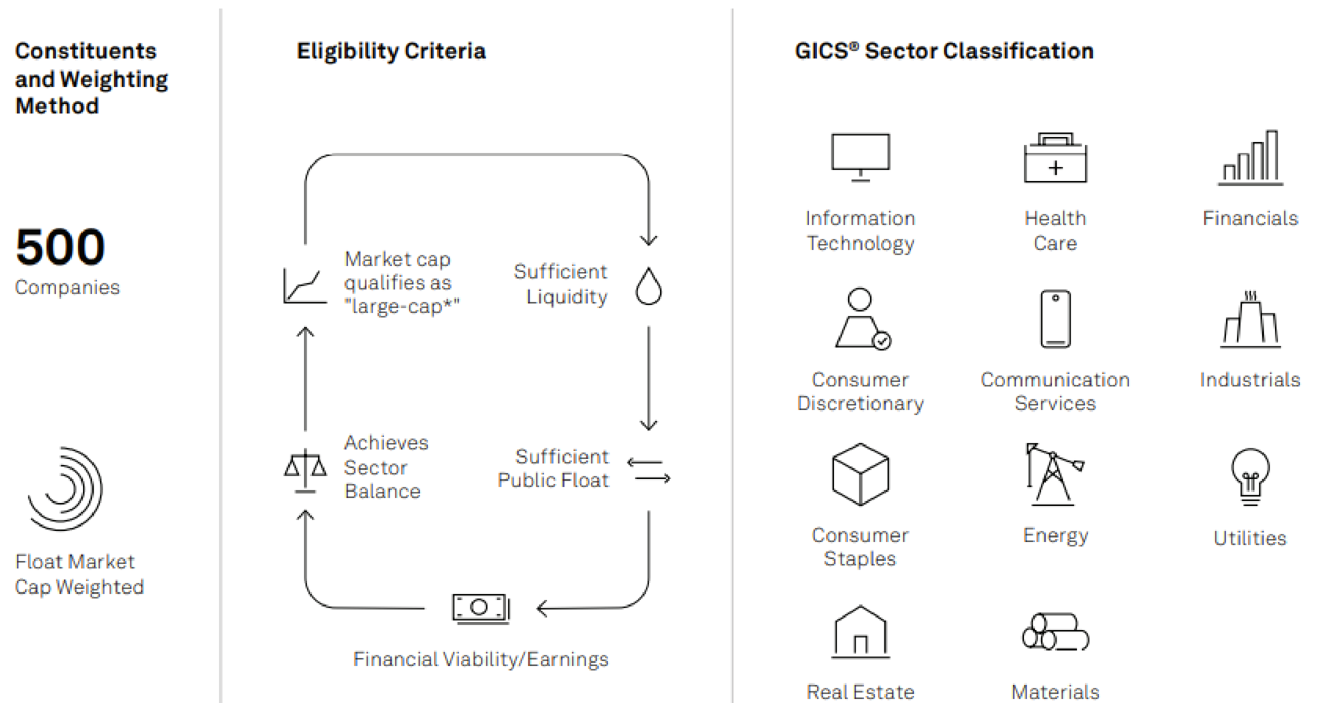


Figure 1: showing the process of how companies are selected for the S&P 500 index. Source: (Standard & Poor).

The illustration shows that the S&P 500 index consists of 500 U.S. large cap companies, and their weightings is determined by its float-adjusted market cap. To be eligible certain criteria regarding liquidity and tradability must be met. Furthermore, the index committee are using the GICS sector classification to provide details on the sectors' weightings in the index and to make sure the index keeps its relevance (Standard & Poor). More detailed information about inclusion criteria for the S&P 500 index can be found in the S&P U.S. Indices Methodology (Standard & Poor, 2023).

4. Methodology

This section presents the methodology applied to answer the research question. The methodology section presents how data has been conducted and processed in order to get the desired results.

4.1 Data Collection

4.1.1 ETF Statistics

Passive index funds' share of the S&P 500 index has been calculated by using multiple data sources. Initially, data on the top 200 ETFs based on AUM and filtered by all asset classes (VettaFi, u.d.). Hereafter, the top 200 equity ETFs are filtered to those being passive and covering the S&P 500 index, totaling 15 ETFs. The AUM of these 15 ETFs is divided by the AUM of the top 200 ETFs among all asset classes resulting in that 20.19% of all ETFs covering all asset classes' assets are placed in the S&P 500 index. Assuming the share to be constant it is possible to estimate passive index funds' share of the S&P 500 index by multiplying the AUM of all ETFs globally (Statista, 2023) by 20.19% divided by the S&P 500 market cap over time (YCharts, u.d.).

$$Share\ of\ ETFs'total\ AUM\ covering\ the\ S\&P\ 500 = \frac{AUM\ of\ ETFs\ covering\ the\ S\&P\ 500\ index}{AUM\ of\ all\ ETFs}$$

$$S\&P\ 500\ ETF\ ownership\ = (S\&P\ 500\ total\ market\ cap)\ x\ (Share\ of\ ETFs'\ total\ AUM\ covering\ the\ S\&P\ 500)$$

$$S\&P\ 500\ ETF\ ownership\ share = \frac{S\&P\ 500\ ETF\ ownership}{S\&P\ 500\ total\ market\ cap}$$

4.1.2 Price Discovery

To estimate whether American equities have an increasing tendency to be mispriced I have calculated the spread between US small and large cap's price-to-sales multiple over time. As a proxy for the Russel 2000 and S&P 500 index I have used the IShares Russel 2000 ETF and SPDR S&P 500 ETF Trust respectively. These are the two biggest ETFs covering each index based on AUM (FactSet 28-02-2023).

4.1.3 Data on Individual Stocks

To calculate whether stocks tend to have an increasing co-movement, increased trading volume and a narrower bid-ask spread after inclusion in the S&P 500 index, trading data on individual stocks has been gathered. Trading data includes daily data on stock open, high, low, and closing prices, and volumes traded. What is common for

these individual stocks is that they are all members of the index today and they were all added to the index during 2015-2017, which means a total of 66 stocks were added (Appendix, 1). Hereafter all stocks that didn't have 5 years of daily trading data, corresponding to approximately 2518 observations, both before and after the date of inclusion was filtered away. The missing data could be due to e.g., IPO dates, de-listings, spinoffs, or acquisitions. This narrows the sample of stocks for further analysis down to 53.
The average values of statistics regarding the individual stocks will be used as a sample to calculate general trends regarding general repercussions of stocks joining the S&P 500 index.

All the sample stocks inclusion dates within the required period are plotted in the figure below. All blue lines represent a stock being added to the index on the specific date. It should be noticed that there tends to be some clustering, meaning that in some periods more stocks are added to the index within a short period of time than in others. For instance, during the first six months of 2016 and from March to July 2017, where 15 and 17 stocks were added to the index respectively (Appendix, 1).

Figure 2: Timeline of Sample Stocks' Index Inclusion Dates

Figure 2: showing a timeline of inclusion dates for the 53 sample stock. Each date is marked with a blue line. Data from Appendix 1.

The next figure, shows an illustration of the data collection methodology and the implications of such:

- Black line: represents the inclusion dates of sample stock companies from early 2015 to 2017 year-end.

- Red areas: represents the five years of data collected before and after the date for samples stocks' inclusion to the S&P 500 index.
- Grey line: represents a global macro event affecting all sample stocks.
- Numbers to the right: represent the number of trading days before and after the index inclusion date (date 0). 252, 756, and 1259 correspond to the trading days of one, three, and five years. The plus and minus sign signal whether the number of trading days after (+) and prior to (-) the sample stock index inclusion date.

Figure 3: Illustration of the Data Collection Process

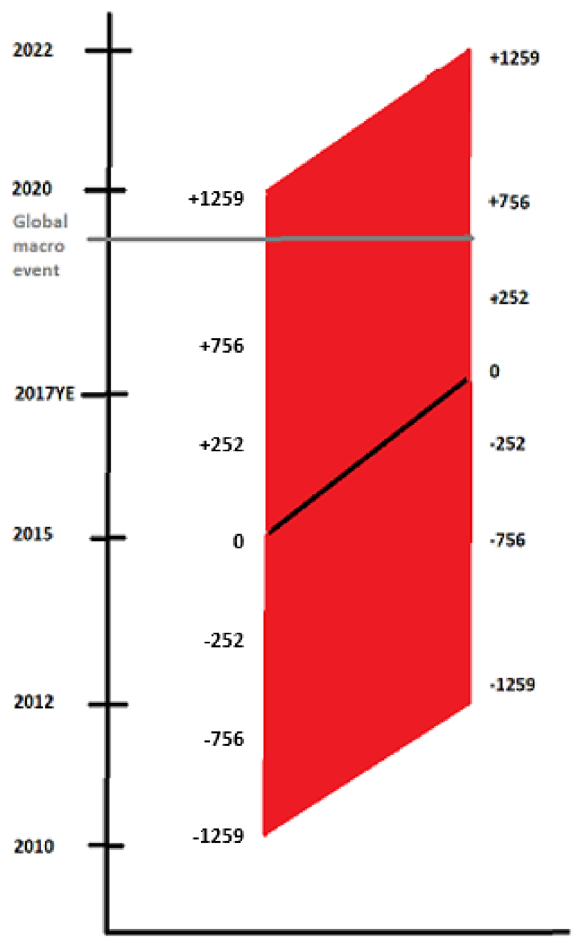


Figure 3: Illustrating the data collection process and its implication relative to a global macro event. Source: own analysis.

To put the illustration into perspective, then an analysis of a sample stock joining the S&P 500 index late 2017(right end of the black line) would be more sensitive to the global macro event (grey line) than a stock added to the index early 2015(left end of the black line. This is illustrated by the cross between the upper red square and the grey line. For a stock joining the index late 2017, there is only a minor part of the collected data which occur prior to the global macro event, whereas for a stock joining early 2015 it is only a minor part of the collected data which happens after the global macro event.

The 53 sample stocks used for analysis and the 503 S&P 500 stocks as of 2022 year-end (Wang, 2023) are distributed among the following sectors:

Table 1: Sector Distribution of Sample Stocks and S&P 500 Index

GICS Sector	Sample Stocks	Sample Weighting	S&P 500	Weighting
Communication Services	3	5.66%	25	4.97%
Consumer Discretionary	4	7.55%	56	11.13%
Consumer Staples	1	1.89%	33	6.56%
Energy	1	1.89%	23	4.57%
Financials	5	9.43%	67	13.32%
Health Care	11	20.75%	63	12.52%
Industrials	6	11.32%	70	13.92%
Information Technology	8	15.09%	76	15.11%
Materials	2	3.77%	29	5.77%
Real Estate	10	18.87%	31	6.16%
Utilities	2	3.77%	30	5.96%
Total	53	100%	503	100.00%

Table 1: shows the 53 sample stocks' distribution and weightings among the different GICS sectors relative to the S&P 500. Source: Own analysis and (Wang, 2023).

The stocks in the table above are equally weighted for comparison reasons. It is known that the S&P 500 weightings are dynamic and based on the specific companies' market cap, but as daily data on those weightings are not available, it assumed that the relative market cap of companies in the sample and the S&P 500 are constant and equal. The sample has an overweight towards health care, real estate, and information technology. The weightings of the sample differ from the weightings of the S&P 500, meaning that the results of the analysis should be looked at with caution. For instance, the real estate sector accounts for 18.87% of the sample and just 6.16% of the index by the end of 2022, which affects the relationship between the sample and the index by making the sample more sensitive to real estate sector specific events.

4.2 Analytical Methods

To answer the research question, several analytical tools are used because key statistics regarding volatility, co-movement, trading volume, and bid-ask spread are to be calculated. Either because they are not directly observable or because they have to be computed from raw data.

4.2.1 Notation

Going forward the time period relative to the stocks index inclusion dates will be referred to as ‘t-5’, ‘t+5’ or ‘t±5’ with ‘t’ being a sample stock’s index inclusion date. The ‘+’ or ‘-‘ sign refer to whether the time is after or before the index inclusion date and the integer indicates the amount of years.

t-5 refers to the 5-year period before the index inclusion date.

t+5 refers to the 5-year period after the index inclusion date.

t±5 refers to the 5-year period before and after the index inclusion date.

A year has 251.8 trading days, whereby t±1, t±3, and t±5 have 252, 756, and 1259 daily observations.

4.2.2 Volatility Measure

Garman Klass is used to measure volatility of the S&P 500 index and the individual stocks. It integrates open, low, high and closing prices of a security. The Garman Klass volatility estimator is described by the following formula from (PortfoliosLab, 2023):

$$\sigma_{GK} = \sqrt{\frac{1}{2T}\sum_{t=1}^{T} \ln\left(\frac{h_t}{l_t}\right)^2 - \frac{2\ln 2 - 1}{T}\ln\left(\frac{c_t}{o_t}\right)^2}$$

Where:

$$T = Number\ of\ days\ in\ the\ sample\ perio$$

$$O_t = Open\ price\ on\ day\ t$$

$$h_t = High\ price\ on\ day\ t$$

$$l_t = Low\ price\ on\ day\ t$$

$$c_t = Close\ price\ on\ day\ t$$

Both the rolling 10- and 30-day Garman Klass volatility estimator will be calculated. The multiday periods are used because the analysis contains 10 years of daily data. First, the 10- and 30-day Garman Klass volatility is estimated for both the SPDR S&P 500 ETF Trust, and the 53 stocks added to the index during 2015-2017. Secondly, the stocks’ excess volatility relative to the SPDR S&P 500 Index ETF Trust is extrapolated to adjust for any general increase or decrease in the market volatility. Lastly, the equally weighted average values of the 53 stocks are used to estimate sample stocks' general change in volatility for the periods t±1, t±3, and t±5.

A z-test is used to determine whether the mean values of two populations are statistically different. This project use a confidence interval 95% to determine whether results are of statistical significance. The null hypothesizes of the Two Sample Z-test is:

$$H_0 = Two\ population\ means\ are\ equal\ (\mu1\ =\ \mu2)$$

$$H_1 = Two\ population\ means\ are\ not\ equal\ (\mu1\ \neq\ \mu2)$$

To uncover whether the results from the z-test are due to a few outlier observations or sectors having a major impact on the mean values, a Wilcoxon Signed Rank Test is used. The Wilcoxon Signed Rank Test compares the ranked median values of two datasets and has the following null hypothesis (Lamorte, 2017):

$$H_0 = The\ median\ difference\ is\ zero$$

$$H_1 = The\ median\ difference\ is\ not\ zero$$

The 10- and 30-day Garman Klass volatility data is split into the following groups: healthcare, real estate, information technology, and the remaining sectors are pooled together. The reason why those three GICS sectors are chosen is because they represent more than 60% of the sample and each of the three sectors represents 8 or more sample stocks. Doing individual analysis on sectors with fewer stocks would make the results more sensitive single stock behavior.

4.2.3 Return Co-Movement

The co-movement of the sample stock returns, and the returns of the S&P 500 index have been calculated for the periods t±1, t±3, and t±5. Because the inclusion date varies in the range from 2015-2017, the statistics; average return, standard deviation, variance, and correlation with the index, have been calculated individually for each of the 53 stocks. For each stock the data source (FactSet) provides two return calculations, normal returns and gross returns. Both returns are included in the analysis to minimize the impact of dividend payments, stock splits etc., which means these calculations are done 106 times.

To test whether the co-movement has changes in the three periods t±1, t±3, and t±5, the z-test is used to test whether the changes in mean values are of statistical significance.

4.2.4 Trading Volume

The correlation between the sample stocks' trading volume and the trading volume of the SPDR S&P 500 ETF Trust is calculated for the periods t±1, t±3, and t±5. As mentioned in the previous section, the inclusion date varies in the range from 2015-2017, why the statistics; average trading volume, standard deviation, variance, and correlation with the index, has been calculated individually for each of the 53 stocks.

To test whether the co-movement changes in the three periods t±1, t±3, and t±5, a z-test is used to test whether the changes are of statistical significance.

4.2.5 Bid-Ask Spread Estimator

The bid-ask spread shows the difference between the highest price a buyer is willing to pay, and a seller is willing to accept (CFI, 2023). This means that the true price of a security is somewhere in between the bid and ask price (Ødegaard, 2023). Figure 4 illustrates this situation. The six upper black dots represent the ask prices for a period with six consecutive trades, while the lower six black dots represent the bid prices for equivalent period. The white does illustrate the true price which is somewhere in between the bid and ask price while the arrows indicate the sequence of trades.

Figure 4: Trading Process

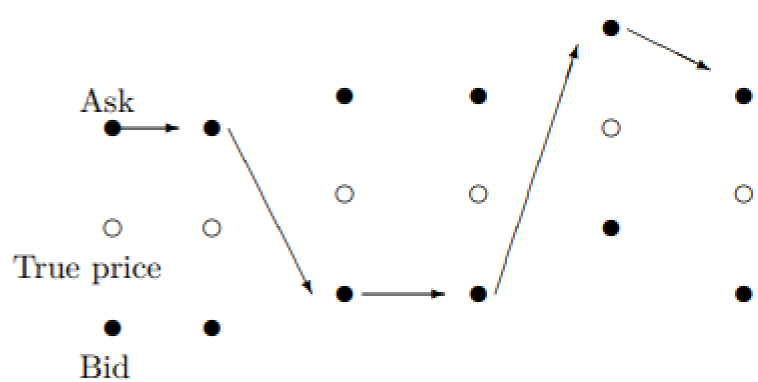


Figure 4: Showing a trading process of six trades with bid and ask prices and the true prices in between. Source: (Ødegaard, 2023)

It is expected that the bid-ask spread narrows after stocks are added to the S&P 500 index, because the increasing awareness of the stock might attract more buyers and sellers and thereby increasing the downward pressure on the ask price and the upward pressure on the bid price. Intraday data on historical bid-ask spreads is not available meaning it will have to be estimated. (Ødegaard, 2023) whose work relies on (Corwin & Schultz, 2012), presents a way to estimate the bid-ask spread by using the daily high and low prices. The idea behind it is that the daily high prices tend to be buyer-initiated, while the daily low prices are seller-initiated. Therefore, the high-low ratio reflects both the volatility and bid-ask spread (Corwin & Schultz, 2012). When no intraday trade and quote date are available, it is necessary to use some estimator to get a sense of the bid-ask spread. (Corwin & Schultz, 2012) has tested their bid-ask spread estimator, where the estimated spread is a function of the high-low ratio for 2 consecutive days and has a correlation coefficient with the true spread more accurate than other spread estimators (Roll, 1984).

The following equations comes from (Ødegaard, 2023) and shows how he deduce the spread estimator:

$$H_{t,t+1}^{0} = (High_t^0, High_{t+1}^0)$$

$$L_{t,t+1}^{0} = (Low_t^0, Low_{t+1}^0)$$

One will then calculate sample estimates

$$\hat{\gamma} = \left[\ln\left(\frac{H_{t,t+1}^0}{L_{t,t+1}^0}\right)\right]^2$$

and

$$\hat{\beta} = \sum_{j=0}^{1}\left[\ln\left(\frac{H_{t+j}^0}{L_{t,+j}^0}\right)\right]^2$$

Or more simply

$$\hat{\beta} = \left(\ln\left(\frac{H_t^0}{L_t^0}\right) + \ln\left(\frac{H_{t+1}^0}{L_{t+1}^0}\right)\right)^2$$

Under some simplifying assumptions one gets a closed form expression for the spread

$$\alpha = \frac{\sqrt{2\beta} - \sqrt{\beta}}{k_2\left(3 - 2\sqrt{2}\right)} + \sqrt{\frac{\gamma}{k_2^2(3 - 2\sqrt{2})}}$$

The volatility is estimated as:

$$\sigma_{HL} = \frac{\sqrt{\frac{\beta}{2}} - \sqrt{\beta}}{\left(k_2\left(3 - 2\sqrt{2}\right)\right)} + \sqrt{\frac{\gamma}{k_2^2\left(3 - 2\sqrt{2}\right)}}$$

Where the constants k_1 and k_2 are calculated as

$$k_1 = 4ln(2)$$

$$k_2 = \sqrt{\frac{8}{\pi}}$$

Instead of estimating 53 individual bid-ask spreads, the 53 sample stocks' *pooled* bid-ask spread is calculated. To do this, an equally weighted portfolio with the average values of all sample stocks' daily high and low prices are created. Due to the formula setup the estimates might be negative as overnight changes to the stock price impact the results. This model assumes no overnight changes to the share price. If the spread is negative, the negative values are set to be zero, so it does not impact the results.

To test whether the bid-ask spread changes in the three periods before sample stocks are added to the index relative to the equivalent period after, the z-test is used to test whether the changes are of statistical significance.

5. Analysis

5.1 Market Share of Passive Index Funds

The following table shows the development of S&P 500 index' market cap and the AUM of all ETFs globally. The fourth column reveals the estimated amount of assets placed in passive ETFs covering the S&P 500 index, and the fifth column shows the passive ETFs share of the S&P 500 index.

Table 2: Estimation of the ETF Ownership Share of the S&P 500

Year	S&P 500 Mkt. Cap (T)	AUM of Global ETFs (T)	AUM of S&P 500 ETFs (T)*	ETF share of S&P 500*
2022	$ 32.13	$ 9.55	$ 1.93	6.00%
2021	$ 40.36	$ 10.02	$ 2.02	5.01%
2020	$ 31.66	$ 7.74	$ 1.56	4.93%
2019	$ 26.76	$ 6.19	$ 1.25	4.67%
2018	$ 21.03	$ 4.68	$ 0.95	4.49%
2017	$ 22.82	$ 4.69	$ 0.95	4.15%
2016	$ 19.27	$ 3.42	$ 0.69	3.59%
2015	$ 17.90	$ 2.90	$ 0.58	3.27%
2014	$ 18.25	$ 2.67	$ 0.54	2.96%
2013	$ 16.49	$ 2.28	$ 0.46	2.79%
2012	$ 12.74	$ 1.77	$ 0.36	2.81%
2011	$ 11.39	$ 1.36	$ 0.27	2.40%
2010	$ 11.43	$ 1.31	$ 0.27	2.32%
2009	$ 9.93	$ 1.04	$ 0.21	2.12%
2008	$ 7.85	$ 0.72	$ 0.14	1.84%
2007	$ 12.87	$ 0.81	$ 0.16	1.27%
2006	$ 12.73	$ 0.58	$ 0.12	0.92%
2005	$ 11.25	$ 0.42	$ 0.08	0.75%
2004	$ 11.29	$ 0.28	$ 0.06	0.51%
2003	$ 10.29	$ 0.20	$ 0.04	0.40%

* Estimated

Table 2: showing the mkt. cap of S&P 500, AUM of ETFs globally, ETFs covering the S&P 500, and the estimated ETF share of S&P 500's total mkt. cap. Source: (Statista, 2023) and own analysis.

During the period 2003 to 2022, the market cap of the S&P 500 has more than tripled and the AUM of global ETFs has increased by more than 47x. A significant share of the inflows to the global ETFs has been going into ETFs covering the S&P 500 index, which has caused ETFs to account for an increasing market share. ETFs share of the S&P 500 market cap has increased from 0.40% in 2001 to 6.00% by the end of 2022.

It should be noted that stocks in the S&P 500 Index could be part of multiple indices, which has a blurring effect on the index-inclusion effects directly related to the inclusion into the S&P 500 index. S&P Global's Annual Survey of Assets states that passively managed assets[2] make up \$5.4T of the S&P 500 index by the end of 2020, corresponding to a 13% market share (S&P Global, 2021). Thereby, ~\$4T of the S&P 500 index market capitalization comes from passive institutional funds, mutual funds, and other index-replicating investment products, meaning the passive holdings of the S&P 500 is more likely to be 2.5x higher than the ETFs' holdings estimated in table 2. The following figure visualizes the estimated share of the S&P 500 Index owned by ETFs and the estimated share held by all passive investment instruments when keeping the relative distance of 2.5x from before constant.

Figure 5: ETF and Multiple Passive Instruments Estimated* Ownership Share of the S&P 500 Index

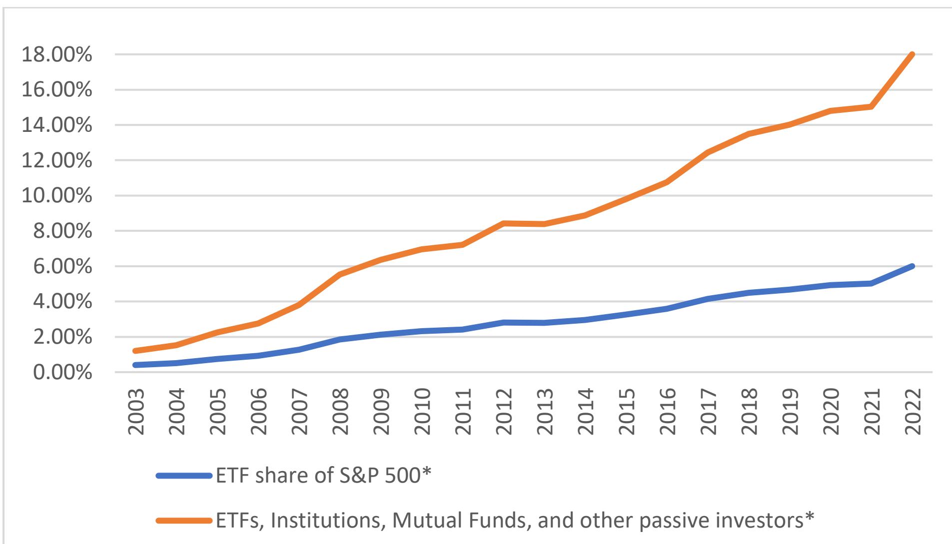


Figure 5: Visualizing the estimated share of S&P 500 mkt. cap placed in ETFs and in multiple passive investment products like ETFs, institutions, mutual funds, and other index-replicating products over time. Source: own analysis.

[2] ETFs, Institutional funds, Mutual Funds, and other investable products seeking to replicate the performance of the S&P 500 index.

5.2 Relationship Between Passive Index Funds and Market Volatility

The following figure shows the excess volatility of the 53 sample stocks added to the S&P 500 index during 2015-2017 on average relative to the index. The x-axis covers daily observations for a 10-year period, t±5 and the y-axis show the excess volatility relative to the volatility of the SPDR S&P 500 ETF Trust.

Figure 6: 10- and 30-Day Garman Klass Volatility Estimates

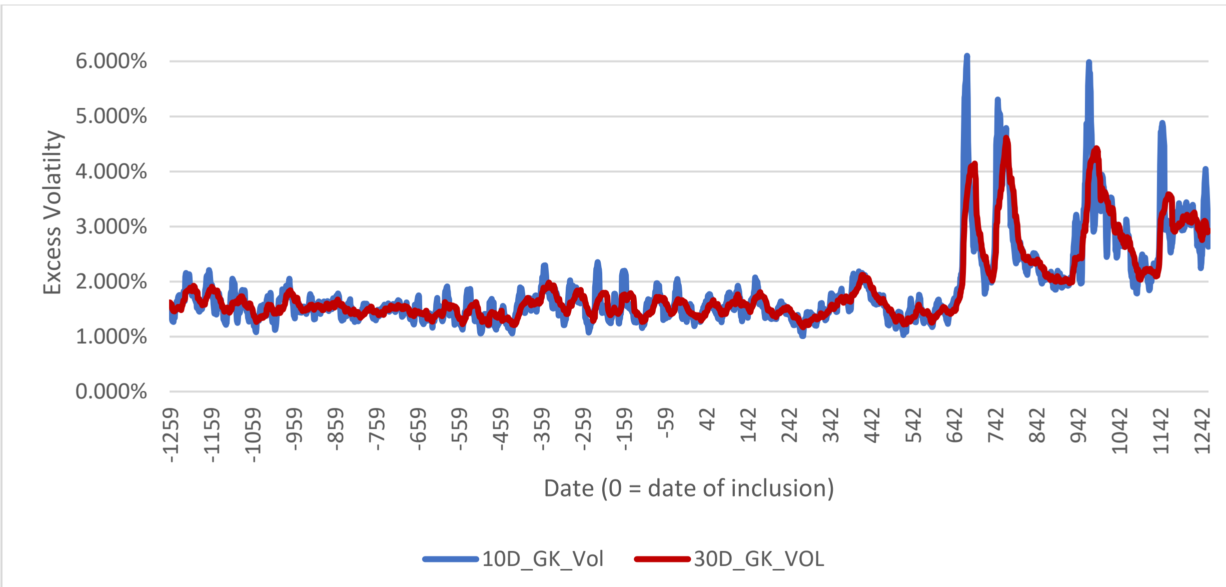


Figure 6: Showing the estimated 10-day and 30-day Garman Klass Volatility. Source: own analysis.

The graph indicates a stable trend until the 642nd day or about two and a half years after index inclusion, whereafter the excess volatility tends to go up.

The following table shows the results of multiple z-tests comparing the sample's 10-day Garman Klass volatility mean values for the periods t±1, t±3, and t±5. With p-values below 0.05 the tests conclude that the mean volatility is significantly higher in t+3 and t+5. Regarding the period t±1, the p-value is >0.05 whereby there is no statistical difference.

Table 3: 10-Day Garman Klass Volatility, Z-Test

Time periods	*t-1*	*t+1*	*t-3*	*t+3*	*t-5*	*t+5*
Mean (annualized)	5.618356	5.558276	5.548011	6.342743	5.61077	7.926277
Known Variance (annualized)	0.002687	0.001111	0.002154	0.02027	0.001956	0.031726
Observations	252	252	756	756	1259	1259
z	0.810042		-7.63794		-23.4321	
P(Z<=z) one-tail	0.208958		0		0	

Table 3: Showing output from z-tests comparing mean 10-day Garman Klass volatility for different periods. Source: own analysis.

The following table shows the results of multiple z-tests comparing the sample's 30-day Garman Klass volatility mean values for the periods t±1, t±3, and t±5. With p-values below 0.05 the tests conclude that the mean volatility is higher in period t+3 and t+5. Regarding the period t+1, the p-value is >0.05 whereby there is no statistical difference.

Table 4: 30-Day Garman Klass Volatility, Z-Test

Time periods	*t-1*	*t+1*	*t-3*	*t+3*	*t-5*	*t+5*
Mean (annualized)	5.660872	5.558768	5.547702	6.342743	5.617478	7.87353
Known Variance (annualized)	0.000748	0.004983	0.000932	0.011859	0.000859	0.025146
Observations	252	252	756	756	1259	1259
z	1.120662		-8.3786		-25.9829	
P(Z<=z) one-tail	0.131216		0		0	

Table 4: Showing output from z-tests comparing mean 30-day Garman Klass volatility for different periods. Source: own analysis.

The output of a Wilcoxon Signed Rank Test performed on 10-day and 30-day Garman Klass volatility data is presented in the next two tables. The results show that the median volatility values have changed across most groups, because most p-values are below 0.05. The only exception is regarding the health care group where both the t±1 for the 10-day and 30-day Garman Klass volatility have p-values above 0.05, whereby the difference in median values is insignificant.

Table 5: 10-Day Garman Klass Volatility, Wilcoxon Signed Rank Test

x-variable	y-variable	P-value
10dHC(t-5)	10dHC(t+5)	0.0000
10dHC(t-3)	10dHC(t+3)	0.0000
10dHC(t-1)	10dHC(t+1)	0.5511
10dRE(t-5)	10dRE(t+5)	0.0000
10dRE(t-3)	10dRE(t+3)	0.0037
10dRE(t-1)	10dRE(t+1)	0.0000
10dIT(t-5)	10dIT(t+5)	0.0000
10dIT(t-3)	10dIT(t+3)	0.0000
10dIT(t-1)	10dIT(t+1)	0.0000
10dR.g(t-5)	10dR.g(t+5)	0.0000
10dR.g(t-3)	10dR.g(t+3)	0.0000
10dR.g(t-1)	10dR.g(t+1)	0.0001

HC = Healthcare, RE = Real Estate, IT = Information Technology, R.g = Remaining

Table 5: Showing the output from Wilcoxon Signed Rank test testing whether median values of different sectors' 10-day Garman Klass volatility are significantly different. Source: own analysis.

Table 6: 30-Day Garman Klass Volatility, Wilcoxon Signed Rank Test

x-variable	y-variable	P-value
30dHC(t-5)	30dHC(t+5)	0.0000
30dHC(t-3)	30dHC(t+3)	0.0000
30dHC(t-1)	30dHC(t+1)	0.1059
30dRE(t-5)	30dRE(t+5)	0.0000
30dRE(t-3)	30dRE(t+3)	0.0002
30dRE(t-1)	30dRE(t+1)	0.0000
30dIT(t-5)	30dIT(t+5)	0.0000
30dIT(t-3)	30dIT(t+3)	0.0000
30dIT(t-1)	30dIT(t+1)	0.0000
30dR.g(t-5)	30dR.g(t+5)	0.0000
30dR.g(t-3)	30dR.g(t+3)	0.0000
30dR.g(t-1)	30dR.g(t+1)	0.0001

HC = Healthcare, RE = Real Estate, IT = Information Technology, R.g = Remaining

Table 6: Showing the output from Wilcoxon Signed Rank test testing whether the median values of different sectors' 30-day Garman Klass volatility are significantly different. Source: own analysis.

5.3 Relationship Between Passive Index Funds and Market Efficiency

5.3.1 Price Discovery

The figure below shows a trend towards an increasing spread between the price-to-sales multiple of the IShares Russel 2000 ETF and the SPDR S&P 500 ETF Trust. The space between the light and the dark yellow lines indicates an increasing spread between the small and large cap stocks in absolute terms, while the grey line shows the price-to-sales multiple of Russel 2000's as % of S&P 500 suggesting that the percentage spread is widening too.

Figure 7: Price-to-Sales Multiple of US Small and Large Cap

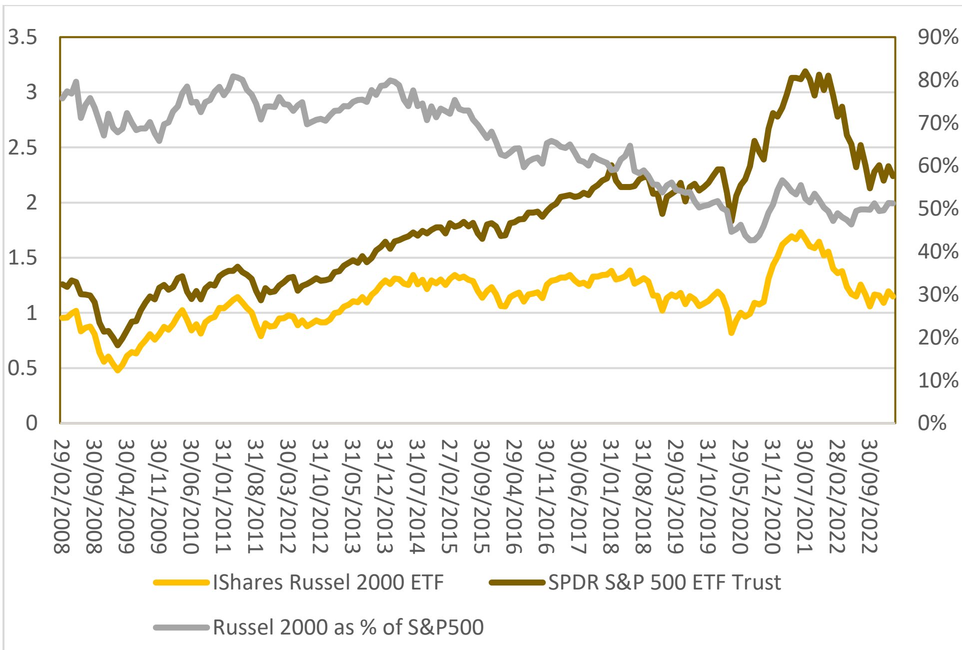


Figure 7: Showing the Price-to-Sales multiple and the relative difference between iShares Russel 2000 ETF and SPDR S&P 500 ETF Trust. Calculated with data from FactSet.

5.3.2 Repercussions of Stocks Added to the S&P 500 Index

5.3.2.1 Return Co-Movement

The following table shows descriptive statistics regarding the sample of stocks added to the S&P 500 index during 2015-2017. The statistics include the average daily return, standard deviation, variance, and correlation with the S&P 500 for the periods t±1, t±3, and t±5.

Table 7: Sample Stock Returns, Descriptive Statistics

Timeline (years)	t-1	t+1	t-3	t+3	t-5	t+5
Avg. Daily Return	0.0013	0.0006	0.0010	0.0005	0.0009	0.0007
Standard Deviation	0.0166	0.0160	0.0163	0.0183	0.0172	0.0211
Variance	0.0003	0.0003	0.0003	0.0004	0.0003	0.0005
Correlation w. index	0.4901	0.4606	0.5023	0.5140	0.5407	0.5464

Table 7: Showing descriptive statistics regarding sample stock returns and return correlation with the index for different periods before and after index inclusion. Calculated with data from FactSet.

Table 7 indicates that stocks tend to move differently during the periods t±1, t±3, and t±5. Regarding both the periods t±3 and t±5 it is shown that the average daily return tends to decrease, the standard deviation and variance tends to increase, while the correlation with the index tends to increase. The period t±1 has some different characteristics, where the average daily return tends to decrease, while the standard deviation, variance, and the correlation with index tends to decrease.

The next table show the output from a z-test performed on the return co-movement with the S&P 500 index:

Table 8: Return Co-movement, Z-Test

Time periods	*t-1*	*t+1*	*t-3*	*t+3*	*t-5*	*t+5*
Mean	0.490107	0.460561	0.502347	0.51405	0.540743	0.546376
Known Variance	0.00032	0.00029	0.0003	0.00038	0.00033	0.00048
Observations	106	106	106	106	106	106
z	12.31626		-4.62063		-2.03771	
P(Z<=z) one-tail	0		0		0.02079	

Table 8: Showing output from z-test comparing mean return correlation coefficients between sample stocks and the S&P 500 index. Source: own analysis.

The z-tests confirm that the return co-movement between the sample stocks and the S&P 500 changes upon index inclusion as all p-values are below 0.05. During the period t±1 the correlation coefficient decreases while it increases in both period t±3 and t±5.

5.3.2.2 Trading Volume

The table below sums up descriptive statistics of the sample stocks' volumes traded for the periods t±1, t±3, and t±5. The statistics include 1-, 3-, and 5-year average daily volume, standard deviation, variance, and correlation with the S&P 500 index before and after the inclusion to the index.

Table 9: Sample Stocks' Trading Volume, Descriptive Statistics

Timeline (years)	t-1	t+1	t-3	t+3	t-5	t+5
Avg. Daily volume	2,453,451.92	2,908,327.25	2,044,326.66	2,541,337.82	2,068,853.78	3,036,324.12
Standard deviation	2,778,144.23	1,680,268.68	2,058,116.02	2,840,467.18	1,852,680.33	2,028,276.39
Variance (millions)	27,315,833.18	27,916,218.10	16,334,069.42	32,465,902.13	13,033,412.87	8,395,143.19
Correlation w. index	0.1539	0.3705	0.1643	0.3636	0.1646	0.3850

Table 9: Showing descriptive statistics regarding sample stocks trading volume for three periods before and after index inclusion. Source: own analysis.

Table 9 indicates that traded volumes change during all three periods t±1, t±3, and t±5. For the period t±5, average daily volumes, standard deviation, and the correlation with the index tend to go up, while the variance decreases. Regarding t±3 it is shown that that all parameters; the average daily volume, the standard deviation and variance, and the correlation with the index tend to increase. For t±1 then the average daily volume, variance, and correlation with the index increases, while the standard deviation decreases.

The following two tables, 10 and 11, present z-tests comparing the mean values of the three periods t±1, t±3, and t±5. Table 10 compares the sample stocks' average daily trading volume. Table 11 compares the trading volume correlation between the sample stocks and the SPDR S&P 500 ETF Trust. All p-values are below 0.05 verifying the increase in both trading volume and the sample stocks trading correlation with the SPDR S&P 500 ETF Trust are of statistical significance.

Table 10: Average Daily Trading Volume, Z-Test

Time periods	*t-1*	*t+1*	*t-3*	*t+3*	*t-5*	*t+5*
Mean	2453452	2909107	2044938	3049014	2068854	3036307
Known Variance	4.61169E+12	5.57147E+11	1.70483E+12	6.46668E+11	1.08599E+12	6.37224E+11
Observations	252	252	756	756	1259	1259
z	-3.18		-18.00		-26.15	
P(Z<=z) one-tail	0		0		0	

Table 10: Showing output from z-tests comparing sample stocks' average daily trading volume for three periods before and after index inclusion. Source: own analysis.

Table 11: Trading Volume Correlation with Index, Z-Test

Time periods	*t-1*	*t+1*	*t-3*	*t+3*	*t-5*	*t+5*
Mean	0.153921	0.370477	0.164345	0.363623	0.16460123	0.384999
Known Variance	2.73E-13	2.79E-13	1.63E-13	3.25E-13	1.30334E-13	3.25E-13
Observations	53	53	53	53	53	53
z	-2121355		-2076765		-2378755.05	
P(Z<=z) one-tail	0		0		0	

Table 11: Showing output from z-tests comparing sample stocks' mean trading volume correlation with the trading volume of the SPDR S&P 500 ETF Trust for three periods before and after index inclusion. Source: own analysis.

5.3.2.3 Bid-Ask Spread

Figure 8 shows the relationship between the daily high and low price of the sample stocks for period t±5. T-5 corresponds to day -1259-0, and t+5 corresponds to day 0-1259. Lows are marked with an orange line and highs are marked with a blue line.

The relationship between the daily high and low price seems relatively stable until day ~600, or about 2.5 years after inclusion, where the gap widens, and the daily high-prices rise more relative to the low-prices.

Figure 8: Sample Stocks' Daily High and Low Prices

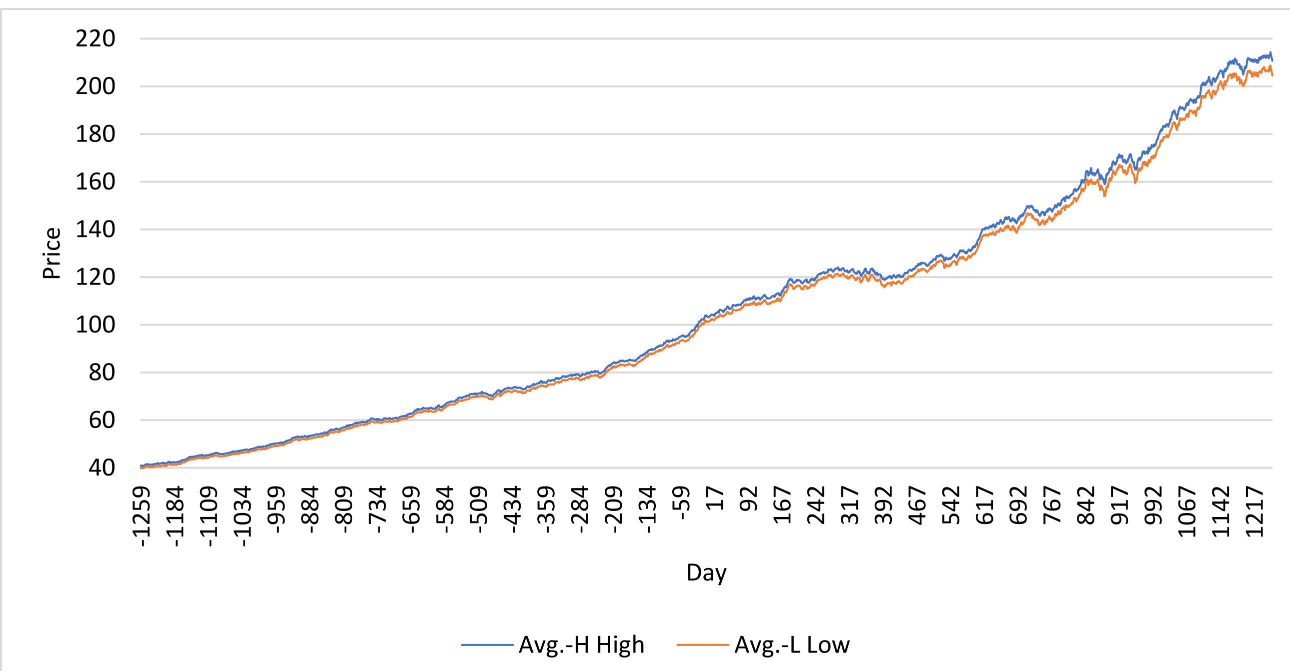


Figure 8: Showing the sample stocks' average daily high and low prices. Calculated with data from FactSet.

To provide a better view of the difference between daily high and low prices relative to the average share price, which according to figure 8, has been increasing, the next figure shows the percentage gap between the daily high and low prices. It confirms a more volatile and increasing gap from about day 600 and onwards.

Figure 9: The % Difference Between Daily High and Low Prices

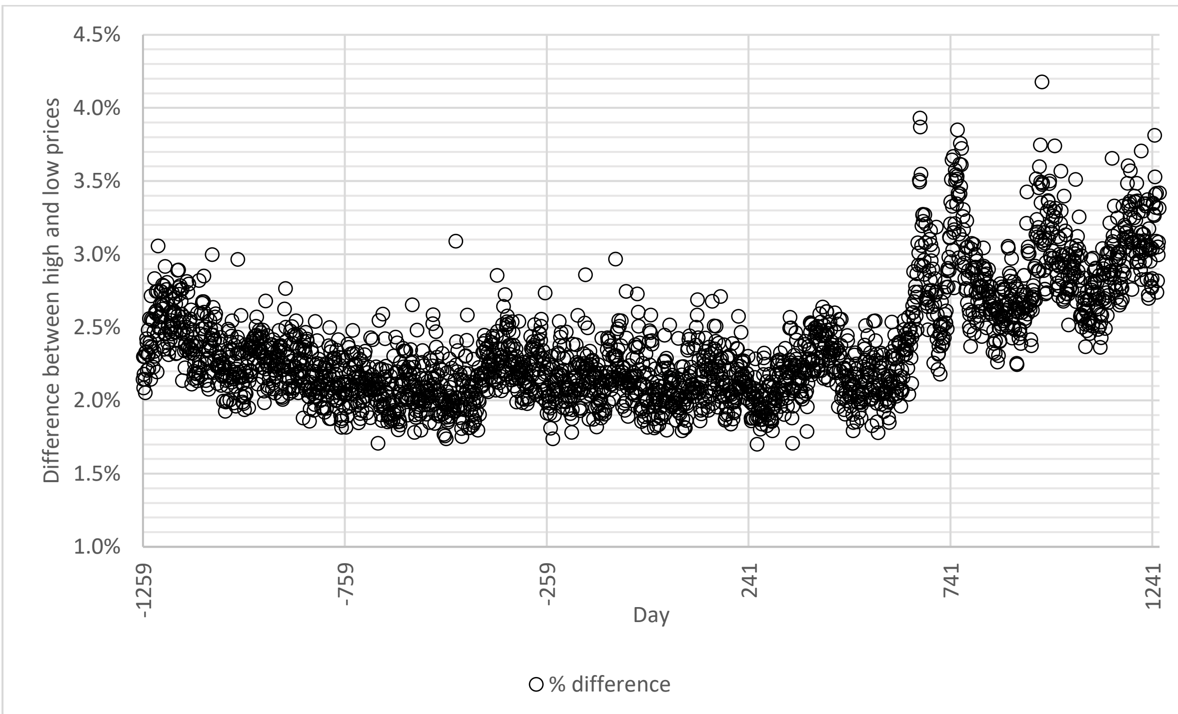


Figure 9: Showing the % difference between sample stocks' daily high and low prices. Source: own analysis.

The next figure displays the estimated daily bid-ask spread for the same period t±5 or 2518 days. It confirms the trend from the figure above, where it is obvious that the spread start to increase quite significantly from ~600 days after inclusion.

Figure 10: Bid-Ask Spreads

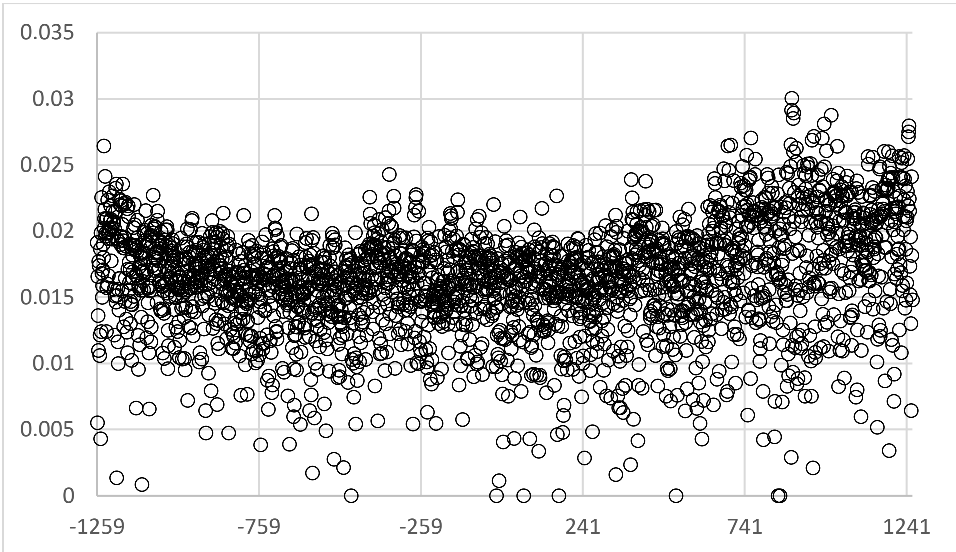


Figure 10: Showing a scatterplot of the estimated bid-ask spreads. Source: own analysis.

Table 12, containing multiple z-tests, shows that the mean bid-ask spread changes significantly during the periods t±3 and t±5 because the p-values are below 0.05. For these periods it is shown that the bid-ask spread widens and the variance increases. The period t±1 shows no statistically significant change in bid-ask spread.

Table 12: Bid-Ask Spread, Z-Test

Time periods	*t-1*	*t+1*	*t-3*	*t+3*	*t-5*	*t+5*
Mean	0.015368	0.014893	0.015300	0.015808	0.015727	0.016961
Known Variance	0.000012	0.000014	0.000013	0.000016	0.000013	0.000022
Observations	252	252	756	756	1258	1258
Z	1.485743		-2.599060		-7.410190	
P(Z<=z) one-tail	0.068674		0.004674		0.000000	

Table 12: Showing output from z-tests comparing sample stocks' average bid-ask spreads for three periods before and after index inclusion. Source: own analysis.

6. Discussion

This section contains a discussion of the results of the analysis. The results will be processed with the aim of knowing what causes the achieved results of the analysis. Finally, it will be discussed how these findings differ from the findings of existing research.

6.1 Passive Index Funds and Market Volatility

The analysis shows that the estimated 10-day and 30-day Garman Klass volatility increases in period t±3 and t±5. The change in volatility for the period t±1 is statistically insignificant. But when grouping the sample stocks into sectors, the Wilcoxon Signed Rank Test shows that the median values are different regarding all t±1, t±3, and t±5 periods for all sectors: *healthcare, real estate, information technology, and remaining*, with t±1 for the healthcare sector being the only exception. This suggests that the insignificance is derived by the sample stocks from the healthcare sector.

It can be observed in figure 11, that the excess Garman Klass volatility increases and tends to spike five times from about 600 days after inclusion. Because the excess Garman Klass volatility is calculated as the average of all 53 stocks volatility means these spikes could be due to a few outliers having a strong impact on the results, the sample stocks has been grouped into sectors. Then it is possible to observe whether the spikes a caused by just a few sample stocks.

The following two figures (11 and 12), show the excess volatility for stocks within health care, real estate, information technology, and remaining sectors marked with blue, red, green and purple respectively:

Figure 11: 10-Day Garman Klass Volatility Estimates

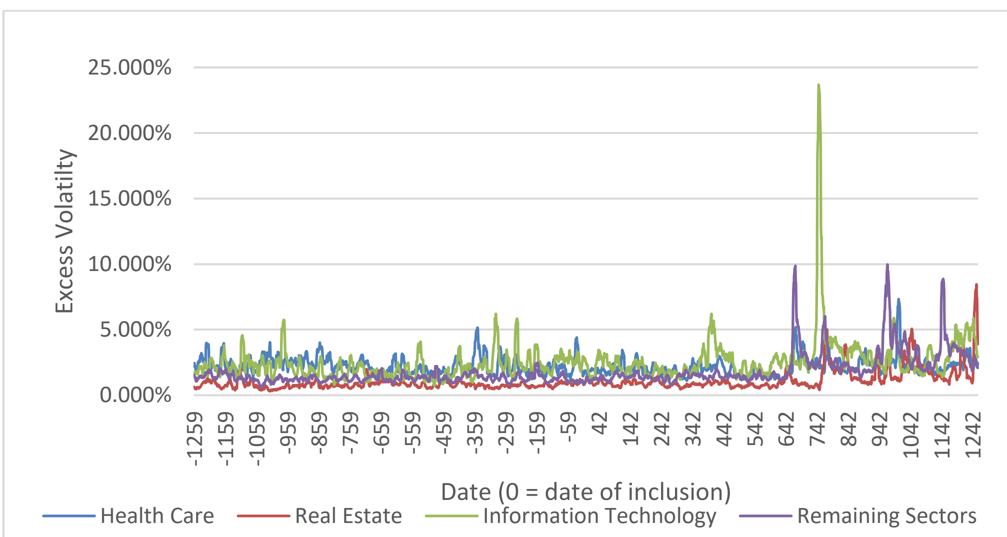


Figure 11: Showing 10-day Garman Klass volatility for the sectors: health care, real estate, information technology, and remaining sectors. Source: own analysis.

Figure 12: 30-Day Garman Klass Volatility Estimates

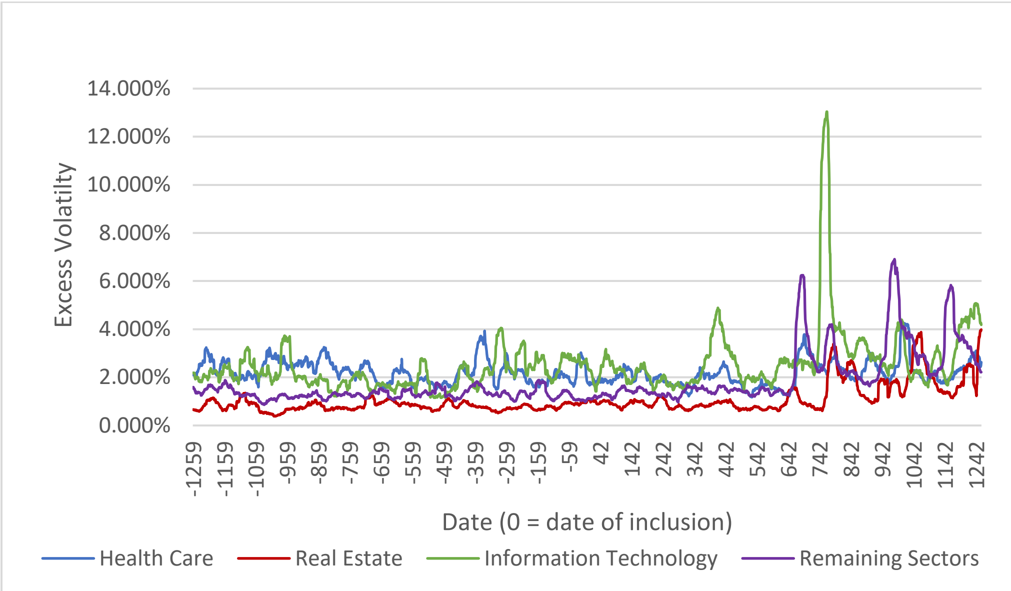


Figure 1: Showing 30-day Garman Klass volatility for the sectors: health care, real estate, information technology, and remaining sectors. Source: own analysis.

It is shown that the first, third, and fourth spike is mainly caused by the *remaining sector*, while the second and largest spike to a higher extent is carried out by the *information technology sector*. The fifth spike seems to be a result of stocks from both the *information technology* and *real estate sector*.

The spikes suggest that some exogenous events affect different industries at different times. It is in this regard relevant to pay attention to more details around the inclusion dates to gain more knowledge on what causes this proven change in volatility.

Table 13 presents a detailed overview of the sample stocks' sector distribution and inclusion dates. It is noteworthy that stocks within the *information technology* sector have been added to the S&P 500 index during the second quarter of both 2015, 2016, and the first and second quarter of 2017 and it still cause such a big spike in volatility after ~742 days when thinking about their varying inclusion dates.

Table 13: Sample Stocks' Sector distribution and Index Inclusion Dates

	2015				2016				2017				
GICS sector	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Total
Communication Services			1				1		1				3
Consumer Discretionary			1			2					1		4
Consumer Staples				1									1
Energy											1		1
Financials		1				1			2	1			5
Health Care	2		1		2	1	1		1	1	2		11
Industrials			2	1		2					1		6
Information Technology		1				1			2	4			8
Materials						2							2
Real Estate	1	1			3	1		1	2		1		10
Utilities					1	1							2
Total	3	3	5	2	6	11	2	1	8	6	6	0	53

Table 13: Showing the distribution of sample stocks' index inclusion dates and sectors. Source: own analysis.

Looking further into the details of the table it appears that six of a total of eight information technology sample stocks were added to the S&P 500 index during the first and second quarter of 2017, meaning that some exogenous event happening during the first half of 2020, would have an abnormal impact on stocks from the information technology sector.

The magnitude of the spikes caused by sample stocks from the other sectors, *real estate* and *remaining* sectors, are substantially smaller. The spike could still be because of some exogenous events, but the lower magnitude would in that regard be, because the event is 'less influential' or because the inclusion dates are more diversified. Another possibility is that some critical company specific event moves the entire sector, when *health care*, *real estate*, *information technology*, and *remaining* sectors account for just 11, 10, 8, and 24 stocks respectively. It appears that 37 of 53 sample stocks were added to the S&P 500 index during the five most crowded quarters within a seven-month time span from 2016 Q1 to 2017 Q3. Occasionally, the most volatile period, day 642-1160 spans about 8 months, meaning that one event potentially could trigger all the spikes at once because of the rolling inclusion dates.

On February 19th, 2020, the S&P 500 closed at a record high of 3,386.15 and 32 days later on March 23rd, the index had bottomed at a decline of 34% relative to the all-time high (Jason, 2020). On April 8th the relative decline was reduced to just 19%, but in between the two dates, six trading days were added to the list of the top 30 best and worst daily returns of the S&P 500 index. Three negative daily returns was ranging from -7.6% to -12% and three positive daily returns ranged from 7% to 9.4%, suggesting one of the most volatile periods in the history of the S&P 500 index (Jason, 2020; Statista, 2022).

So how do the volatility spikes in figures 11 and 12 compare with the sector specific inclusion dates presented in table 13? The following table 14, an extension of table 13, shows the index inclusion dates of the sample stocks and their sectors together with the time span to the COVID-outbreak in terms of quarters and trading days.

Table 14: Sample Stocks' Sector distribution and Index Inclusion Dates Relative to COVID Outbreak

	2015				2016				2017				
GICS sector	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Total
Communication Services			1				1		1				3
Consumer Discretionary			1			2					1		4
Consumer Staples				1									1
Energy											1		1
Financials		1				1			2	1			5
Health Care	2		1		2	1	1		1	1	2		11
Industrials			2	1		2					1		6
Information Technology		1				1			2	4			8
Materials						2							2
Real Estate	1	1			3	1		1	2		1		10
Utilities					1	1							2
Total	3	3	5	2	6	11	2	1	8	6	6	0	53

Q's till the COVID crash (2020 Q1)	20	19	18	17	16	15	14	13	12	11	10	9	
Trading days (63 trading days per Q)	1260	1197	1134	1071	1008	945	882	819	756	693	630	567	

Table 14: An extension of table 13, showing the sample stocks' sectors and inclusion dates together with the time span to the COVID-outbreak in both quarters and trading days. Source: own analysis.

The first volatility spike on figure 11 happens about 650 days after the inclusion date. This spike is driven by the *remaining* sectors. During the quarter 650 days prior to the COVID-outbreak, 2017 Q3, there are three of six index additions belonging to the *remaining* group. Hereby it seems like the first spike in volatility is triggered by the COVID-outbreak.

The second spike is nearly two and half times bigger than the first. It is mainly triggered by the *information technology* sector. The spike happened ~742 days after the inclusion of 6 information technology stocks. The sample has 8 information technology stocks in total, meaning the COVID-outbreak has a big impact on the *information technology* sector. Thereby it seems like the COVID-outbreak caused the second spike by having a major impact on the volatility of those six *information technology* additions.

The third and fourth spike are both caused by the *remaining* sector, about 950 and 1142 days after inclusion respectively. In the quarter containing the 950th day before the COVID-outbreak, 2016 Q2, there were 11 new additions to the S&P 500 of which eight stocks belong to the *remaining* sector. The quarter containing the 1142nd day before the COVID-outbreak, 2015 Q3, has five index inclusions of which four belong to the *remaining* sector. Hence volatility of sample stocks in the *remaining* sector are very sensitive to any consequences of the COVID-outbreak.

The fifth spike happening about 1245 days after inclusion was caused by the *real estate* and the *information technology* sector. The quarter 1245 days before the COVID-outbreak, 2015 Q1, has just three additions to the S&P 500 index of which one belongs to the *real estate* sector and one belongs to the *information technology* sector, suggesting this could be a single-stock event causing the volatility. The following table shows a fragment of the estimated rolling 10-day Garman Klass excess volatility for sample stocks fitting in the *real estate* sector.

Table 15: Subset of Estimated 10-Day Garman Klass Volatility for Real Estate Sample Stocks

Date	EQIX-US	O-US	EXR-US	FRT-US	UDR-US	DLR-US	MAA-US	REG-US	ARE-US	SBAC-US	Average
1244	1.2%	-0.7%	1.5%	2.0%	3.1%	0.8%	0.8%	0.9%	0.2%	1.2%	1.1%
1245	1.3%	11.8%	1.5%	1.9%	3.0%	0.7%	0.8%	0.7%	0.1%	0.8%	2.3%
1246	2.0%	11.8%	1.3%	2.1%	3.0%	0.7%	0.7%	0.7%	0.1%	0.9%	2.3%
1247	1.7%	34.0%	1.3%	2.4%	3.0%	0.9%	0.7%	0.8%	0.2%	0.9%	4.6%
1248	1.8%	49.9%	1.2%	2.5%	3.0%	1.0%	0.4%	1.9%	0.0%	0.8%	6.3%
1249	2.2%	51.9%	1.2%	2.2%	2.8%	0.9%	0.6%	2.1%	0.1%	0.8%	6.5%
1250	2.5%	56.4%	1.8%	2.3%	2.6%	0.9%	0.6%	2.2%	0.2%	0.7%	7.0%
1251	2.3%	60.3%	1.8%	2.3%	2.7%	0.9%	0.6%	2.2%	0.4%	0.8%	7.4%
1252	2.7%	65.8%	1.7%	2.0%	2.4%	0.9%	0.7%	2.2%	0.5%	0.9%	8.0%
1253	2.8%	66.3%	1.5%	2.1%	2.4%	1.0%	0.8%	2.3%	0.5%	0.8%	8.1%
1254	2.7%	66.8%	1.7%	2.2%	1.3%	1.0%	0.9%	2.2%	0.5%	0.9%	8.0%
1255	2.8%	71.3%	1.5%	2.7%	0.7%	0.9%	1.0%	2.6%	0.4%	0.6%	8.5%
1256	4.1%	60.2%	1.5%	2.9%	0.7%	1.0%	1.3%	2.6%	0.4%	0.7%	7.5%
1257	2.7%	60.8%	1.5%	2.6%	0.7%	0.8%	1.4%	2.4%	0.2%	0.7%	7.4%
1258	3.8%	39.6%	1.6%	2.5%	0.7%	0.5%	1.5%	2.4%	0.3%	0.6%	5.3%
1259	5.1%	24.6%	1.7%	2.5%	0.6%	0.5%	1.7%	1.2%	0.4%	0.6%	3.9%

Table 15: Showing a section of the estimated rolling 10-day Garman Klass excess volatility of sample stocks belonging to the real estate sector. Source: own analysis.

The table shows that the estimated excess volatility was significantly higher in the highlighted column regarding Realty Income ($O-US). Realty Income was added to the index on April 7th or very early in the second quarter of

2015. The fifth spike had a magnitude of about 8%, and on date 1255 after inclusion one of ten real estate stocks had an excess volatility of 71.3%. This means that seven percentage points of the sector's 8.5% spike was due to just one stock.

The same thing cannot be said about the *information technology* sample stock Qorvo Inc. ($QRVO-US), which was added to the S&P 500 index on June 11, 2015 (appendix 1). Table 16 shows a fragment of the estimated rolling 10-day Garman Klass excess volatility for sample stocks fitting in the *information technology* sector:

Table 16: Subset of Estimated 10-Day Garman Klass Volatility for Information Technology Sample Stocks

Date	QRVO-US	GPN-US	SNPS-US	AMD-US	DXC-US	IT-US	ANSS-US	CDNS-US	Average
1244	4.4%	1.1%	6.2%	16.0%	9.9%	3.2%	2.6%	0.3%	5.5%
1245	4.6%	0.9%	6.4%	16.0%	9.9%	3.0%	1.8%	0.4%	5.4%
1246	5.2%	0.9%	10.2%	15.8%	9.8%	3.2%	1.6%	0.4%	5.9%
1247	5.2%	0.9%	9.9%	12.7%	4.8%	2.9%	1.3%	0.3%	4.8%
1248	3.8%	0.9%	9.9%	12.9%	3.7%	2.8%	1.6%	0.3%	4.5%
1249	3.9%	0.9%	8.6%	13.1%	3.6%	2.5%	2.1%	0.4%	4.4%
1250	4.4%	0.8%	8.7%	13.5%	3.3%	1.9%	2.1%	0.5%	4.4%
1251	4.2%	0.8%	8.9%	14.1%	3.2%	1.7%	2.1%	0.5%	4.4%
1252	3.8%	0.6%	8.8%	15.4%	2.0%	1.3%	2.3%	0.6%	4.3%
1253	4.3%	0.7%	6.6%	15.4%	1.9%	1.1%	2.3%	0.6%	4.1%
1254	4.6%	0.6%	6.9%	14.1%	1.9%	1.1%	2.3%	0.5%	4.0%
1255	4.5%	0.4%	6.8%	10.2%	2.0%	0.9%	2.3%	0.6%	3.5%
1256	4.2%	0.4%	6.8%	10.0%	2.3%	1.1%	2.1%	0.3%	3.4%
1257	3.5%	0.4%	3.2%	11.3%	2.6%	0.9%	2.0%	0.4%	3.0%
1258	3.3%	0.4%	4.3%	9.6%	2.0%	1.0%	2.1%	0.4%	2.9%
1259	3.4%	0.4%	4.1%	9.3%	3.0%	0.8%	1.8%	0.5%	2.9%

Table 16: Showing a section of the estimated rolling 10.day Garman Klass excess volatility of sample stocks belonging to the information technology sector. Source: own analysis.

The table shows that the fifth spike in volatility for the *information technology* sector was not caused by one single stock. A few stocks have elevated volatility levels, meaning that some periods of stress might cross each other, because only the timeline for the first information technology sample stock inclusion date fits with the COVID-outbreak.

To sum it up, there seems to be evidence indicating that the first, second, third and fourth spike are largely a result of the COVID-outbreak. The fifth spike was partially caused by a single stock's extreme volatility during the COVID-outbreak. COVID-related spikes were mainly caused by the *remaining* and *information technology* sector. This aligns well with the findings of (Curto & Serrasquiro, 2022), who study how COVID-19 affected the return volatility of different sectors in the S&P 500. They find the highest volatility during the COVID-crisis to be in the in the sectors: information technology/telecom services, consumer discretionary, industrials, and consumer staples. These sectors can be translated into the *information technology* and *remaining* sectors used for analysis in this report.

Robustness check

Because the volatility spikes seem to be related to COVID, it will be analyzed how volatility changes in the period t±2. In this way, as much data as possible will be used for analysis with no without interfering with the COVID-crisis.

The next two tables show the output of a z-test comparing the mean 10- and 30-day Garman Klass volatility for the period t±2. The z-tests conclude no statistically significant difference in either 10- or 30-day Garman Klass volatility.

Table 17: 10-Day Garman Klass Volatility, Z-Test

Time periods	*t-2*	*t+2*
Mean (annualized)	0.015531553	0.015453574
Known Variance (annualized)	7.18477E-06	5.54006E-06
Observations	504	504
z	0.490756574	
P(Z<=z) one-tail	0.311799314	

Table 17: Showing output from a z-test comparing sample stocks' average 10-day Garman Klass volatility for the 2-year periods before and after index inclusion. Source: own analysis.

Table 18: 30-Day Garman Klass Volatility, Z-Test

Time periods	*t-2*	*t+2*
Mean (annualized)	0.015513384	0.01547067
Known Variance (annualized)	3.17228E-06	3.81537E-06
Observations	504	504
z	0.362756694	
P(Z<=z) one-tail	0.358393322	

Table 18: Showing output from a z-test comparing sample stocks' average 30-day Garman Klass volatility for the 2-year periods before and after index inclusion. Source: own analysis.

Diving further into the sample stocks' volatility it would be of interest to find out if this nonsignificant change is due to the biases of the z-test. A few sample stocks with a large decrease in volatility could offset the increase from more stocks with a smaller increase.

The following table presents descriptive statistics regarding the 10- and 30-day Garman Klass volatility for the period t±2:

Table 19: 10- and 30-Day Garman Klass Volatility, Descriptive Statistics

Garman Klass Volatility	10-day	30-day
All stocks average t-2 GK volatility	1.553%	1.551%
All stocks average t+2 GK volatility	1.545%	1.547%
Stocks with increased volatility	30	30
- sum of increase	12.24%	12.10%
- average increase	0.41%	0.40%
Stocks with decreased volatility	23	23
- sum of decrease	-12.65%	-12.33%
- average decrease	-0.55%	-0.54%

Table 19: Showing descriptive statistics regarding sample stocks 10- and 30-day Garman Klass volatility for the periods ±2 years. Source: own analysis.

It is shown in the top of the table that on average the 10- and 30-day Garman Klass volatility decreases, but the general decrease is caused by a minor number of stocks. 30 stocks report an increase in volatility, while just 23 stocks report a decrease. The average decrease in volatility is larger than the average increase. Therefore, the initial analysis concluding no statistically significant difference does not necessarily paint the right picture, because most sample stocks experience an increase in volatility upon index inclusion.

6.1.2 Summary of Volatility Findings

The analysis involved z-tests to examine the changes in 10- and 30-day Garman Klass volatility. The results indicate an increase in volatility during the periods t±3 and t±5. However, for the t±1 period, the z-test shows an insignificant change in mean volatility values. This finding contradicts the Wilcoxon Signed Rank test, which suggests a significant change in median values for all sectors in all periods except for the healthcare sector during the t±1 period.

To investigate what causes the volatility spikes occurring approximately 642 trading days after the index inclusion date, sector-specific calculations were performed for 10- and 30-day Garman Klass volatility. A pattern emerged between the inclusion dates of sample stocks and the sectors associated with each volatility spike, indicating that the COVID-outbreak played a significant role in all volatility spikes.

To ensure the reliability of the results and minimize interference from the COVID-outbreak, a robustness check was conducted by adjusting the time period to t±2. The z-test for this new period revealed no statistically significant change in volatility. However, to determine whether this test was influenced by outliers that had a greater impact on mean values, descriptive statistics were examined for the same t±2 period. These statistics showed that the majority of stocks (30) experienced an increase in both 10- and 30-day Garman Klass volatility, while 23 sample stocks witnessed a decrease. The average decrease among the 23 stocks exceeded the average increase among the 30 stocks, resulting in an offsetting effect that confirmed the bias observed in the previous z-test.

6.2 Passive index funds and market efficiency

Table 2 shows that in the period 2008 to 2022, the estimated share of S&P 500 stocks combined market cap laying in ETFs has increased +3x and the literature review do argue that this increase has an impact on market efficiency as stocks trading pattern in terms of increased return co-movement, increased trading volume, and a narrowing bid-ask spread change upon being added to the S&P 500 index.

6.2.1 Price Discovery

The price-to-sales multiple of Russel 2000 and S&P 500 stocks, representing US small and large cap, has grown significantly from 2008 to the end of 2022, but they haven’t grown the same. The spread between Russel 2000 and S&P 500 stocks’ price-to-sales multiples has become wider over time. Russel 2000 stocks’ price-to-sales multiple has grown from about 1 to 1.2 during the period 2008-2022, while S&P 500’s price-to-sales multiple has grown from about 1.25 to 2.25(figure 7). This development shows that either the markets have become less efficient and increasingly misprice large cap stock relative to small cap stocks over time or there is some intangible asset attached to S&P 500 stocks.

6.2.2 Return Co-Movement

In the analysis, it is shown that sample stocks’ average daily returns change after a they join the S&P 500 index, but even more importantly, when having a market efficiency point of view, their return co-movement change upon index inclusion. Table 7 shows that when stocks join the S&P 500 index, the return co-movement with the index increases for the periods t±3 and t±5. Oppositely, it shows a decrease in return co-movement for period t±1. This means that the systematic risk, which cannot be diversified away goes up during t±3 and t±5 and goes down during t±1. This decrease could be a result of inclusion-effects happening prior to the effective inclusion date. (Petajisto, 2011; Harris & Gurel, 1986; Lynch & Mendenhall, 1997; Beneish & Whaley, 1996; Schleifer, 1986) argue the inclusion-effects start when the new S&P 500 additions and deletions are announced.

After studying the index change announcements of the sample stocks from S&P, it is found that the sample stock inclusions tend to be announced 3.25 trading days before the effective day. The index change announcements ranged from happening one to six trading days prior to the event. Therefore, if 6 trading days before and after the inclusion date are excluded from the return co-movement correlation calculations for t±1, the correlation coefficient should be without the impact of the short-term sample stock volatility.

The following table shows mean co-movement between sample stocks and the SPDR S&P 500 ETF Trust for the adjusted period t±1*, when the six days before and after the index inclusion date is excluded.

Table 20: Co-Movement, Z-Test

Time periods	*t-1**	*t+1**
Mean	0.4944	0.4589
Known Variance	0.0153	0.0236
Observations	106	106
z	1.8512	
P(Z<=z) one-tail	0.0321	

*excluding six trading days prior/after index inclusion

Table 20: Showing the results of a z-test comparing the sample stocks average co-movement for the period t±1. Source: own analysis.*

The results of the z-test show that return co-movement for the periods t-1* increases and t+1* decreases compared with the initial return co-movement of t-1 and t+1. Thereby the gap between the return co-movement of t-1* and t+1* is wider than if those six days were not excluded from both periods. These results does not necessarily support the findings of (Petajisto, 2011; Harris & Gurel, 1986; Lynch & Mendenhall, 1997; Beneish & Whaley, 1996; Schleifer, 1986), who found index inclusion effects to start by the announcement of the index changes.

To ascertain if these findings is representative of all sample stocks or just a result of the z-test being biased by outliers, the following table with descriptive statistics for the periods t±1, t±3, and t±5 has been made:

Table 21: Co-Movement, Descriptive Statistics

Time Period	*t±1*	*t±3*	*t±5*
All stocks average (t-) co-movement	0.494371	0.502347	0.540743
All stocks average (t+) co-movement	0.458917	0.51405	0.546376
Stocks with increased co-movement	21	30	38
-sum of increase	4.747225	6.197	7.345783
- average increase	0.226058	0.206567	0.19331
Stocks with decreased co-movement	32	23	15
- sum of decrease	-8.50541	-4.95647	-6.7487
- average decrease	-0.26579	-0.2155	-0.44991

Table 21: Showing descriptive statistics of sample stocks co-movement coefficient with the SPDR S&P 500 ETF Trust for the three periods t±1, t±3, and t±5 years. Source: own analysis.

The table shows that the sample stocks' average co-movement decreased in period t±1 and increased in period t±3 and t±5. It further shows that the decrease fits with most sample stocks reporting a decreasing development in period t±1. In period t±3 and t±5 most sample stocks report an increasing development in return co-movement. Furthermore, it is shown that the impact on the mean from average increase is less than impact of the average decrease, whereby the mean values to some extend offset each other, supporting the suspicion that the test might be biased.

6.2.3 Trading Volume

It is concluded in the analysis that daily trading volume increases during period t±1, t±3, and t±5. During these periods, the average daily trading volume increased by 18.6%, 49.1%, and 46.8%[3] for the periods t±1-, t±3-, and t±5 respectively. The percentage change in trading volume is summarized in table 22:

Table 22: Percentage Increase in Trading Volume Upon Index Inclusion

	Percentage Increase
t±1	18.6%
t±3	49.1%
t±5	46.8%

Table 22: Showing the sample stocks' percentage increase in trading volume for the periods ±1, t±3, and t±5. Source: own analysis.

[3] Calculated by dividing average daily trading volume of the respective period after the index inclusion date with the equivalent period before and subtracting one from the result.

The correlation coefficient between the sample stocks' average daily trading volume and the average daily trading volume of the SPDR S&P 500 ETF Trust changes too. From the three periods prior to the index inclusion date to after, the trading volume correlation coefficient increased from 0.15 to 0.37, 0.16 to 0.36, and 0.15 to 0.38 for the periods t±5-, t±3-, and t±1 respectively. The trading volume correlation coefficients is summarized in table 23:

Table 23: Sample Stocks and SPDR S&P 500 ETF Trust Trading Volume Correlation Coefficient

Time Period	t-	t+
t±1 year	0.15	0.37
t±3 years	0.16	0.36
t±5 years	0.16	0.38

Table 23: Showing the sample stocks' trading volume correlation coefficient for the periods t±1, t±3, and t±5. Source: own analysis.

These findings are highly correlated with findings from previous studies. (Edmister, Graham, & Pirie, 1996) similarly find the daily trading volume to increase significantly during the year after a stock is added to the S&P 500 index. (Pruitt & Wei, 1989) find the volume increase upon index inclusion to be positively correlated with changes in institutional holdings. A study of the short-term index inclusion effects, shows that for an index addition the trading volume increases 89% and 29% on the day and week following the index change announcement relative to the past eight week average and that this effect grew during the period 1978 to 1983 (Harris & Gurel, 1986). If adjusting the analysis to fit similar time periods as of (Harris & Gurel, 1986), the average trading volume for the week following the average announcement, which averagely happened 3.25 trading days prior to inclusion, would be 281% higher than the average of the past eight weeks or 40 trading days. The following table shows a subset of the sample stocks' average trading volume around the inclusion date:

Table 24: Subset of Data Regarding Sample Stocks Average Daily Trading Volume

Trading day relative to the inclusion date	Average volume
-9	2816705
-8	2797617
-7	2168572
-6	2723757
-5	3390914
-4	4184910
-3	3659756
-2	4915427
-1	34828083
1	6515224
2	4220562
3	3541859
4	2925306
5	3059592
6	2698710
7	2587958
8	2438527
9	2587099
10	3928067

Highlighted period is **281%** higher than the previous 40 trading days average.

Table 24: Showing a subset of the sample stocks average daily trading volume. Source: own analysis.

A 281% increase is 9.7x higher than the effect found by (Harris & Gurel, 1986), suggesting that the level of institutional investing has surged. This notion is substantiated by a study from PwC concluding that an increasing share of pension funds' AUM is placed in passive index funds and this trend is forecasted to continue until 2025 (PwC, 2020). This information points towards the increase in stocks trading volume after index inclusion has grown over time somewhat explained by an increasing demand from passive investors, which previous analysis confirms as the popularity of passive investing has escalated.

Robustness check

The next table shows descriptive statistics regarding sample stocks trading volume for the periods t±1, t±2, t±3, and t±5. There is an increasing trend in the number of stocks with increasing trading volume as the time-period expands. But there seems to be a spike between t±2 and t±5. From t±2 and t+3, the number of sample stocks with an increasing trading volume surged from 33 to 45 and in t±5, the number of stocks with an increasing trading volume declined to 35.

Table 25: Trading Volume, Descriptive Statistics

Time Period	*t±1*	*t±2*	*t±3*	*t±5*
All stocks average (t-) trading volume	2,453,452	2,083,046	2,044,327	2,068,854
All stocks average (t+) trading volume	2,908,327	3,097,485	2,541,338	3,036,307
Stocks with increased trading volume	23	33	45	35
-sum of increase	29,086,285	57,446,656	28,566,341	57,390,943
- average increase	1,264,621	1,740,808	634,808	1,639,741
Stocks with decreased trading volume	30	20	8	18
- sum of decrease	-4,977,893	-3,681,403	-2,224,749	-6,115,912
- average decrease	-165,930	-184,070	-278,094	-339,773

Table 25: Showing descriptive statistics regarding sample stocks trading volume for the periods t±1, t±2, t3, and t±5. Source: own analysis.

Figure 13 shows the daily trading volume and the share price of the SPDR S&P 500 ETF Trust. It shows that volumes are especially elevated during the first and second quarters of 2020, whereafter they go back to pre-COVID levels. The two first quarters of 2020 affect the period t±3 for 20[4] sample stocks because of their index inclusion dates, which helps explain the elevated number of stocks with increased trading volume from t±2 to t±3. Because the trading volume in figure 13 seems to go back to normal levels, it makes sense that the number of stocks with increased trading volume during the periods ±2 and ±5 is more alike with 33 and 35 stocks respectively.

[4] 20 sample stocks are added to the S&P 500 index during 2017 and has 2020 Q1 as part of their ±3-year data (see table 13).

Figure 13: Price & Trading Volume of SPDR S&P 500 ETF Trust

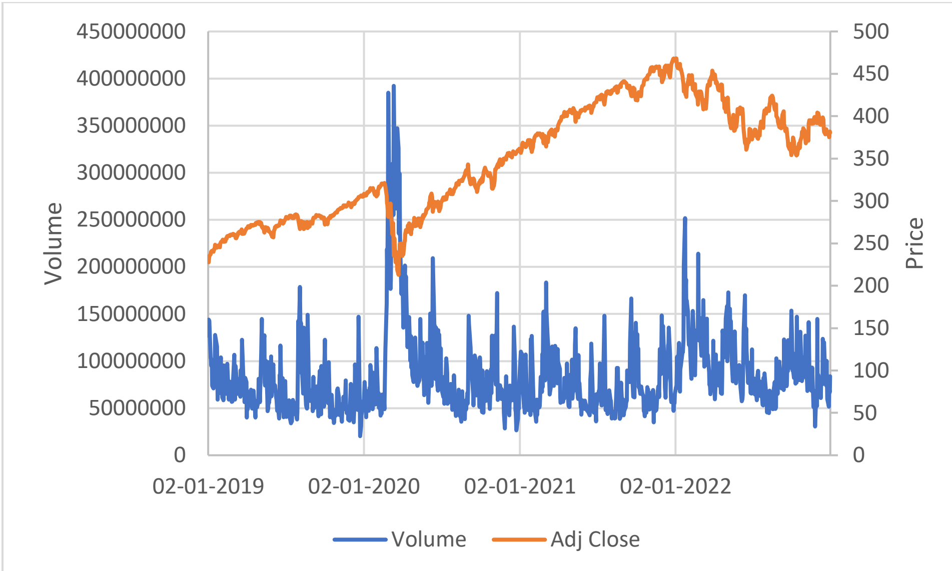


Figure 13: Showing the daily trading volume and adjusted close prices of the SPDR S&P 500 ETF Trust. Made with data from Yahoo Finance.

6.2.4 Bid-Ask Spread

The analysis of the bid-ask spread shows that the sample stocks' estimated bid-ask spread increases during the periods t±3 and t±5. In period t±1, there is no significant difference in the mean bid-ask spreads. Therefore, the analysis shows that trading costs increase during periods longer than one year. Both periods, t±3- and t±5, are affected by the COVID-outbreak in March 2020 as it happened about two and a half years, or about 630 trading days, after the newest sample stock inclusion date. Thereby, the data used to estimate the bid-ask spread of period t±5 is more affected by the COVID-crisis as it contains more data collected after the outbreak. This could be the reason why the spread for the period t+5 of 0.01691 is larger than the spread of period t+3 of 0.015808.

To adjust for the COVID-impact, similar calculations as in the analysis have been made, but instead of estimating the actual spread, the sample stocks excess bid-ask spread relative to the spread of the SPDR S&P 500 ETF Trust has been calculated. This is done by estimating the spread of all 53 sample stocks individually and subtracting the estimated spread from the SPDR S&P 500 ETF Trust as a proxy of the market. In this way, the results are adjusted for general increases in the bid-ask spread of the market.

Figure 14 shows a scatterplot of the daily excess bid-ask spreads for a 10-year period, t-5 and t+5.

Figure 14: Excess Daily Bid-Ask Spreads

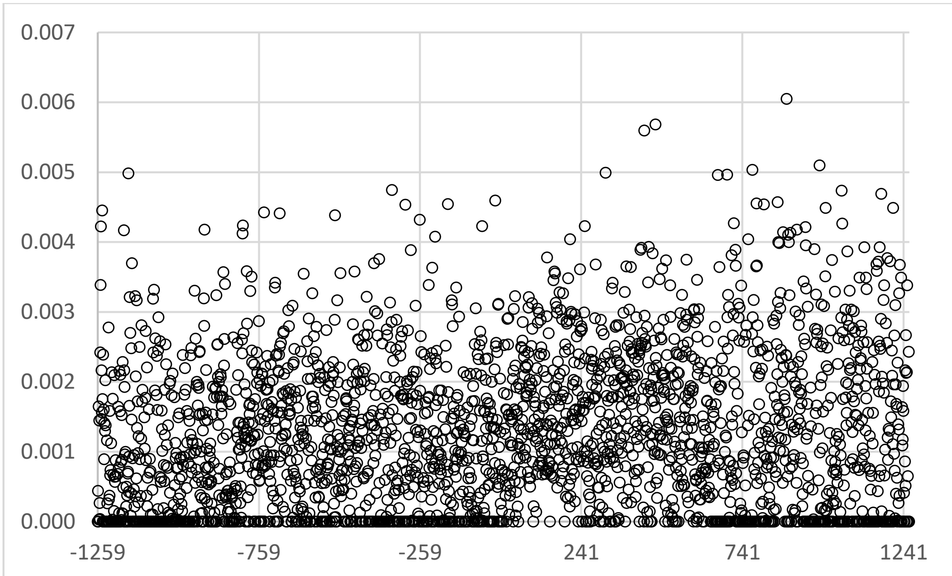


Figure 14: Showing a scatterplot of the sample stocks excess daily bid-ask spreads relative to the SPDR S&P 500 ETF Trust. Source: own analysis.

The upward trend in excess daily bid-ask spreads seems less clear than the upward trend in the raw bid-ask spreads estimates plotted in figure 10. The next table shows z-tests comparing the sample stocks' mean excess bid-ask spreads during the three periods t±1, t±3, and t±5:

Table 26: Excess Daily Bid-Ask Spread, Z-Test

Time periods	*t-1*	*t+1*	*t-3*	*t+3*	*t-5*	*t+5*
Mean	0.00103	0.00160	0.00112	0.00150	0.00106	0.00143
Known Variance	0.0000010	0.0000008	0.0000010	0.0000011	0.0000010	0.0000013
Observations	252	252	756	756	1259	1259
z	-6.79773		-7.23708		-8.61982	
P(Z<=z) one-tail	0		0		0	

Table 26: Showing the output of z-tests comparing the sample stocks' mean excess bid-ask spreads for three periods before and after the index inclusion date.

The table shows that the sample stocks' excess daily bid-ask spread relative to that of the SPDR S&P 500 ETF Trust increases in all three z-tests, because all p-values are below 0.05. But looking at the periods after the inclusion dates, the estimated spreads now have a decreasing trend the longer the time horizon.

To figure out whether the periodical increases in the spread are due to specific sectors or because of any COVID impact, z-tests comparing the same three time periods plus an additional ±2-year period has been made for the sectors: *health care, real estate, information technology, and remaining.* The results of the z-tests are presented in the following table. The green color highlights the p-values below 0.05, whereby the change is significant. The red color highlights the mean excess daily bid-ask spread with an increasing development.

Table 27: Sample Stocks Sector Distributed Excess Bid-Ask Spread

Sectors:	Health care		Real estate		Inf. technology		Remaining	
Period: t±5	**(t-)**	**(t+)**	**(t-)**	**(t+)**	**(t-)**	**(t+)**	**(-)**	**(t+)**
Mean	0.0034	0.0036	0.0039	0.0039	0.0025	0.0031	0.0029	0.0035
Known Variance	3.2E-06	3.5E-06	4.6E-06	4.4E-06	2.7E-06	4.5E-06	1.1E-06	2.0E-06
Observations	1259	1259	1259	1259	1259	1259	1259	1259
z	3.09282		0.416637		-8.144		11.7856	
P(Z<=z) one-tail	0.00099		0.338472		0		0	
Period: t±3	**(t-)**	**(t+)**	**(t-)**	**(t+)**	**(t-)**	**(t+)**	**(t-)**	**(t+)**
Mean	0.0035	0.0036	0.0040	0.0039	0.0024	0.0029	0.0029	0.0031
Known Variance	3.0E-06	3.2E-06	4.7E-06	4.0E-06	2.5E-06	3.7E-06	1.1E-06	1.5E-06
Observations	756	756	756	756	756	756	756	756
z	0.54907		0.926707		4.59955		3.78735	
P(Z<=z) one-tail	0.29148		0.177039		0		0	
Period: t±2	**(t-)**	**(t+)**	**(t-)**	**(t+)**	**(t-)**	**(t+)**	**(t-)**	**(t+)**
Mean	0.0035	0.0035	0.0041	0.0039	0.0023	0.0029	0.0029	0.0030
Known Variance	3.2E-06	2.8E-06	5E-06	4.1E-06	2.3E-06	3.1E-06	1.2E-06	1.2E-06
Observations	504	504	504	504	504	504	504	504
z	-0.1953		1.411322		5.02024		1.17842	
P(Z<=z) one-tail	0.42258		0.079075		0		0.119315	
Period: t±1	**(t-)**	**(t+)**	**(t-)**	**(t+)**	**(t-)**	**(t+)**	**(t-)**	**(t+)**
Mean	0.0034	0.0034	0.0040	0.0038	0.0022	0.0028	0.0030	0.0031
Known Variance	2.9E-06	2.3E-06	4.8E-06	3.5E-06	2.0E-06	3.1E-06	1.1E-06	1.2E-06
Observations	252	252	252	252	252	252	252	252
z	-0.33215		1.213345		-4.2545		-1.69943	
P(Z<=z) one-tail	0.369889		0.112499		0		0.044619	

Table 27: Showing output from z-tests comparing the sample stocks' sector specific excess bid-ask spread for the periods t±1, t±2, t±3, and t±5. Source: own analysis.

The table shows that half of the z-tests provide p-values below 0.05, meaning there is no significant changes in the mean excess daily bid-ask spreads for the other half of the analyzed periods and sectors. It seems like the statistically significant changes in excess bid-ask spreads are concentrated around the sectors *information* *technology* and *remaining*. All four periods regarding information technology sample stocks have an increasing excess daily bid-ask spread while three of four periods regarding the remaining sectors similarly provide a significant increase in the mean excess daily bid-ask spread. No sectors report a significant decreasing excess daily bid-ask spread in any of the analyzed time intervals. A potential COVID-impact on the excess daily bid-ask spread cannot be left out, since 5 boxes in the table above prove a statistically significant change in the COVID-affected periods ±3 and ±5 years, when just 3 boxes demonstrate a change in the periods t±1 and t±2, which was not affected by COVID.

To get a sense of whether these statistical and non-statistical differences in mean values are caused by outliers having a meaningful impact on the mean, it could be of interest to look into some descriptive statistics presented by table 28:

Table 28: Excess Daily Bid-Ask Spread, Descriptive Statistics

Time Period	*t±1*	*t±2*	*t±3*	t±5
All stocks average (t-) bid-ask spread	0.00319	0.00324	0.00322	0.00323
All stocks average (t+) bid-ask spread	0.00334	0.00333	0.00339	0.00364
Stocks with increased bid-ask spread	32	31	36	43
-sum of increase	0.01477	0.01888	0.01888	0.03043
- average increase	0.00046	0.00061	0.00052	0.00071
Stocks with decreased bid-ask spread	21	22	17	10
- sum of decrease	-0.01314	-0.00976	-0.01015	-0.00868
- average decrease	-0.00063	-0.00044	-0.00060	-0.00087

Table 28: Showing descriptive statistics regarding sample excess bid-ask spread for the periods t±1, t±2, t±3, and t±5. Source: own analysis.

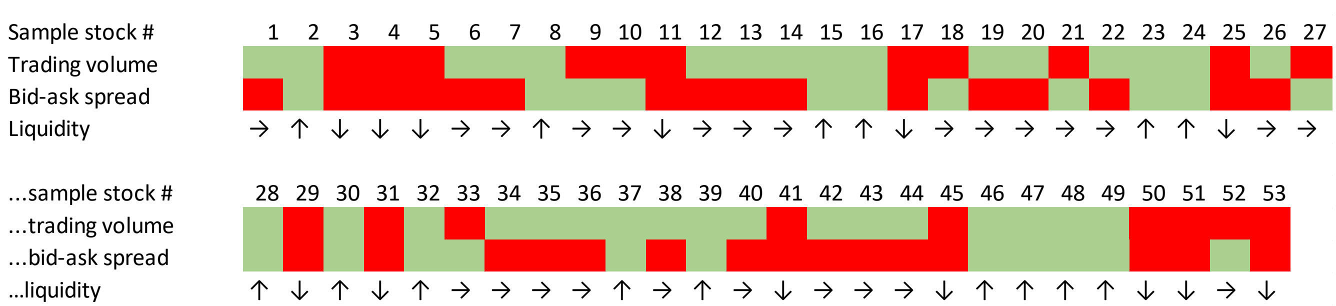
The table shows that among all time periods most sample stocks have an increasing excess daily bid-ask spread upon inclusion to the S&P 500 index. In the periods t±1 and t±2, the number of stocks issuing a decreasing excess daily bid-ask spread is 21 and 22 respectively. This is more than twice the number of stocks with a decreasing excess daily bid-ask spread in the period t±5, suggesting trading costs increase over time. The development suggests some COVID-impact as the number of stocks with a narrowing excess daily bid-ask spread declines quite heavily during t±3 and t±5, which occasionally is when the sample stock data starts to be affected by the COVID-outbreak.

6.2.5 Liquidity

As mentioned in the literature review, trading volume and bid-ask spread can be used as a gauge of liquidity. This report's general takeaways of the analysis and discussion of these subjects suggest a greatly increasing trading volume and mixed results regarding the daily bid-ask spread upon index inclusion. A higher trading volume indicates an increased liquidity and an increasing bid-ask spread indicates the opposite (Febrian, 2008). Descriptive statistics show that the majority of stocks have an increasing trading volume during the periods t±2, t±3, and t±5 (table 25). The excess daily bid-ask spread is similarly increasing for most stocks (table28), which has the opposite impact on the liquidity measure.

The following table provides an overview of the development of the 53 sample stocks' trading volume and excess daily bid-ask spread and the impact on the stock's liquidity for the period t±2. The sample stocks are arranged on a timeline, so sample stock number 1 was added to the S&P 500 index early 2015, while sample stock number 53 was added late 2017. The green color means the development in trading volume and excess daily bid-ask spread upon inclusion moves towards an increased liquidity and opposite for the red color. The two variables' combined impact on liquidity is expressed with arrows pointing in the direction, up for an increase, down for a decrease, and flat for in between.

Figure 15: The Impact of Sample Stocks Change in Volume and Bid-Ask Spread on Liquidity



	#
Liquidity ↑	15
Liquidity →	25
Liquidity ↓	13

Figure 15: Showing how the change in trading volume and bid-ask spread during the ±2-year period relative to index inclusion impacts the liquidity of the 53 sample stocks. Source: own analysis.

It is shown that just 15 stocks have both variables pointing towards an increased liquidity upon index inclusion when analyzing the period t±2. Oppositely, 13 sample stocks have both variables indicating a decrease in liquidity, while 25 sample stocks have contradicting results. The individual inclusion dates do not seem to have any meaningful influence on the results, because the red and green colors are relatively mixed from sample stock 1 to 53.

6.2.6 Summary of Market Efficiency Findings

The analysis of price discovery reveals significant changes in the price-to-sales multiple of Russel 2000 and S&P 500 stocks from 2008 to 2022. During this period, the Russel 2000 price-to-sales multiple increased from 1 to 1.2, while the S&P 500 price-to-sales multiple rose from 1.25 to 2.25. As a result, the gap between the two indices widened, indicating a potential decrease in market efficiency. This suggests that large-cap stocks may be increasingly mispriced compared to small-cap stocks over time, unless there are intangible assets attached to S&P 500 stocks. Occasionally, the share of ETF holdings in the S&P 500 index tripled during the same period, 2008-2022.

Examining the return co-movement, the analysis demonstrates that the return co-movement increases in the periods t±3 and t±5, while it decreases for the period t±1. It is worth noting that the literature argues the index inclusion-effects start with the announcement rather than the effective inclusion date. To provide further insight adjusting for the inclusion effects happening before the effective day, a z-test is performed on the period t±1, excluding the six trading days before and after the effective date, which results in a broader change in return co-movement. To ensure the robustness of the findings, descriptive statistics are calculated for the 53 sample stocks during the periods t±1, t±3, and t±5. The statistics reveal that as the time period lengthens, more sample stocks exhibit an increasing co-movement.

In terms of trading volume, the analysis shows a significant increase in mean trading volume upon index inclusion, particularly in the week following the announcement date. This surge in volume surpasses previous studies and limits the diversification benefits for investors. Descriptive statistics further highlight that, over time, more sample stocks exhibit an increasing trading volume upon inclusion. However, there is a notable spike in the number of stocks with increased trading volume for the period t±3, likely attributed to the COVID-outbreak in March 2020.

The analysis of bid-ask spread indicates an increasing spread for the periods t±3 and t±5, whereas the change is insignificant for the period t±1. To account for the impact of COVID-19 and general bid-ask spread increases, the excess daily bid-ask spread relative to the spread of the SPDR S&P 500 ETF Trust is calculated for the same periods. Interestingly, all three periods demonstrate an increasing excess bid-ask spread. Further analysis using z-tests on the periods t±1, t±2, t±3, and t±5 for the four sectors indicates that the increasing mean excess daily bid-ask spreads primarily stem from the *information technology* sector and *remaining* sectors. Moreover, descriptive statistics reveal that the majority of sample stocks report an increasing excess daily bid-ask spread, with a greater number of stocks exhibiting this trend as the time period extends.

The assessment of liquidity using trading volume and excess daily bid-ask spread as gauges presents mixed results. Both variables suggest an increase in liquidity for 15 sample stocks, while indicating a decrease for 13 sample stocks. The variables present contradicting results for 25 sample stocks, leaving their liquidity impact uncertain.

7. Conclusion

The analysis reveals a clear trend of passive index funds, especially ETFs covering the S&P 500 index, gaining market share during the past couple of decades. From 2003-2022, the market cap of the S&P 500 has experienced significant growth, more than tripling in value. Concurrently, the assets under management of global ETFs have increased by over 47 times and even more for those covering the S&P 500 explicitly. This development of passive investing has had a meaningful impact on the market volatility and efficiency.

The analysis reveals a relationship between the rise of passive investments and market volatility. The sample stocks contain stocks added to the S&P 500 index during the period 2015-2017. The study examines the excess volatility of these stocks before and after their inclusion in the index. The findings indicate a noticeable increase in both the 10- and 30-day mean volatility for the periods t±3 and t±5, which is highly connected to a series of correlations between the methodology regarding data collection, the sectors experiencing a substantial increase in volatility and the COVID-outbreak. However, there is no statistically significant difference in mean volatility during the periods t±1 or t±2. Further analysis of the sample stocks based on their sector classification reveals that the sample stocks' mean values might not be the best measure as the median values suggest a significant change in volatility regarding all sectors and most time intervals. Inspecting, the sample stocks individually then 30 and 23 stocks experienced an increase and decrease in volatility respectively during the period t±2.

The analysis of the relationship between passive investing and market efficiency reveals increasing inefficiencies. Firstly, the price discovery relationship between American small and large cap valuations shows an increasing spread, suggesting the market potentially misprices large cap stocks relative to small cap stocks. Secondly, index-inclusion effects provide mixed results from an efficiency point of view. The sample stocks return co-movement with fellow S&P 500 constituents increases over time reducing the diversification benefits. Additionally, the trading volume increases for the majority of the sample stocks in the periods t±2, t±3, and t±5. The excess daily bid-ask spread seems to increase for most stocks and the longer the time periods the more sample stocks experience an increase. With trading volume and bid-ask spread as a measure of liquidity it is concluded that only 15 of the 53 sample stocks experience an increase in liquidity during the period t±2.

To sum it up, most stocks experience an increase in volatility even during the non-COVID-affected time periods. In the same period return co-movement and trading volume goes up for most sample stocks, while the estimated excess daily bid-ask spread widens, whereby only a limited number of stocks document an increased liquidity.

8. Bibliography

Anadu, K., Kruttli, M., McCabe, P., & Osambela, E. (2020, May 15). *The Shift from Active to Passive Investing: Potential Risks to*. Retrieved April 28, 2023, from www.federalreserve.gov: https://www.federalreserve.gov/econres/feds/files/2018060pap.pdf

Barberis, N., Schleifer, A., & Wurgler, J. (2005). Comovement. *Journal of Financial Economics*, 283-317. Retrieved May 3, 2023, from https://scholar.harvard.edu/shleifer/publications/comovement

Ben-David, I., Franzoni, F., & Moussawi, R. (2018). Do ETFs Increase Volatility? *The Journal of Finance*, 2471-2535. Retrieved May 2, 2023, from https://onlinelibrary.wiley.com/doi/10.1111/jofi.12727

Beneish, M. D., & Whaley, R. E. (1996). An Anatomy of the "S&P Game": The Effects of Changing the Rules. *The Journal of Finance*, 1909-1930. Retrieved May 2, 2023, from https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1996.tb05231.x

Beneish, M. D., & Whaley, R. E. (2002). S&P 500 Index Replacements. *The Journal of Portfolio Management*, 51-60. Retrieved May 2, 2023, from https://www.pm-research.com/content/iijpormgmt/29/1/51

Bianco Research. (2018, August 20). *A Brief History of Equity Mutual Funds*. Retrieved May 12, 2023, from www.biancoresearch.com: https://www.biancoresearch.com/a-brief-history-of-equity-mutual-funds-2/

BlackRock. (2017). *Index Investing Supports Vibrant Capital Markets.* Retrieved from https://www.blackrock.com/corporate/literature/whitepaper/viewpoint-index-investing-supports-vibrant-capital-markets-oct-2017.pdf

BlackRock. (n.d.). *Different Types of ETFs*. Retrieved May 5, 2023, from www.blackrock.com: https://www.blackrock.com/sg/en/etfs-simplified/types-of-etfs

Bogusslavsky, V., & Muravyev, D. (2022, October). Who Trades at the Close? Implications for Price Discovery and Liquidity. 59. Retrieved May 2, 2023, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3485840

Bolla, L., Kohler, A., & Wittig, H. (2016). Index-Linked Investing-A Curse for the Stability of Financial Markets around the Globe? *The Journal of Portfolio Management*, 26-43. Retrieved May 3, 2023, from https://www.semanticscholar.org/paper/Index-Linked-Investing%E2%80%94A-Curse-for-the-Stability-of-Bolla-Kohler/aff2d787edd82680c81845ee7ffa2d021a937c68

Brogaard, J., Heath, D., & Huang, D. (2019, November). The Rise of ETF Trading and the Bifurcation of Liquidity. 45. Retrieved May 2, 2023, from https://iextrading.com/docs/heath_etf-liquidity-bifurcation.pdf

Carneiro, L. M., Junior, W. E., & Yoshinaga, C. E. (2022). The implications of passive investments for active fund management: International evidence. *Global Finance Journal*, 9. Retrieved March 8, 2023, from https://www.sciencedirect.com/science/article/pii/S1044028321000211

CFI. (2023, May 15). *Bid and Ask*. Retrieved from www.corporatefinanceinstitute.com: https://corporatefinanceinstitute.com/resources/equities/bid-and-ask/

CFI. (2023, March 22). *Volume vs Liquidity*. Retrieved May 22, 2023, from www.corporatefinanceinstitute.com: https://corporatefinanceinstitute.com/resources/capital-markets/volume-vs-liquidity/

Claessens, S., & Yafeh, Y. (2011, March 1). *Additions to Market Indices and the Comovement of Stock Returnds Around the World.* Retrieved May 11, 2023, from www.imf.org: https://www.imf.org/en/Publications/WP/Issues/2016/12/31/Additions-to-Market-Indices-and-the-Comovement-of-Stock-Returns-Around-the-World-24688

Clifford, C. P., Fulkerson, J. A., Jordan, B. D., & Waldman, S. R. (2014). Risk and Fund Flows. *The Financial Review*, 43-62. Retrieved April 28 2023, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1752362

Collins Advisors. (n.d.). *The History of the Mutual Fund, the Emergence of the ETF, and Picking the Right for You*. Retrieved May 12, 2023, from www.mlcollins.com: https://mlcollins.com/the-history-of-the-mutual-fund/

Corwin, S. A., & Schultz, P. (2012). A Simple Way to Estimate Bid-Ask Spreads from Daily High and Low Prices. *The Journal of Finance*, 719. Retrieved March 8, 2023, from https://deliverypdf.ssrn.com/delivery.php?ID=601116105008118123105018071071074105042086062072079074028086121029086079031107119120107060033115123035020097127029097014079016043086006050076119020087065073124017061086091112123064027007072117027028104071121010

Curto, J. D., & Serrasquiro, P. (2022). The impact of COVID-19 on S&P 500 sector indices and FATANG stocks volatility: An expanded APARCH model. *Finance Research Letters*. Retrieved May 16, 2023, from https://reader.elsevier.com/reader/sd/pii/S1544612321003020?token=791E712DCA7E7DF391FC1BD486EE2E1642F4E1DC2697C00D88E92F97A66924DBDF7F9301D035D6F1C8BA712C5EC05964&originRegion=eu-west-1&originCreation=20230516082709

Da, Z., & Shive, S. (2016). Exchange Traded Funds and Asset Return Calculations. 56. Retrieved May 11, 2023, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2158361

Deville, L. (2008). Exchange Traded Funds: History, Trading; and Research. *Handbook of Financial Engineering*, 67-98. Retrieved May 12, 2023, from https://link.springer.com/content/pdf/10.1007/978-0-387-76682-9_4.pdf

Edmister, R. O., Graham, A. S., & Pirie, W. L. (1996). Trading Cost Expectations: Evidence from S&P 500 Index Replacement Stock Announcements. *Journal of Economics and Finance*, 75-85. Retrieved May 9, 2023, from https://link.springer.com/article/10.1007/BF02920893

Edmister, R. O., Graham, A. S., & Pirie, W. L. (n.d.). Trading Cost Expectations: Evidence from S&P 500 Index Replacement Stock Announcements. *Journal of Economics and Finance*.

Erwin, G. R., & Miller, J. M. (2005). The liquidity effects associated with addition of a stock to the S&P 500 index: evidence from bid/ask spreads. *The Financial Review*, 131-146. Retrieved May 9, 2023, from https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-

6288.1998.tb01612.x?casa_token=0UVt2AvuL7YAAAAA:KebF5aAXbQQn50CEkD25LqA5Uqz2IDAiRQFsk P1HDQz_SuUQQ3BVVPvzQvp4UXF_1MOO2PjGW2u31aY

Fama, E. F. (1970). Efficient Capital Markets: A Review Of Theory and Empirical Work. *The Journal of Finance*, 383. Retrieved from https://www.jstor.org/stable/2325486?origin=crossref

Febrian, E. (2008). Liquidity Measurement Based on Bid-Ask Spread, Trading Frequenzy, and Liquidity Ratio: The Use of GARCH Model on Jakarta Stock Exchange (JSX). *Economic Journal*, 14. Retrieved May 23, 2023, from https://deliverypdf.ssrn.com/delivery.php?ID=727098123007093064090076112127018072052025093049028051091114073078022020070110078018029005000115006016107030023096093017101077019007013060013097121066024021000075021013085074000102089065105085023095073007107011

Goetzmann, W. N., & Massa, M. (2003). Index Funds and Stock Market Growth. *The Journal of Business*, 1-28. Retrieved April 28, 2023, from https://www.jstor.org/stable/pdf/10.1086/344111.pdf?refreqid=excelsior%3A7d2d9316c93608447a5fd5c3e322e8ae&ab_segments=&origin=&initiator=&acceptTC=1

Graves, J., & Pratt, K. (2022, September 08-03-2023). *Active vs Passive: Which Investing Approach is Right For You?* Retrieved from www.forbes.com: https://www.forbes.com/uk/advisor/investing/passive-investing-vs-active-investing/

Green, R. C., & Berk, J. (2004). Mutual Fund Flows and Performance in Rational Markets. *Journal of Political Economy*, 1269-1295. Retrieved April 28, 2023, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=612465

Grégoire, V. (2020). The rise of passive investing an index-linked comovement. *North American Journal of Economics and Finance*, 51. Retrieved May 11, 2023, from https://hal.science/hal-03489917v1/preview/S1062940819302992.pdf

Harris, L., & Gurel, E. (1986). Price and Volume Effects Associated with Changes in the S&P 500 List: New Evidence for the Exitence of Price Pressures. *The Journal of Finance*, 815-829. Retrieved May 2, 2023, from https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1986.tb04550.x

Hegde, S. P., & McDermott, J. B. (2003). The liquidity effects of revision to the S&P 500 index: an empirical analysis. *Jounrla of Financial Markets*, 413-459. Retrieved May 9, 2023, from https://www.sciencedirect.com/science/article/pii/S1386418102000460

Honghui, C., Singal, V., & Whitelaw, R. F. (2016). Comovement revisited. *Journal of Financial Economics*, 624-644. Retrieved May 3, 2023, from https://www.sciencedirect.com/science/article/pii/S0304405X16300988

Jason, J. (2020, April 8). *The Coronavirus Stock Market: A Market Gone Wild*. Retrieved May 16, 2023, from www.forbes.com: https://www.forbes.com/sites/juliejason/2020/04/08/the-coronavirus-stock-market-a-market-gone-wild/?sh=2c7e0b16a31f

Kamara, A., Lou, X., & Sadka, R. (2008). The Divergence of Liquidity Commonality in the Cross-Section of Stocks. *The Journal of Financial Economics*, 444-466. Retrieved May 3, 2023, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=943040

Kamara, A., Lou, X., & Sadka, R. (2010). Has the U.S. Stock Market Become More Vulnerable over Time? *Financial Analyst Journal*, 41-52. Retrieved May 3, 2023, from https://www.jstor.org/stable/27809161

Lamorte, W. W. (2017, May 4). *Wilcoxon Signed Rank Test*. Retrieved April 19, 2023, from https://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_nonparametric/BS704_Nonparametric6.html: https://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_nonparametric/BS704_Nonparametric6.html

Lynch, A. W., & Mendenhall, R. R. (1997). New Evidence on Stock Price Effects Associated with Changes in the S&P 500 Index. *The Journal of Business*, 351-383. Retrieved May 2, 2023, from https://econpapers.repec.org/article/ucpjnlbus/v_3a70_3ay_3a1997_3ai_3a3_3ap_3a351-83.htm

Morck, R., & Yang, F. (2001). The Mysterious Growing Value of S&P 500 Membership. *NBER Working Paper No. 8654*. Retrieved May 2, 2023, from https://www.nber.org/system/files/working_papers/w8654/w8654.pdf

Patel, N., & Welch, I. (2017). Extended Stock Returns in Response to S&P 500 Index Changes. *The Review of Asset Pricing Strategies*, 172-208. Retrieved May 3, 2023, from https://academic.oup.com/raps/article-abstract/7/2/172/3787841

Petajisto, A. (2011). The index premium and its hidden cost for index funds. *Journal of Empirical Finance*, 271-288. Retrieved 05 2, 2023, from https://pages.stern.nyu.edu/~jwurgler/papers/indexing13.pdf#:~:text=Trillions%20of%20dollars%20are%20invested%20through%20index%20funds%2C,as%20index-linked%20investing%20continues%20to%20grow%20in%20popularity.

PortfoliosLab. (2023, April 4). *Garman Klass Volatility*. Retrieved from www.portfolioslab.com: https://portfolioslab.com/tools/garman-klass

Pruitt, S. W., & Wei, K. C. (1989). Institutional Ownership and Changes in the S&P 500. *The Journal of Finance*, 509-513. Retrieved May 22, 2023, from https://www.jstor.org/stable/pdf/2328603.pdf?casa_token=ppAqQhTPkFwAAAAA:UEAAu-H13sSxsgtbodXpRamaN9Q1u7ZYKAub9Aa7fGvRNSvJsomPLoxMJQiGQkBpLlxJnnxrspJTFQujOGc5qjIZMr6b9tnpQ9QCwWNAn3EWATdv

PwC. (2020). *Evolution of foreign investments by pension funds 2020 edition*. Retrieved May 22, 2023, from www.pwc.lu: https://www.pwc.lu/en/market-research-centre/docs/pwc-mrc-beyond-their-borders-2020.pdf

Roll, R. (1984). A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market. *The Journal of Finance*, 1127-1139. Retrieved May 11, 2023, from https://onlinelibrary.wiley.com/doi/10.1111/j.1540-6261.1984.tb03897.x

S&P Global. (2021). *Annual Survey of Assets*. Retrieved from www.spglobal.com: https://www.spglobal.com/spdji/en/documents/index-news-and-announcements/spdji-indexed-asset-survey-2020.pdf?utm_source=pdf_research

Schleifer, A. (1986). Do Demand Curves for Stocks Slope Down. *The Journal of Finance*, 579-590. Retrieved May 2, 2023, from https://scholar.harvard.edu/sites/scholar.harvard.edu/files/shleifer/files/demandcurves.pdf

Sharpe, W. F. (1991). https://www.tandfonline.com/doi/abs/10.2469/faj.v47.n1.7?journalCode=ufaj20. *Financial Analyst Journal*, 7-9. Retrieved March 8, 2023, from https://www.tandfonline.com/doi/abs/10.2469/faj.v47.n1.7?journalCode=ufaj20

Siblis Research. (n.d.). *Total Market Value of the U.S. Stock Market*. Retrieved April 17, 2023, from www.siblisresearch.com: https://siblisresearch.com/data/us-stock-market-value/

Standard & Poor. (2023, March). *S&P U.S. Indices Methodology.* Retrieved April 17, 2023, from www.spglobal.com: https://www.spglobal.com/spdji/en/documents/methodologies/methodology-sp-us-indices.pdf?utm_source=pdf

Standard & Poor. (n.d.). *S&P 500, The Gauge of the U.S. Large-Cap Market.* Retrieved April 17, 2023, from www.spglobal.com: https://www.spglobal.com/spdji/en/documents/additional-material/sp-500-brochure.pdf

Statista. (2022, January 11). *S&P 500 performance during major crashes as of August 2020*. Retrieved May 16, 2023, from www.statista.com: https://www.statista.com/statistics/1175227/s-and-p-500-major-crashes-change/

Statista. (2023, March 08-03-2023). *Development of assets of global exchange traded funds(ETFs) from 2003 to 2022 (in billion U.S. dollars)*. Retrieved from www.Statista.com: www.statista.com

Statista. (2023, March 8). *Number of exchange traded funds (ETFs) worldwide from 2003 to 2022*. Retrieved from www.statista.com: www.statista.com

Steliaros, M. E., & Rossi, G. D. (2022). The Shift from Active to Passive and its Effect on Intraday Stock Dynamics. *Journal of Banking and Finance*, Vol. 143. Retrieved May 2, 2023, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4262415

Stobierski, T. (2018, March 30). *401(k) Basics: When it Was Invesnted and How It Works*. Retrieved May 12, 2023, from www.northwesternmutual.com: https://www.northwesternmutual.com/life-and-money/your-401k-when-it-was-invented-and-why/

Sullivan, R. N., & Xiong, J. X. (2012). How Index Trading Increases Market Vulnerability. *Financial Analyst Journal*, 70-84. Retrieved May 3, 2023, from https://www.tandfonline.com/doi/abs/10.2469/faj.v68.n2.7?casa_token=1ze9Ftj2VyIAAAAA:Q7H7naM2M5hXPpfIAaDZj6NK42CosSsivRkwUSCJsuov_-t84aKxL2YTtM4ksyHftO--EAW8oNmXMw

Sushko, V., & Turner, G. (2018). *The Implications of Passive Investing for Securities Markets.* Bank for International Settlements.

Trainer, D. (2020, June 1). *The Hidden Dangers Of Passive Investing*. Retrieved May 12, 2023, from www.Forbes.com: https://www.forbes.com/sites/greatspeculations/2020/06/01/the-hidden-dangers-of-passive-investing/?sh=1ff887834d96

Vanguard. (n.d.). *What is the history of ETFs?* Retrieved May 12, 2023, from www.vanguardmexico.com: https://www.vanguardmexico.com/en/learn/explore/etf-fundamentals/basics/what-is-the-history-of-etfs

VettaFi. (n.d.). *ETF Screener*. Retrieved March 16, 2023, from www.etfdb.com: https://etfdb.com/screener/#page=2&asset_class=equity&sizes=Large-Cap%2CMega-Cap&regions=U.S.&active_or_passive=Passive

Vijh, A. M. (1994). S&P 500 Trading Strategies and Stock Betas. *The Review of Financial Studies*, 215-251. Retrieved May 3, 2023, from https://academic.oup.com/rfs/article-abstract/7/1/215/1568661?redirectedFrom=fulltext

Wang, F. (2023, February 3). *2023 GICS Changes: S&P 500 Impact Analysis*. Retrieved April 4, 2023, from www.indexologyblog.com: 2023 GICS Changes: S&P 500 Impact Analysis

Wurgler, J. (2011). On the Economic Consequences of Index-Linked Investing. *Challenges to Business in the Twenty-First Centurey*, 1-27. Retrieved April 28, 2023, from https://pages.stern.nyu.edu/~jwurgler/papers/indexing13.pdf#:~:text=Trillions%20of%20dollars%20are%20invested%20through%20index%20funds%2C,as%20index-linked%20investing%20continues%20to%20grow%20in%20popularity.

Wurgler, J., & Zhuravskaya, E. (2002). Does Arbitrage Flatten Demand Curves for Stocks? *The Journal of Business*, 583-608. Retrieved May 2, 2023, from https://www.jstor.org/stable/10.1086/341636

YCharts. (n.d.). *S&P 500 Market Cap*. Retrieved March 17, 2023, from www.ycharts.com: https://ycharts.com/indicators/sp_500_market_cap

Ødegaard, B. A. (2023, April 20). *Trading costs - The High Low Spread Estimator*. Retrieved from www.ba-odegaard.no: https://ba-odegaard.no/teach/notes/liquidity_estimators/corwin_schultz_high_low_estimator/lectures_high_low.pdf

Appendix

1) Stocks Added to the S&P 500 Index During 2015-2017

Symbol	Security	GICS Sector	Date added	Founded	Comment
SWKS	Skyworks Solutions	Information Technology	12-03-2015	2002	Excluded
AAL	American Airlines Group	Industrials	23-03-2015	1934	Excluded
KHC	Kraft Heinz	Consumer Staples	06-07-2015	2015 (1869)	Excluded
PYPL	PayPal	Information Technology	20-07-2015	1998	Excluded
FOX	Fox Corporation (Class B)	Communication Services	18-09-2015	2019	Excluded
NWS	News Corp (Class B)	Communication Services	18-09-2015	2013 (1980)	Excluded
HPE	Hewlett Packard Enterprise	Information Technology	02-11-2015	2015	Excluded
SYF	Synchrony Financial	Financials	18-11-2015	2003	Excluded
CFG	Citizens Financial Group	Financials	29-01-2016	1828	Excluded
FTV	Fortive	Industrials	01-07-2016	2016	Excluded
HLT	Hilton Worldwide	Consumer Discretionary	19-06-2017	1919	Excluded
IQV	IQVIA	Health Care	29-08-2017	1982	Excluded
NCLH	Norwegian Cruise Line Holdings	Consumer Discretionary	13-10-2017	2011 (1966)	Excluded
HCA	HCA Healthcare	Health Care	27-01-2015	1968	
HSIC	Henry Schein	Health Care	17-03-2015	1932	
EQIX	Equinix	Real Estate	20-03-2015	1998	
O	Realty Income	Real Estate	07-04-2015	1969	
QRVO	Qorvo	Information Technology	11-06-2015	2015	
JBHT	J.B. Hunt	Industrials	01-07-2015	1961	
AAP	Advance Auto Parts	Consumer Discretionary	09-07-2015	1932	
ATVI	Activision Blizzard	Communication Services	31-08-2015	2008	
UAL	United Airlines Holdings	Industrials	03-09-2015	1967	
VRSK	Verisk	Industrials	08-10-2015	1971	
ILMN	Illumina	Health Care	19-11-2015	1998	
CHD	Church & Dwight	Consumer Staples	29-12-2015	1847	
WTW	Willis Towers Watson	Financials	05-01-2016	2016	
EXR	Extra Space Storage	Real Estate	19-01-2016	1977	
FRT	Federal Realty	Real Estate	01-02-2016	1962	
AWK	American Water Works	Utilities	04-03-2016	1886	
UDR	UDR, Inc.	Real Estate	07-03-2016	1972	
CNC	Centene Corporation	Health Care	30-03-2016	1984	
HOLX	Hologic	Health Care	30-03-2016	1985	
ULTA	Ulta Beauty	Consumer Discretionary	18-04-2016	1990	
GPN	Global Payments	Information Technology	25-04-2016	2000	
ALK	Alaska Air Group	Industrials	13-05-2016	1985	
DLR	Digital Realty	Real Estate	18-05-2016	2004	
LKQ	LKQ Corporation	Consumer Discretionary	23-05-2016	1998	
AJG	Arthur J. Gallagher & Co.	Financials	31-05-2016	1927	
TDG	TransDigm Group	Industrials	03-06-2016	1993	
ALB	Albemarle Corporation	Materials	01-07-2016	1994	
LNT	Alliant Energy	Utilities	01-07-2016	1917	
MTD	Mettler Toledo	Health Care	06-09-2016	1945	
CHTR	Charter Communications	Communication Services	08-09-2016	1993	
COO	CooperCompanies	Health Care	23-09-2016	1958	
MAA	Mid-America Apartment Communities	Real Estate	02-12-2016	1977	
IDXX	Idexx Laboratories	Health Care	05-01-2017	1983	
INCY	Incyte	Health Care	28-02-2017	1991	
CBOE	Cboe Global Markets	Financials	01-03-2017	1973	
REG	Regency Centers	Real Estate	02-03-2017	1963	
DISH	Dish Network	Communication Services	13-03-2017	1980	
SNPS	Synopsys	Information Technology	16-03-2017	1986	
RJF	Raymond James	Financials	20-03-2017	1962	
AMD	AMD	Information Technology	20-03-2017	1969	
ARE	Alexandria Real Estate Equities	Real Estate	20-03-2017	1994	
DXC	DXC Technology	Information Technology	04-04-2017	2017	
IT	Gartner	Information Technology	05-04-2017	1979	
RE	Everest Re	Financials	19-06-2017	1973	
ALGN	Align Technology	Health Care	19-06-2017	1997	
ANSS	Ansys	Information Technology	19-06-2017	1969	
BKR	Baker Hughes	Energy	07-07-2017	2017	
MGM	MGM Resorts	Consumer Discretionary	26-07-2017	1986	
RMD	ResMed	Health Care	26-07-2017	1989	
AOS	A. O. Smith	Industrials	26-07-2017	1916	
PKG	Packaging Corporation of America	Materials	26-07-2017	1959	
SBAC	SBA Communications	Real Estate	01-09-2017	1989	
CDNS	Cadence Design Systems	Information Technology	18-09-2017	1988	

Table 28: Showing an overview of stocks added to the S&P 500 index during 2015 to 2017. Source: press releases from Standard & Poor (can be handed upon request)