Passive Efficiency

Exploring Stocks' Return, Risk, & Efficiency Through Index Investing in Europe

Markus Lexander & Filip Sjöberg

Department of Economics

Master's Programme in Finance

Supervisor: Dr. Thomas Fischer

May 2024

Abstract

This study investigates the consequences of stocks being included in an index, in terms of returns, risk, and pricing efficiency. We examine the European stock market, utilizing the broad STOXX Europe 600 index, which, to a large extent, has been overseen by previous research. Over the recent period 2013-2023, we document all additions to the index and collect fund ownership data, classified as index (passive) and active, on all stocks before and after the index inclusion. Employing a two-way fixed effect panel regression, we examine potential effects in relation to fund ownership. The study suggests that active fund ownership is significantly related to positive stock returns, as seen by both the arithmetic and geometric mean return. Concerning stocks' risk, measured by volatility and beta, we find a significant positive connection to index fund ownership. Regarding pricing efficiency, tested through the variance ratio test, we observe insignificant results, consistent with previous research. In essence, our findings suggest that inclusion in an index does not harm the efficiency of the pricing mechanism. However, a potential side effect is an increase in idiosyncratic risk as index ownership rises.

Keywords: European stock market, index investing, fund ownership, pricing efficiency, fixed effect panel regression

JEL Classification: G12, G14

Acknowledgments

We would like to especially thank our supervisor, Dr. Thomas Fischer, for his support and help during the process of writing this thesis. With his advice, we steered in the right direction and overcame any problem that we encountered. In addition, we would also like to thank our fellow students for pointing out faults and areas of improvement.

Table of Contents

List of Figures

List of Tables

1 Introduction

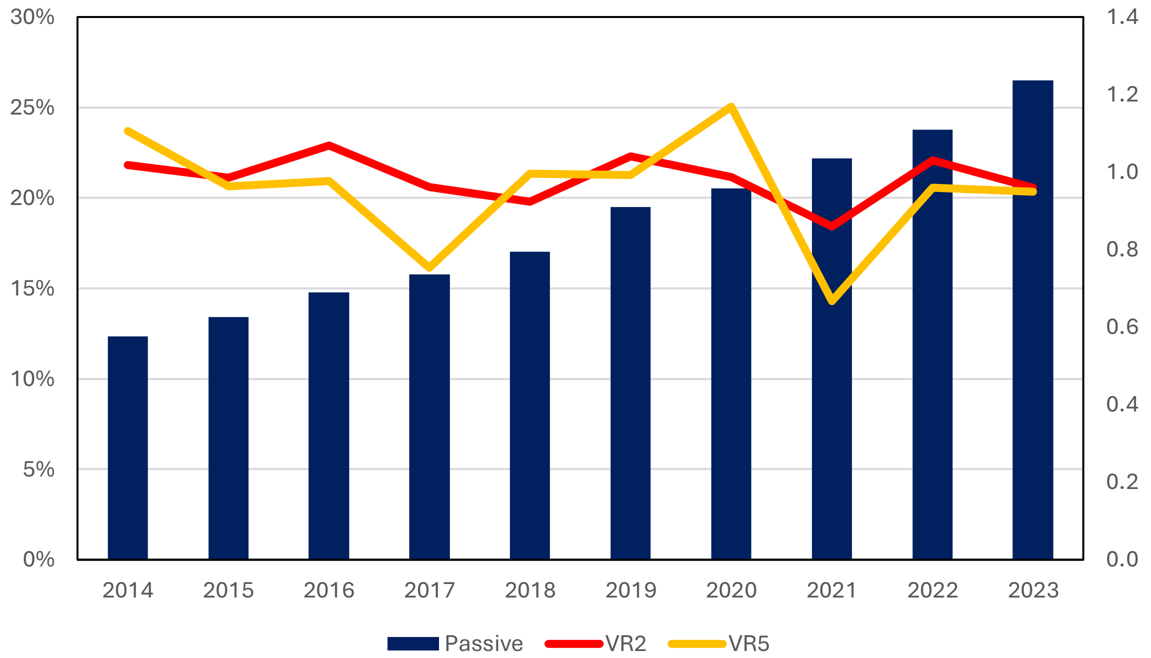
Over the past decade, the investment landscape has witnessed a significant shift towards index funds and a growing preference for passive investment strategies over active management (Coles *et al.*, 2022; Investment Company Institute, 2024). Index funds, which aim to replicate the performance of a specific market index by mirroring its composition, have gained popularity due to their lower costs and simplified management approach (Chabakauri and Rytchkov, 2020; Investment Company Institute, 2024). Moreover, passive investment strategies have often been shown to create superior returns over active strategies, adjusting for fees (Carhart, 1997; Fama and French, 2010; Jaquart *et al.*, 2023). This raises questions, however, as to what effects this might have on individual stocks, as the ownership structure changes from active to passive investors.

This study investigates the consequences of stocks being included in an index, with regard to returns, risk, and pricing efficiency. We examine the European stock market, using the broad STOXX Europe 600 index, which, often, has been overseen for the larger American stock market in previous research. Over the recent period 2013-2023, we document all additions to the index and collect fund ownership data, classified as index and active, on all stocks before and after the index inclusion. Next, we perform a two-way fixed effect panel regression to test for possible effects in relation to fund ownership. From our tests, we find that active fund ownership is significantly related to positive stock returns. Concerning stocks' risk, we find a significant positive connection to index fund ownership. Regarding pricing efficiency, we document insignificant results, consistent with previous research. In short, our findings imply that being included in an index does not harm the pricing efficiency. However, a side effect may be that the stocks' risk increases, as passive ownership increases.

Index mutual funds were introduced in the 1970s and were later followed by exchange-traded funds (ETFs) in the 1990s (Chabakauri & Rytchkov, 2021; Investment Company Institute, 2024). By the end of 2023, U.S. index mutual funds and ETFs amassed a total of $13.3 trillion in net assets, representing 48 percent of the assets in long-term funds – a significant increase from 19 percent by the end of 2010 (Investment Company Institute, 2024). Despite this growth, index funds are still relatively minor players in the U.S. stock markets, holding only 18 percent of the value of U.S. stocks (year-end 2023). In contrast, actively managed funds collectively hold 13 percent, with the majority of the market remaining in the hands of other investors such as hedge funds, pension funds, and individuals (Investment Company Institute, 2024).

Figure 1.1

European Passive Ownership vs. Market Efficiency



Note. The left y-axis shows the percent ownership by passive funds in the European stock market, as reflected by a broad European ETF (Europe-Domiciled OE Funds & ETFs ex MM ex FOF ex Feeder). The right y-axis measures the variance ratio (VR) test, with lags of two and five days. The tests are done on the same ETF as the ownership data. The period covers ten years, from 2014 to 2023. Source: Morningstar (2024).

Given the growing popularity and influence of passive investments, it is crucial to understand the long-term effects of these dynamics. Most existing studies have focused on the American equity market, with particular attention on the Russell 1000 and 2000 indices, to examine the effects of increased passive index investing (Ben-David *et al.*, 2018; Coles *et al.*, 2022; Glosten *et al.*, 2021). Figure 1.1 illustrates that a similar trend of increasing passive fund investments has developed in the European stock market as well. The data, collected from Morningstar (2024) for the period from 2014 to 2023, is presented alongside results from two variance ratio (VR) tests, with lags of two and five days. The tests are performed on the same underlying ETF that the ownership data is based on. The VR test tests for market efficiency, specifically whether the price adheres to a random walk hypothesis, where a value equal to one indicates a random walk, and hence an efficient market. Both VR lines exhibit some variation, particularly the VR5, which fluctuates more than the VR2. A discernible trend is present, although weak, with values starting slightly above one and converging towards one by 2023. Figure 1.1 indicates that while passive investing has increased substantially in Europe, the equity market has simultaneously become slightly more efficient. Although not a clear trend, this emerging pattern underscores the importance of further research into how passive investing influences the stock market and individual stocks.

There has been a growing body of research on the influence of investor composition on stocks' return, risk, and pricing efficiency. However, there is no clear consensus in the current literature. Regarding the influence of passive investing on stock returns, opinions among researchers are divided. Vijh and Wang (2022) find a negative relationship between stock returns and inclusion in the S&P 500 index, which contradicts the results of similar studies in the area (see *e.g.,* Chen *et al.* (2004) and Patel and Welch (2017)). Similarly, the influence of passive investing on stocks' risk has been studied in various ways. Da and Shiva (2018) examine the consequences of ETF trading activity and find that it creates excess comovement, leading to an increase in stocks' riskiness. Ben-David *et al.* (2018) investigate the relationship between indexing and risk using the Russell 1000 and 2000 indices. The authors identify a significant relationship between increased passive investing and volatility, noting that an increase in passive investing results in negative autocorrelation in stock prices. Additionally, Chabakauri and Rytchkov (2021) suggest that the simultaneous price pressure exerted on all stocks by indexing could increase return correlations.

The impact of passive investing on pricing efficiency also remains a topic of debate. Coles *et al.* (2022) argue that the rise in passive investing may not necessarily degrade price efficiency. Conversely, studies by Bond and Garcia (2022) and Breugem and Buss (2019) propose that passive funds' mechanical buying of index stocks, without assessing their intrinsic value, could reduce the market's responsiveness to new information. On the other hand, Baltussen *et al.* (2019) suggest that the presence of index funds and ETFs, which offer investors diversified and easily tradable portfolios, may lead to more efficient incorporation of market-wide information.

Considering the rapid growth of passive investing, it is crucial to understand the long-term effects this might have on individual stocks. It is important to note that there are some discrepancies in the literature regarding the definition of passive investing. This thesis will specifically study the effect of index investing, *i.e.*, index fund ownership, and compare it with other fund ownership structures labeled as active. Furthermore, most existing studies focus on the American equity market, particularly the Russell 1000 and 2000 indices, to examine the effects of increased passive index investing (see *e.g.*, Ben-David *et al.* (2018), Coles *et al.* (2022), and Glosten *et al.* (2021)). However, there appears to be a lack of research on the rise of index investments in the European equity market, and its subsequent effects on stocks' return, risk, and pricing efficiency, especially in more recent times. To the best of our knowledge, no previous research focus on the increase in index investing in the European stock market, particularly the STOXX Europe 600 index, and its effects on stocks' return, risk, and price efficiency. The goal of this thesis is to contribute to the existing literature by investigating how passive investing has influenced stocks' return, risk, and pricing efficiency from 2013 to 2023. Given this, we formulate the research question to be:

RQ: How has the rise of index fund ownership influenced stocks' return, risk, and pricing efficiency, within the European equity market (STOXX Europe 600 index) during the period from 2013 to 2023?

To test this in practice, we collect fund ownership data, classified as either index (proxy for passive fund ownership) or active, on all stocks that got included in the STOXX Europe 600 index during the period 2013-2023. We collect this data for five quarters per stock, two before and two after the index inclusion, to see for possible effects. Next, we perform both Welch's *t*-test and the Mann-Whitney *U* test to confirm a significant increase in fund ownership after being included in the index. To test for stocks' return, risk, and pricing efficiency, we utilize a two-way fixed effects panel regression, that allows us to discern any difference before and after the inclusion. To see if the effects are different during different years, we separate the full ten-year period into four periods, among them the pandemic years 2020-2022. First, we establish that there is a significant increase in index ownership after being included in the index. From our regressions, we find a significant positive relationship between active fund ownership and stock returns. Further, we show that increased index fund ownership is positively related to stocks' risk. In relation to pricing efficiency, we do not find any significant relationship, with regards to either index or active fund ownership. As to the different periods, we do not find any clear pattern, although the latter five years show higher significance than the first five years.

The remainder of the thesis is structured in the following order. Section 2 discusses the main contributions from previous research with regard to index investing. Section 3 describes the data we use and the research design. Specifically, it discusses the panel regression we employ to test for the consequences of being included in the index. Section 4 presents our empirical results and discusses them in a broader sense, relating them to both our hypothesis and the previous literature, as well as mentioning possible limitations. Finally, Section 5 concludes our thesis by summarizing the main points and implications and points out new directions for further research.

2 Literature Review

This section investigates the previous literature's findings on passive investing and its effect on individual stocks. We differentiate between theoretical and empirical literature, which separates the previous research. Some papers, though, such as that by Coles *et al.* (2022), apply both an empirical and theoretical approach. Note that this thesis focuses specifically on index investing, while many previous studies use a broader definition of passive investing. Given too few articles focus only on index investing, we include previous studies covering more ownership types than index investing, such as ETFs.

2.1 Theoretical Predictions

In the theoretical literature, a model is typically developed from scratch to explore the impacts of increased index investing on individual stocks. These models suggest possible outcomes, providing predictions rather than definitive results, distinguishing them from empirical studies. Coles *et al.* (2022) build a framework that relies on Grossman and Stieglitz's (1980) original model. Grossman and Stieglitz's (1980) study does not actually incorporate index investing (as index investing was only in its early phases at the time), but creates a solid framework for the way in which information is discounted into asset prices. In Coles *et al.*'s (2022) model, the authors rely on three types of investors: passive index investors, active publicly-informed investors, and active privately-informed investors. The authors argue that investors will only choose to invest more in information production, *i.e.*, become more active, whenever it is profitable to do so. From their model, they postulate four different hypotheses, which they, later, empirically test:

1) the volatility of the stocks' price increases the larger the fraction of index investors,
2) an increase in index investing does not change the fraction of active investors,
3) an increase in index investing leads to a decrease in firm-specific information production,
4) an increase in index investing leads to no change in price efficiency.

To test these hypotheses, Coles *et al.* (2022) utilize the Russel 1000 and 2000 indices as an exogenous change in index investing, by studying when stocks move from the small-cap index (Russel 2000) to the large-cap index (Russel 1000). When moving to the larger index, the amount of index capital will decrease, since the stock will have a lower index weight in the large-cap index, than it had in the smaller index. They use a time frame from 2007 to 2016 and apply an event study method to discern the effect of moving between the two indices, examining a period before and after the index inclusion. Coles *et al.* (2022) find the following: When index investing increases, information production is lower, although, the price efficiency, remains unchanged. These results, the authors relate to their theoretical model, which also postulates that passive investing would not reduce price efficiency (see hypothesis 4 above).

Chabakauri and Rytchkov (2021) investigate the effect of increased index investing on the market equilibrium; in particular, the effect of investors changing from trading individual stocks to indices. In their model, they find that increased index investing leads to decreased volatility of market returns, while the effect on correlations is ambiguous. Jaquart *et al.* (2023) simulate a hypothetical financial market to examine the effect of different levels of active and passive investments on market efficiency. They find that an increase in the fraction of active investments is associated with an increase in fundamental market efficiency, which does not hold for the passive investment. Moreover, Jaquart *et al.* (2023) present a possible link between passive investment to market failure, and show that it might facilitate price bubbles. Buss and Sandaresan (2023), on the contrary, establish that increased index investing leads to higher pricing efficiency, as seen by the informativeness of the stock price.

Similar to Chabakauri and Rytchkov (2021), Bond and Garcia (2022) develop a model to investigate the effect of index investing on the market equilibrium. In addition, they also perform a welfare analysis to analyze the consequences of increased indexing. Their model relies on a standard rational expectation setting, similar to that of Grossman and Stieglitz (1980), also used by Coles *et al.* (2022). Bond and Garcia (2022) study, particularly, the effects of index investing becoming cheaper, as witnessed during the last decades. The authors show that a reduced cost will lead to increased indexing, which will reduce the price efficiency of the index, while increasing the relative price efficiency of individual stocks. Baruch and Zhang (2022) develop a model based on a CAPM framework and investigate the consequences of more investors becoming indexers. They find that this shift diminishes the price efficiency of stocks and increases the stocks' risk.

2.2 Empirical Evidence

Even though index mutual funds have been established since the 1970s (Bogle, 2016), the academic investigations on its rise and their effect on stocks have only been developed more recently (Coles *et al.*, 2022; Nandal & Kumar, 2021). The previous research focus on different subjects, such as index investing's effect on different market characteristics of stocks (Jiang *et al.*, 2022; Liebi, 2020). The characteristics vary between different studies, yet some examples are: volatility, pricing efficiency, and returns. Three findings shown to have a connection to index investing are: 1) excess comovements, 2) increased return volatility, and 3) reduced stock liquidity. The origin of these effects is so-called basket trading, meaning the mass-buying or selling of index constituents as a result of reconstitutions (Ahn & Patatoukas; 2022). The first finding is shown by Sullivan and Xiong (2012) and Da and Shive (2018). Return volatility is investigated by Krause *et al.* (2014) and Ben-David *et al.* (2018). The effect of decreased stock liquidity is found by Israeli *et al.* (2017). All of these studies identify the downside of increased index investing, since it impedes price discovery, thereby decreasing the price efficiency of stocks.

Ahn and Patatoukas (2022), like many articles that study the U.S. market (see *e.g.*, Coles *et al.* (2022), Ben-David *et al.* (2018), and Glosten *et al.* (2021)), employ the Russel 1000 and 2000 indices as an exogenous variation, to study the effects of increased index investing. Ahn and Patatoukas (2022) focus, specifically on information arbitrage and price discovery. They find that index investing leads to an increase in the speed of price adjustments to news for, in particular, micro-cap stocks. The same effect is, however, not found for large- or mid-cap stocks. Moreover, in contrast to other papers, Ahn and Patatoukas (2022) establish that increased index investing does not lead to increased volatility. Israeli *et al.* (2017) also studies the U.S. market, although they focus on ETFs instead, and especially highlight the relationship between ETF ownership and the pricing efficiency of stocks. The authors find that an increase in ownership may lead to higher trading costs and worse information production for the underlying securities. As such, the pricing efficiency decreases with increased passive ownership. Another study that finds evidence of decreased pricing efficiency due to increased ETF ownership is that by Chen *et al.* (2024). Chen *et al.* (2024) evaluate the Chinese market during a 10-year time frame from 2012 to 2021. In contrast to Israeli *et al.* (2017) and Chen *et al.* (2024), Glosten *et al.* (2021) find evidence of a positive connection between ETF ownership and stocks' pricing efficiency during a short time horizon. Glosten *et al.* (2021) first construct a theoretical model to base their hypothesis, and then test them using the Russel 1000/2000 reconstitution framework. The authors highlight a distinction between weak and strong informational environments for the underlying securities. With a weak informational environment, they see a positive relationship with increased ETF ownership, whereas no effect is found for stocks with a stronger informational environment.

The relationship between increased index ownership and stock returns is an area with relatively fewer previous studies, although some exist. Generally, the studies are based on a before-and-after analysis of stocks that get included in an index and their subsequent returns. Vijh and Wang (2022) studies the U.S. market by investigating stocks that move between the large-cap stock index S&P 500 and the mid-cap stock index S&P 400. They note that while the majority of prior studies (see *e.g.*, Chen *et al.* (2004) and Patel and Welch (2017)) find a positive association with inclusion in the S&P 500, Vijh and Wang (2022) observe a reverse relationship, where inclusion in the S&P 500 correlates with negative returns. The authors attribute this to lower institutional or passive ownership in the S&P 500 compared to the S&P 400, as stocks in the former constitute a smaller portion of the index. Thus, Vijh and Wang (2022) find evidence supporting a positive relationship between returns and passive ownership. Similarly, Chang *et al.* (2015) investigate prices, or returns, after index changes, using the Russel 1000 & 2000 framework. They find that when a stock moves from the Russel 2000 to the Russel 1000, it is associated with negative returns, and contrary when moving from the 1000 index to the 2000 index. Chang *et al.* (2015) relate this finding to more institutional capital following the largest stocks in the Russel 2000, than those with the lowest weights in the 1000 index. Hence, this explains the positive, or negative, returns, which is similar to the finding by Vijh and Wang (2022).

Da and Shiva (2018) analyze the U.S. stock market, studying multiple ETFs to assess the consequences of ETF trading activity on individual stock constituents. They find that it will create excess comovement, thereby increasing the risk of the stocks. Staer and Sottile (2018) also investigate return comovement using ETFs and intraday data. The authors find a significant positive relationship between increased ETF trading and return comovement of the stocks. Ben-David *et al.* (2018) focus on the relationship between indexing and risk, utilizing the Russell indices (1000 and 2000) to study the effect of increased index investing on firm-specific volatility. Analyzing data from 1996 to 2006, they identify a significant positive relationship between passive investing and volatility. Further, the authors show that increased passive investing results in negative autocorrelation in stock prices.

Baltussen *et al.* (2019) corroborate Ben-David *et al.*'s (2018) findings by examining 20 major indices worldwide from 1951 to 2016. Baltussen *et al.* (2019) observe a shift in autocorrelation from positive to negative after the 2000s. With this change in autocorrelation, they find a significant connection to the global increase in index investing. Further evidence of a relationship between passive investing (ETFs in this case) and volatility is given by Wang and Xu (2019), in their study on the Chinese stock market. The authors demonstrate that as ETF flows increase, it leads to higher volatility of the component stocks.

2.3 Hypothesis Development

As apparent from the discussion of the previous research, both the pure theoretical and the empirical, different studies find different results, often contradicting each other. Given our chosen methodology and time frame for the thesis, certain sources bear greater similarity, offering more guidance for hypothesis development. Among them are Coles *et al.* (2022), that investigates both efficiency and volatility, Ben-David *et al.* (2018) that focuses on volatility, and Vijh and Wang (2022), that examines the relationship to returns. Taking all previous research into consideration, with a particular focus on these three studies, we propose the following hypotheses for testing:

H1: The increase in index fund ownership leads to an increase in stocks' return.
H2: The increase in index fund ownership leads to an increase in stocks' risk.
H3: The increase in index fund ownership leads to a decrease in stocks' price efficiency.

3 Data & Research Design

This section presents the data utilized in our analysis, detailing where it is sourced, and the methodology employed. We start by discussing the index used (STOXX Europe 600) and its construction, followed by an explanation of the data cleaning process, which reduced the initial stock set from 523 to 387. Subsequently, we outline the applied methodology, particularly the two-way fixed effect panel regression, and proceed to present all variables used in the tests.

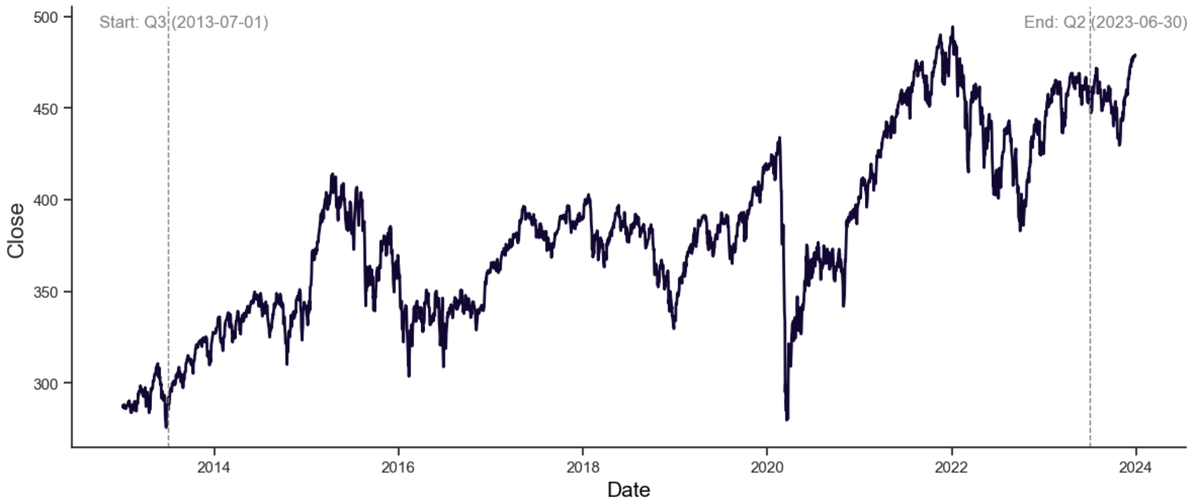
3.1 STOXX Europe 600

The STOXX Europe 600 index is a stock index covering the European market with 600 constituents, and the weighting scheme is based on free float markets capitalization. It covers 17 European countries, including both euro and non-euro countries, comprising companies that represent large, medium, and small capitalizations (STOXX, 2024). The index accounts for approximately 90% of the free-float market capitalization of the European stock market (STOXX, 2024). It is reviewed quarterly, coinciding with the announcements of new additions and deletions. These announcements occur on the first day of every quarter: January 1st, April 1st, July 1st, and October 1st (STOXX, 2024), although the actual inclusion of stocks may vary within the quarter. While the precise timing of inclusions may fluctuate, weak market efficiency is assumed, implying that market effects occur upon announcement (Coles *et al.*, 2022; Fama, 1970; Lo & MacKinlay, 1988).

Figure 3.1 displays the daily closing prices of the STOXX Europe 600 index during the entire sample period, *i.e.*, from 2013 to 2023. This period spans ten years from the first to the last index inclusion, plus an additional two quarters before and after, to capture the full sample window. For the analysis, the sample period is segmented into four phases: 1) the entire ten-year period from 2013 to 2023, 2) the first five years from 2013 to 2018, 3) the second five years from 2018 to 2023, and 4) the pandemic period from 2020 to the end of 2022. Each five-year segment begins at the start of Q3 (July 1st) and concludes at the end of Q2 (June 30th). The time frame for the COVID-19 pandemic spans from January 1, 2020, to December 31, 2022.

Figure 3.1

STOXX Europe 600 Index from January 2013 to December 2023



Note. Daily closing prices of the STOXX Europe 600 index from January 2013 to December 2023 from Yahoo Finance. The dashed vertical lines represent the sample period's start and end points, from the first index inclusion to the last one, respectively.

Over the full period, a total of 523 stocks were included in the index, according to Refinitiv Eikon. However, after the data cleaning process, discussed in Section 3.3, the number of entities used was 387 for the full ten years; 164 for the first five years; 223 for the second five years; and 147 during the pandemic period. The variance in the number of entities between the first and second halves of the sample can be attributed to two primary factors. First, the number of stocks added to the index each quarter varied, as is evident in Appendix A, Table A.1. Second, the data cleaning process entailed the removal of stocks without considering the specific year of their removal, potentially distorting the inclusion count. Consequently, expired stocks, which are more prevalent among earlier inclusions due to the passage of time, disproportionately affect the first five years. This discrepancy explains the 59 fewer entities in the early years (the first five years) compared to the latter half.

3.2 Data Collection

The sample dataset is sourced from Refinitiv Eikon's database and consists of the stocks that were added to the STOXX Europe 600 index between July 1st, 2013, and June 30th, 2023, covering a total of ten years. In total, there were 523 inclusions in the STOXX Europe 600 index during this ten-year time frame (Refinitiv Eikon). Due to several of these stocks having been delisted during the time frame as a result of *e.g.*, buyouts, mergers, or bankruptcies (categorized as "expired" by Refinitiv Eikon), this reduced the final dataset to 387 stocks (see Section 3.3).

For each stock, fund ownership data was sourced from Refinitiv Eikon and categorized as either active or index (passive). The fund ownership data was calculated as a percent of the total shares outstanding for each stock. It is important to note that this data does not encompass 100% of equity ownership, as there are other types of owners such as families, retail investors, or private equity groups. The ownership classification was conducted by Refinitiv Eikon, who label owners according to type, including index funds, which we consider passive. Accordingly, all other types of fund ownership were classified as active. We aggregated the ownership percentages classified as index and active for each stock and quarter. The ownership data was collected five times per stock: during the inclusion event (Q0), two quarters before (Q-2, Q-1), and two after the event (Q+1, Q+2).

As our data collection extends two quarters before and after each index inclusion, the "full" sample period spans 11 years rather than 10. Daily adjusted closing prices for each stock over their respective five-quarter periods were collected from Yahoo Finance. This price data was utilized to calculate average daily percent stock returns, both arithmetic and geometric means, volatility, beta values, and variance ratio tests (see Section 3.5). Additionally, adjusted daily closing prices for the STOXX Europe 600 index over the entire sample period were sourced from Yahoo Finance.

3.3 Data Cleaning

Out of the 523 index inclusions that occurred over the period, not all stocks were available or suitable for our analysis. First, some stocks were no longer available on Yahoo Finance, since they were expired, necessitating their removal from our sample. Moreover, stock tickers used by Refinitiv Eikon differed from those used by Yahoo Finance, requiring manual adjustments to the ticker symbols to retrieve the data. During this step, there were instances where the Yahoo Finance ticker could not be found, likely because the company had been delisted or had changed its name. As a result, 50 stocks, or approximately 9.6% of the initial sample, were removed, leaving us with 473 stocks. Second, some stocks failed to meet the requirement of having pricing data available for all five quarters. The reasons for these discrepancies varied, however generally involved events such as spin-offs from larger companies. In such cases, the stock might have been directly included in the index due to its size or traded volume. The absence of "pre" data, *i.e.*, data from before the index inclusion, made these stocks unsuitable for the analysis, because we would not be able to perform a before-and-after analysis to examine the effect of the stocks' inclusion in the index. In total, 85 stocks, or approximately 16.25% of the initial sample, were removed due to incomplete data across the required quarters. Furthermore, a final exclusion involved one outlier, AIB Group PLC. This stock skewed the data due to being acquired and delisted by the Irish Government in December 2010, and subsequently relisted in 2017 under the same name and ticker. The stock's price data was skewed because when it was relisted, it continued trading from the last closing price in December 2010, resulting in a significant price increase of 1,880%. This outlier distorted calculations of returns, volatility, and beta, necessitating its removal from the dataset.

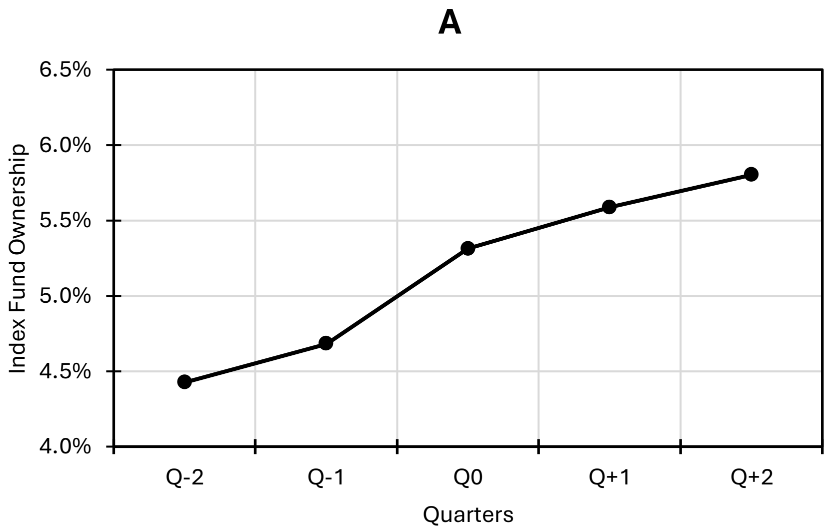
After implementing the described steps, a total of 136 stocks were removed from the initial dataset, resulting in a final sample comprising 387 stocks. With five observations (quarters) per stock, the total number of observations amounted to 1,935. Appendix A, Figure A.1 illustrates the number of stocks used in the sample from each period, as well as all stocks included per quarter. Notably, there is a higher number of stocks excluded at the beginning of the sample period compared to more recent years. This pattern is expected, as stocks included earlier in the sample have been publicly traded and part of the index for a longer duration, thus having a higher likelihood of being delisted. Consequently, a greater number of expired stocks are observed during the first half of the sample period, *i.e.*, between 2013 and 2018.

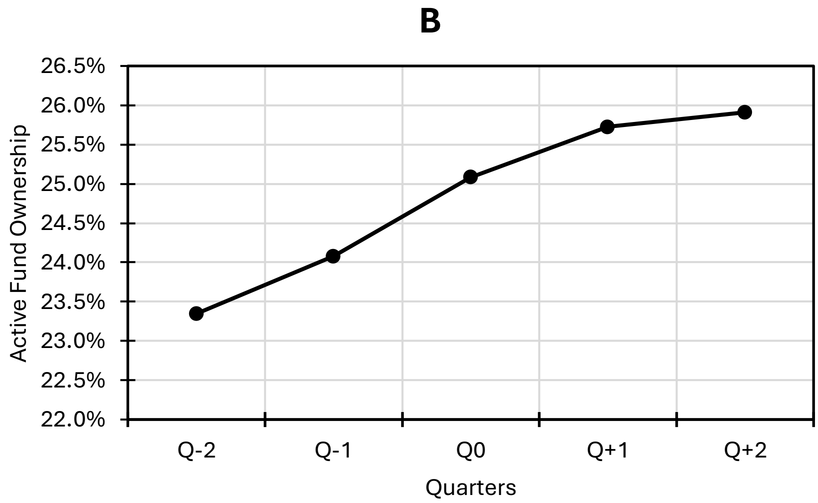
3.4 Fund Ownership Change

Before proceeding with the main analysis of this thesis, it was necessary to confirm a significant change in index and active fund ownership among the sample stocks from one quarter to the next. To test this, we created two datasets aggregating the sum of active and index ownership for each stock across each quarter. The first dataset organized the sum of index ownership across the quarters Q-2, Q-1, Q0, Q+1, and Q+2 in columns, with each row representing a specific stock. The second dataset followed the same structure, although with sums of active ownership.

Figure 3.2

Index & Active Fund Ownership Over the Quarters





Note. A and B illustrate the mean development of "Index Ownership" and "Active Ownership," respectively, representing the proportion of fund ownership of the sample stocks relative to the total ownership of each stock. The quarters denote the five quarters surrounding the quarter of inclusion (Q0) and the two quarters before and after. The y-axis is expressed in the percentage of total ownership. Each point denotes the average value for all 387 stocks across all quarters.

Figure 3.2 illustrates the sum of mean data for the index (A) and active fund ownership (B) for all 387 sample stocks, over the quarters before and after inclusion. For detailed summary statistics on index and active fund ownership, see Appendix B, Table B.1, Panel A and B, respectively. In both A and B, in Figure 3.2, an incremental increase in mean values is discernible over the five quarters. In A, the line for index ownership exhibits the sharpest rise, particularly pronounced when moving from Q-1 to Q0. This suggests that index ownership increased the most during the quarter of index inclusion. Similarly, in B, active ownership also shows an upward trend, albeit not as steep as index ownership. Notably, active ownership starts from a considerably higher baseline compared to index ownership. In relative terms, the increase in index ownership is more substantial, rising approximately 33% from Q-2 to Q+2, whereas active ownership increased about 12% over the same period.

3.4.1 Welch's t-test & Mann-Whitney U Test

From Figure 3.2, A and B, it is evident that both index and active fund ownership increase over the quarters. To assess whether these increases are statistically significant, we employed both Welch's *t*-test and the Mann-Whitney *U* test. These tests are particularly suitable due to the non-normal distribution and unequal variance across the quarter columns (Ahad & Yahaya, 2014; Callaert, 1999; Divine *et al.*, 2018; Mann & Whitney, 1947; Welch, 1947).

Welch's *t*-test is a robust alternative to the Student's *t*-test when dealing with samples that exhibit unequal variances and potentially different sizes (Ahad & Yahaya, 2014; Welch, 1947). The null hypothesis posits that there is no mean difference between the two groups, while the alternative hypothesis suggests that there is a mean difference (Welch, 1947). The test is calculated using the formula:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}} \tag{1}$$

where:

- $\bar{X}_1$ and $\bar{X}_2$ are the sample means.
- s_1^2 and s_2^2 are the sample variances.
- N_1 and N_2 are the sample sizes.

Welch's *t*-test is particularly suitable given the variability in both index and active fund ownership across the different quarters. This enables us to determine if there are statistically significant mean differences in fund ownership, at various time points relative to the index inclusion.

The Mann-Whitney *U* test, or Wilcoxon rank-sum test, is a non-parametric method used to compare two independent groups (Callaert, 1999; Divine *et al.*, 2018; Mann & Whitney, 1947). It tests whether two samples likely originate from the same population, implying similar distributions. This test is particularly valuable when the normality assumption cannot be met (Callaert, 1999; Divine *et al.*, 2018). The null hypothesis implies no difference in distribution between the groups, while the alternative hypothesis indicates a difference (Callaert, 1999; Mann & Whitney, 1947). The test statistic is calculated as follows:

$$U_1 = n_1 \times n_2 + \frac{n_1(n_1 + 1)}{2} - R_1 \tag{2}$$

$$U_2 = n_1 \times n_2 + \frac{n_2(n_2 + 1)}{2} - R_2 \tag{3}$$

where:

- U_1 and U_2 are the test statistics for each group.
- R_1 and R_2 are the sums of ranks for each group.
- n_1 and n_2 are the sample sizes.

To determine the significance, the following z-score was calculated:

$$z = \frac{U - \mu_U}{\sigma_U} \tag{4}$$

where:

$$\mu_U = \frac{n_1 \times n_2}{2} \tag{5}$$

$$\sigma_U = \sqrt{\frac{n_1 \times n_2(n_1 + n_2 + 1)}{12}} \tag{6}$$

Considering that the fund ownership data does not follow a normal distribution, the Mann-Whitney *U* test served as a complementary analysis due to its non-parametric nature, allowing us to evaluate differences without assuming normality in the data distribution.

We structured our analysis to explore changes in fund ownership across all quarters surrounding the index inclusion. Since both tests require two samples for comparison, we designated one quarter as the baseline and compared it against subsequent quarters. Initially, the baseline was set at Q-2, with comparisons made for Q-1, Q0, Q+1, and Q+2. After completing the set of comparisons with Q-2 as the baseline, the baseline was shifted to Q-1, and the process was repeated, starting comparisons from Q0 onwards. This pattern continued for each quarter, moving the baseline sequentially through the timeline. This shift in the baseline quarter allowed us to cover all possible combinations, similar to moving along the upper diagonal of a correlation matrix. The process is illustrated in Table 3.1. Each pair of quarters was tested, ensuring a consistent analysis of how fund ownership varied over time relative to the event of index inclusion. This process was repeated for both Welch's *t*-test and the Mann-Whitney *U* test and conducted separately for index ownership and active ownership.

Table 3.1

Illustration of Welch's t-test & Mann-Whitney U Test Results Interpretation

		End (Comparison)				
		Q-2	Q-1	Q0	Q+1	Q+2
Start (Baseline)	Q-2	Q-2 vs Q-2	Q-2 vs Q-1	Q-2 vs Q0	Q-2 vs Q+1	Q-2 vs Q+2
	Q-1		Q-1 vs Q-1	Q-1 vs Q0	Q-1 vs Q+1	Q-1 vs Q+2
	Q0			Q0 vs Q0	Q0 vs Q+1	Q0 vs Q+2
	Q+1				Q+1 vs Q+1	Q+1 vs Q+2
	Q+2					Q+2 vs Q+2

Note. Illustration of how Welch's *t*-test and Mann-Whitney *U* test were implemented to test for statistical significance between the different quarters for both index and active fund ownership. Both tests require two samples. Here, the start represents the baseline sample, which is compared against the end representing the comparison sample. The diagonal tests are each sample against itself, resulting in a null result. This method ensures a consistent analysis of changes in fund ownership across all relevant time points relative to the index inclusion.

3.5 Panel Data Regression

To investigate the impact of index inclusion on the added stocks, we conducted a two-way fixed effect panel regression in Python (library: linearmodels). This approach considers both cross-sectional and time-series elements, utilizing data from 387 stocks (entities) over a sample period comprising five quarters: two quarters before and two quarters after the quarter of index inclusion. This aggregation resulted in 1,935 observations in total. Our objective was to assess how changes in index and active fund ownership influence stocks' return, risk, and pricing efficiency surrounding the event of stock inclusion.

To address the unbalanced nature of stock inclusions occurring at different times over the ten-year period, we drew inspiration from an event study setup. To balance the panel data, we established a pseudo timeline standardized relative to each inclusion event, encompassing two quarters before and after the announcement (Q-2, Q-1, Q0, Q+1, Q+2). This approach ensured that all stocks were evaluated within the same time frame, facilitating consistent assessment of the inclusion effect across the sample. According to Greene (2020), it is reasonable to treat unbalanced data as a characteristic of random sampling. We applied this consideration to the stocks within each quarter of the pseudo timeline. This uniform time index for each stock enabled the model to capture the effects of changes in independent variables on the dependent variables effectively. For all regressions, we divided the sample period into four separate periods: the full ten-year period (2013-2023), the first five years (2013-2018), the last five years (2018-2023), and the pandemic period (2020-2022). With six dependent variables (average daily stock return (arithmetic mean), average daily stock return (geometric mean), daily stock volatility, quarterly stock beta, variance ratio lag 2, variance ratio lag 5), we conducted a total of 24 regressions. This approach allowed us to systematically analyze the impact of index inclusion on stock performance across different periods and provided comprehensive insights into the relationship between fund ownership change and stock characteristics.

The rationale for employing a two-way fixed effects model over other panel regression models, such as first difference or random effects, lies in its ability to mitigate bias from omitted variables, that vary across entities but not over time, and vice versa (Brooks, 2019; Greene, 2020; Pesaran, 2015; Petersen, 2009; Wooldridge, 2001). The fixed effects model was chosen to control for unobserved heterogeneity within entities (stocks) and across time. In this study, the fixed entity effects are crucial, as they control for all time-invariant factors specific to each stock and remove the influence of intrinsic factors unique to each stock. Examples of such factors are corporate governance structures or strategic positioning, which are not captured by the observed variables, yet could affect the dependent variables (Brooks, 2019; Greene, 2020). The fixed time effects are equally important, as they adjust for factors that affect all stocks simultaneously during a specific quarter relative to the inclusion event. Examples of such are market-wide economic conditions or regulatory changes impacting the markets, in which these stocks operate (Brooks, 2019; Greene, 2020; Wooldridge, 2001).

To ensure that the data fulfilled the requirements for a two-way fixed effects model, and that it was a more suitable option than other panel regression models, we performed the following diagnostic tests: the variance inflation factor (VIF), the Breusch-Pagan test, Durbin-Watson statistics, and the Hausman specification test (Brooks, 2019; Greene, 2020; Wooldridge, 2001). Overall, the test outcomes displayed no multicollinearity nor autocorrelation for all dependent variables. There were some tendencies and instances of heteroscedasticity within certain sub-periods, and the Hausmann test favored the fixed-effect model over the random effect at either the 1% or 5% significance level. An additional note is that we had previously considered including trading volume as a dependent variable, however, it got rejected at this stage, due to being non-linear and displaying high levels of heteroscedasticity.

Given the panel structure of the dataset, with observations nested both within entities (stocks) and time (pseudo-quarters around the event), standard errors needed to be adjusted to account for potential correlations. Clustering standard errors is a method used to address this issue by allowing for arbitrary correlation within clusters while assuming independence (Greene, 2020; Pesaran, 2015; Petersen, 2009; Wooldridge, 2001). This configuration ensures that the standard errors are robust to correlations both within individual entities and across specific times, thereby providing a more reliable foundation for statistical inference (Greene, 2020; Henningsen & Haman, 2007; Pesaran, 2015; Petersen, 2009). Clustering standard errors in panels with a limited number of clusters–either by entity or time–can introduce biases, typically resulting in overly optimistic (downward-biased) standard errors (Greene, 2020; Petersen, 2009; Wooldridge, 2001).

This study employs clustering along two dimensions. First, entity clustering, where observations within the same stock are likely to be correlated. This correlation can arise from firm-specific attributes not captured by the model, such as management practices or strategic initiatives that affect the stock's performance across all periods. Clustering by entity adjusts the standard errors to account for this intra-stock correlation, ensuring that the inference remains valid even if the observations within each stock are not independent (Brooks, 2019; Greene, 2020; Wooldridge, 2001). Second, time clustering, which aligns the data's time dimensions. Due to the pseudo timeline and the event study setup of the analysis, the fixed time effect should hopefully capture shock patterns common to all entities, such as when the index inclusion announcement becomes public. Clustering by time also accounts for common shocks or reactions across all stocks at specific event-relative times, which could introduce correlation among the residuals (Brooks, 2019; Greene, 2020; Wooldridge, 2001).

3.5.1 Dependent Variables

In the two-way fixed effect panel regression, we have defined six dependent variables to empirically test our hypotheses. To test for stock return (H1), we test both average daily stock return via the arithmetic mean, and average daily stock return via the geometric mean. To test for stocks' risk (H2), we use daily stock volatility and quarterly stock beta. To test for the pricing efficiency of the stocks (H3), we apply the variance ratio test with two different lags: two- and five-day lags.

3.5.1.1 Average Daily Stock Return – Arithmetic Mean

The first dependent variable, which examines stock return (H1), is the arithmetic mean of daily stock returns. This metric is calculated as the percentage change of the daily closing prices, P_t, (Equation 8) within each quarter. For each quarter, we summarize all the percent returns and take the arithmetic mean, as displayed in Equation 7.

$$ArithmeticMeanReturn = \bar{R} = \frac{1}{T}\sum_{t=1}^{T} R_t \tag{7}$$

where:

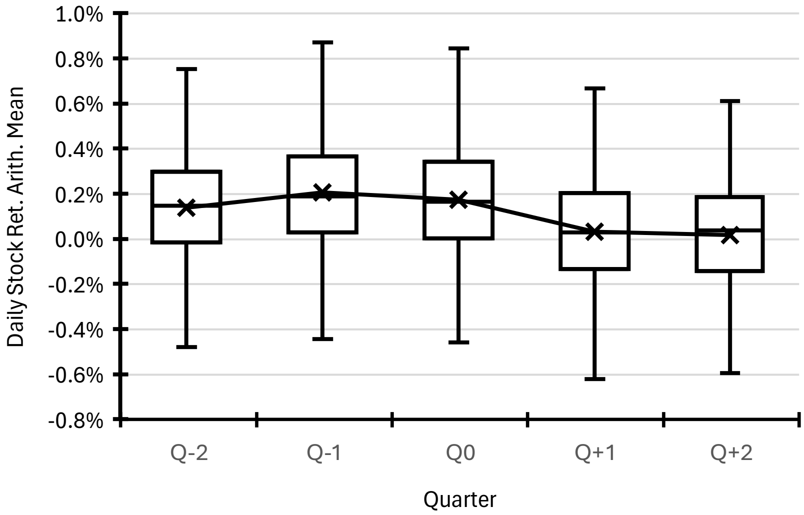
$$R_t = \frac{P_t}{P_{t-1}} - 1 \tag{8}$$

This calculation was performed for all quarters and all stocks, yielding one value for each stock within each quarter. The idea to use this metric was to see how the average daily return changes from one quarter to the next, and if there is a pattern of returns surrounding the index inclusion. Figure 3.3 presents the box plot of the calculated average daily stock returns for each quarter. For detailed descriptive statistics of the variable, over the quarters, see Appendix B, Table B.2, Panel A. The box plot indicates that stock returns were generally positive, yet fluctuated across different quarters, with the mean and median daily return fluctuating between 0.0% and 0.2%.

The plot demonstrates that the entire box, including the whiskers, increases from Q-2 to Q-1, dips slightly during the quarter of inclusion, and then declines after the inclusion. The mean return is slightly higher in Q-1, suggesting some momentum or anticipation effect before the index inclusion. The dip in returns during Q0 could indicate initial volatility or adjustment as the stock is included in the index. The subsequent decline after Q0 might reflect market stabilization, or reversion to the mean after the inclusion event.

Figure 3.3

Average Daily Stock Return (Arithmetic Mean)



Note. Box plot of the average daily stock returns (arithmetic mean) for the sample stocks across various quarters, where the 'x' symbolizes the mean, connected by a trendline. The median is represented by the middle line of the box plot, while the lower and upper bounds depict the 25th and 75th percentiles, respectively. The whiskers extend from the upper and lower bounds by ± 1.5*IQR. Each data point in the box plot represents one stock's daily return (arithmetic mean) for that quarter.

3.5.1.2 Average Daily Stock Return – Geometric Mean

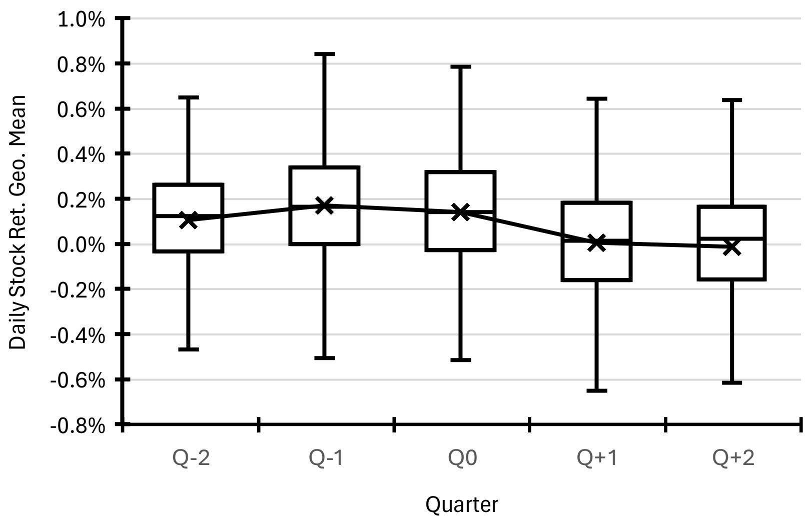
The second dependent variable, which tests for stock return (H1), is the geometric mean of daily stock return. It is calculated by taking the product of daily price return (P_t) across each quarter and then raising it to the power of one, over the number of trading days (T), as displayed in Equation 9.

$$GeometricMeanReturn = \left(\prod_{t=1}^{T}\left(\frac{P_t}{P_{t-1}}\right)\right)^{\frac{1}{T}} - 1 \tag{9}$$

Similar to the arithmetic mean return, this calculation is performed for all quarters and all stocks, yielding one value for each stock within each quarter. The idea to use this metric is to see how the geometric daily return changes from one quarter to the next, and if there is a pattern of returns surrounding the index inclusion. The rationale to examine the geometric mean in addition to the arithmetic mean, is that it captures the compounding effect of returns, providing additional insight into the growth of stock prices over time. This approach does more than simply assess if the daily return was higher; it gives an indication of whether the compounding effect was greater before, during, and after the index inclusion.

Figure 3.4

Average Daily Stock Return (Geometric Mean)



Note. Box plot of the daily geometric mean of stock returns for the sample stocks across various quarters, where the 'x' symbolizes the mean, connected by a trendline. The median is represented by the middle line of the box plot, while the lower and upper bounds depict the 25th and 75th percentiles, respectively. The whiskers extend from the upper and lower bounds by ± 1.5*IQR. Each data point in the box plot represents one stock's daily return (geometric mean) for that quarter.

Figure 3.4 presents the box plot of the calculated geometric daily stock returns for each quarter. For detailed descriptive statistics of the variable, over the quarters, see Appendix B, Table B.2, Panel B. The box plot indicates that stock returns were generally positive before the inclusion, with its mean and median between 0.10% and 0.20%. After the inclusion, the mean and median were around 0.0%. Similar to the arithmetic mean, the slight increase in mean return in Q-1 may suggest an anticipation effect or momentum, before the index inclusion. Also noticeable, is that the whiskers widen in the quarters of inclusion, meaning that there is a greater difference of geometric mean among the sample. However, just like the mean and median stabilizes after the inclusion, the range also narrows. This pattern suggests that the compounding effect of return diminishes after the inclusion effect, giving further support that index inclusion influences stock returns.

3.5.1.3 Daily Stock Volatility

The third dependent variable which examines stock risk (H2), is the daily volatility of stocks within each quarter. It is calculated by taking the population standard deviation of the daily percentage change in closing price within each quarter, illustrated in Equation 10. Note that R_t and $\bar{R}$ are defined in Equations 8 and 7, respectively.

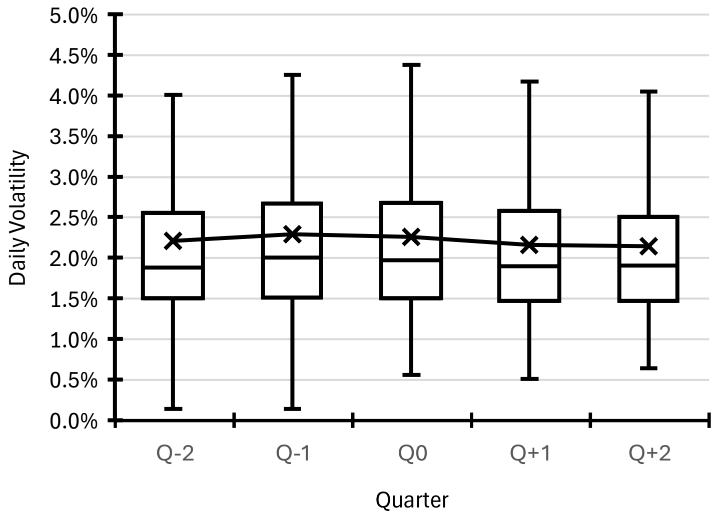
$$Daily\ Stock\ Volatility = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(R_t - \bar{R})^2} \tag{10}$$

This metric is computed for all quarters and stocks to assess volatility before, during, and after the index inclusion. The population standard deviation is chosen because it accounts for the full quarter's data, ensuring that all available data for each stock is included, rather than relying on a sample return size.

Figure 3.5 presents the box plot of the daily stock volatility for each quarter. For detailed descriptive statistics of the variable, over the quarters, see Appendix B, Table B.2, Panel C. The box plot indicates that daily stock volatility is relatively stable, at around 2.25%, although the minimum volatility increases after index inclusion. The median is lower than the mean, slightly below 2.00%, which suggests that the distribution is positively skewed. Moreover, the box fluctuates slightly while the whiskers vary more substantially. There is a clear increase in the minimum volatility after the inclusion, rising from slightly above 0.00% to 0.50%. This change indicates that the range of volatility becomes more concentrated toward the box, and the distribution becomes more positively skewed. Additionally, the range of the whiskers becomes smaller after the inclusion and remains at this level. This pattern suggests that while the mean and median volatilities do not change drastically, there is a reduction in extremely low volatility values, leading to a more consistent volatility range post-inclusion. The increased minimum volatility and reduced whisker range imply that stocks become less prone to very low volatility levels after being included in the index. This could be attributed to higher trading activity and the increased investor attention that often accompanies index inclusion. In short, Figure 3.5 suggests that index inclusion may influence stocks' volatility.

Figure 3.5

Daily Stock Volatility



Note. Box plot of the daily volatility for the sample stocks across various quarters, where the 'x' symbolizes the mean, connected by a trendline. The median is represented by the middle line of the box plot, while the lower and upper bounds depict the 25th and 75th percentiles, respectively. The whiskers extend from the upper and lower bounds by ± 1.5*IQR. Each data point in the box plot represents one stock's daily volatility.

3.5.1.4 Quarterly Stock Beta

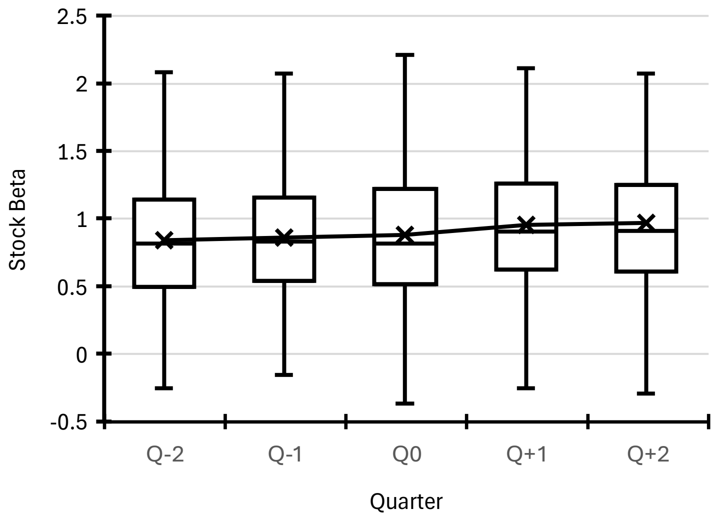
The fourth dependent variable, which examines stock risk (H2), is the quarterly stock beta. Stock beta is calculated by dividing the covariance of the stock returns with market return (the STOXX Europe 600 index in this case) by the market variance (Lakonishok & Shapiro, 1984; Merton, 1980), as shown in Equation 11.

$$Quarterly\ Stock\ Beta = \frac{Cov(R_i, R_m)}{Var(R_m)} \tag{11}$$

This beta is calculated for each stock and each quarter using the daily percent return of the stocks and the STOXX Europe 600, to calculate the covariance and variance used in Equation 11 (Lakonishok & Shapiro, 1984; Merton, 1980). This approach aligns the risk assessment with the specific period of interest, enhancing the relevance of the analysis, with respect to the index inclusion. Additionally, using stock beta allows for a more intuitive comparison of how the riskiness of the stock differs in relation to the market before, during, and after the index inclusion.

Figure 3.6

Quarterly Stock Beta



Note. Box plot of the beta-value for the sample stocks across various quarters, where the 'x' symbolizes the mean, connected by a trendline. The median is represented by the middle line of the box plot, while the lower and upper bounds depict the 25th and 75th percentiles, respectively. The whiskers extend from the upper and lower bounds by ± 1.5*IQR. Each data point in the box plot represents one stock's beta for that quarter.

Figure 3.6 presents the box plot of the calculated stock betas for each quarter. For detailed descriptive statistics of the variable, over the quarters, see Appendix B, Table B.2, Panel D. The box plot indicates that the mean and median stock beta increases from 0.8 to around 1.0 following index inclusion. This suggests that stocks become more correlated with the market after being included in the index, which aligns with expectations that the stocks are integrated into the index. Despite this, there is a wide dispersion around the mean, where the box ranges from 0.5 to 1.25, and the whiskers fluctuate from around -0.25 to above 2.0, where the biggest range occurs at the quarter of inclusion. The expansion of the box and whiskers during the quarter of inclusion reflects higher variability in beta values, likely due to increased trading activity and changes in investor behavior. The contraction before and after inclusion suggests a stabilization of beta values, indicating that the stocks' systematic risk, as measured by beta, aligns more closely with the market index post-inclusion. As beta measures the responsiveness of stocks' return-to-market movements, an increase in beta implies that the stocks might experience greater swings in response to market fluctuations after inclusion. The pattern in Figure 3.6 supports our hypothesis that index inclusion affects the systematic risk of stocks, potentially making them more sensitive to market movements.

3.5.1.5 Variance Ratio Lag 2 & 5

The fifth and sixth dependent variables, which examine the price efficiency (H3) of the stocks, are the variance ratio tests with two- and five-day lags (VR2 and VR5). The VR test focuses specifically on weak-form price efficiency (Fama, 1970; Lo & MacKinlay, 1988). According to Fama (1970), a market is considered efficient if prices "fully reflect" all available information for all market participants. This is consistent with the random walk hypothesis, suggesting that future prices are determined by today's price, p_{t-1}, and an expected change in price, μ, plus a random shock, ε_t. This shock represents new information not previously known to market participants and is expected to be independent and identically distributed (Fama, 1970; Lo & MacKinlay, 1988).

$$p_t = \mu + p_{t-1} + \varepsilon_t, \;\; \varepsilon_t \sim IID(0, Var(r_t)) \tag{12}$$

The benchmark for efficiency is a VR outcome equal to one, implying that price movements are uncorrelated over time and fully reflect all publicly available information (Lo & MacKinlay, 1988). Thus, any deviation from this value indicates a departure from the random walk benchmark (Coles *et al.*, 2022; Lo & MacKinlay, 1988). The VR test utilizes a q-period bias-corrected formula developed by Lo and MacKinlay (1988), presented in Equation 13. This formula calculates the ratio of the variance of log return, r_t, with q lags (Equation 15) to q times the variance of log return (Equation 16), with a correction term for autocorrelation, ρ_k (defined in Equation 18). If the correlation is 0 for all lags, then $VR(q) = 1$, indicating a random walk.

$$VR(q) = \frac{Var(r_t(q))}{q \times Var(r_t)} = 1 + 2\sum_{k=1}^{q-1}\left(1 - \frac{k}{q}\right) \times \rho_k \tag{13}$$

$$\rho_k = 0 \; for \; all \; k \; if \; random \; walk \;\; \Rightarrow VR(q) = 1 \tag{14}$$

where:

$$Var\big(r_t(q)\big) = \; q \times Var(r_t) + \sum_{i=0}^{q-1}\sum_{j \neq i}^{q-1} cov(\varepsilon_{t-i}, \varepsilon_{t-j}) \tag{15}$$

$$Var(r_t) = \frac{1}{T}\sum_{t=1}^{T}(\varepsilon_t - \hat{\mu})^2 \tag{16}$$

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T}(\varepsilon_t) \tag{17}$$

$$\rho_k = \frac{Cov(r_t, r_{t+k})}{\sqrt{Var(r_t)} \times \sqrt{Var(r_{t+k})}} \tag{18}$$

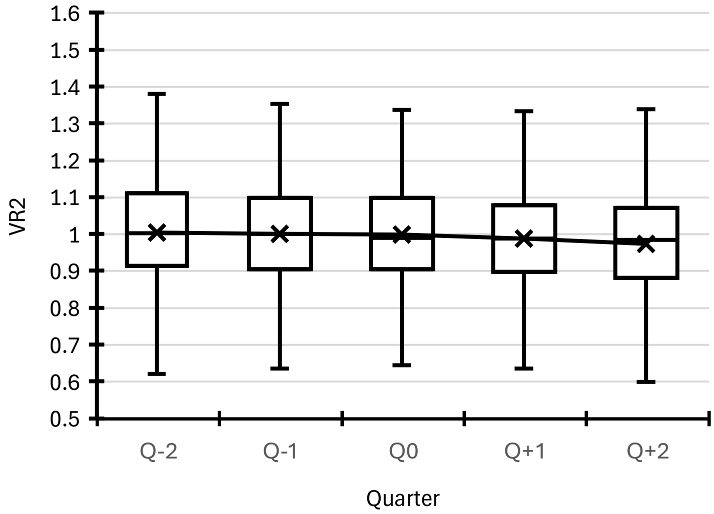
$$r_t = \log(1 + R_t) = \log\left(\frac{P_t}{P_{t-1}}\right) \tag{19}$$

We calculated the VR test for each firm and all quarters surrounding the stocks' index inclusion, setting the number of lags (q), to two (VR2) and five (VR5) trading days. This dual approach allowed us to assess pricing efficiency over both short periods and longer periods.

The VR2 assesses market efficiency over short periods, specifically two trading days. The choice of two lags is to evaluate how short-term price efficiency varies before, during, and after the index inclusion. Figure 3.7 presents the box plot of all the calculated VR2 for each quarter. For detailed descriptive statistics of the variable, over the quarters, see Appendix B, Table B.2, Panel E. The plot reveals that the mean and median VR2 are fairly consistent around one, suggesting that on average, the stock prices exhibit characteristics of a random walk over a two-day period, consistent with weak-form efficiency (Fama, 1970; Lo and MacKinlay, 1988). However, the whiskers fluctuate slightly, displaying a wide range from approximately 0.6 to 1.4, indicating significant variability among individual stocks. This variability suggests that while the overall market may be efficient, on average, in the short term, individual stocks can experience periods of inefficiency. Although this range deviates from one, it seems to stay consistent across all five quarters, which indicates no clear inclusion effect. Despite this, VR2 is chosen as a dependent variable in this analysis to see if the difference in pricing efficiency on the entity level is still captured by the model.

Figure 3.7

Variance Ratio Lag 2

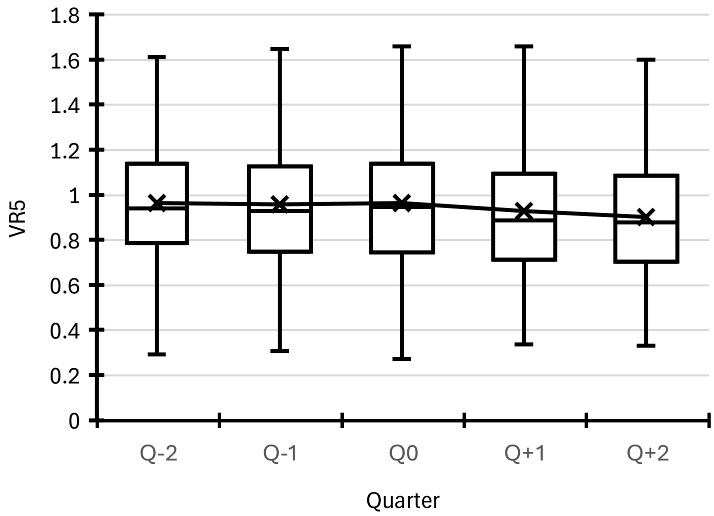


Note. Box plot of the variance ratio test with two days lag (VR2) for the sample stocks across various quarters, where the 'x' symbolizes the mean, connected by a trendline. The median is represented by the middle line of the box plot, while the lower and upper bounds depict the 25th and 75th percentiles, respectively. The whiskers extend from the upper and lower bounds by ± 1.5*IQR. Each data point in the box plot represents one stock's VR2 for that quarter.

The VR5 provides insights into market efficiency over a longer period, covering a full trading week. This calculation is particularly valuable for understanding changes in price efficiency over longer intervals around the index inclusion and serves as a complement to the short-term measure of VR2. Figure 3.8 presents the box plot of all the calculated VR5 for each quarter. For detailed descriptive statistics of the variable, over the quarters, see Appendix B, Table B.2, Panel F. The box plot reveals a greater degree of variability compared to VR2, where the whiskers extend from approximately 0.3 to 1.6 throughout all quarters. The mean and median values remain centered around one and, like VR2, suggest that on average the stock prices exhibit weak-form efficiency over a trading week. However, the mean and median fluctuate more than the VR2, which indicates that prices are less stable for longer periods compared to shorter periods. In addition, the box, *i.e.*, the 25th and 75th percentile displays a wider range between approximately 0.75 to 1.2, which further supports that the VR5 contains an increased variability. This may suggest that pricing efficiency becomes more susceptible to inefficiency as the lag increases and becomes more sensitive to external influences. Given this, the VR5 is chosen as the final dependent variable and serves as a complement to the VR2 to examine pricing efficiency.

Figure 3.8

Variance Ratio Lag 5



Note. Box plot of the variance ratio test with five days lag (VR5) for the sample stocks across various quarters, where the 'x' symbolizes the mean, connected by a trendline. The median is represented by the middle line of the box plot, while the lower and upper bounds depict the 25th and 75th percentiles, respectively. The whiskers extend from the upper and lower bounds by ± 1.5*IQR. Each data point in the box plot represents one stock's VR5 for that quarter.

3.5.2 Two-Way Fixed Effects Model

In the two-way fixed effect panel regression, we define (1) index fund ownership and (2) active fund ownership, as our two independent variables for all regressions. Based on the assumptions and all dependent variables listed in Section 3.5.1, the two-way fixed effects model is defined as follows:

$$y_{i,t} = \beta_1 Active_{i,t} + \beta_2 Index_{i,t} + u_i + v_t + \varepsilon_{i,t} \tag{20}$$

where:

- $y_{i,t}$ represents the dependent variable for each stock i at each quarter time t, capturing the chosen financial metric: average daily stock return (arithmetic mean), average daily stock return (geometric mean), daily stock volatility, quarterly stock beta, VR2, and VR5.
- $Active_{i,t}$ and $Index_{i,t}$ are the independent variables in all regressions, represent active and index fund ownership, respectively, both in decimal form.
- β_1 and β_2 represent the coefficients for $Active_{i,t}$ and $Index_{i,t}$, respectively.
- u_i represents the entity fixed effect, which captures all cross-sectional influences that are time-invariant, accounting for intrinsic characteristics of each stock, such as sector influences or company-specific attributes.
- v_t represents time fixed effect, which captures all factors that affect $y_{i,t}$ that vary over time (the five quarters) yet are constant across all entities. This includes macroeconomic conditions, market-wide shocks, inclusion effect, and trends that affect all stocks at specific times but are not directly observed.
- $\varepsilon_{i,t}$ is the idiosyncratic error term, which varies over time and entities and captures all the unexplained variability in $y_{i,t}$. $\varepsilon_{i,t}$ is assumed to be normally distributed, encompassing random noise and measurement errors.
- $i = 1, \dots, N$ indexing the entities (stocks), where $N = 387$.
- $t = 1, \dots, T$ indexing time (quarters), where $T = 5$.

Note: The traditional regression intercept was omitted, due to redundancy within the fixed effect model, where the fixed entity and time effects adjust for baseline levels (Brooks, 2019; Greene, 2020; Wooldridge, 2001). This omission avoids unnecessary parameter estimation and focuses analysis on the influence of the variables of interest.

4 Results & Discussion

This section presents the outcomes from the two-way fixed effects panel regressions examining the effect on stocks' return, risk, and pricing efficiency, in relation to active and passive fund ownership. Initially, we identify changes in fund ownership following the inclusion in the index. Next, we investigate the effect of index inclusion on stocks' return, risk, and pricing efficiency. We report results for all four periods that we perform our tests on. For all results, we relate them to the previous research and our initial hypothesis. Finally, we discuss the possible limitations of our method.

4.1 Effect on Fund Ownership

Table 4.1 displays results from Welch's *t*-test and the Mann-Whitney *U* test in Panels A and B, respectively. Both tests assess significant changes in passive funds ownership from one quarter to another. Panel A indicates significant changes in Welch's *t*-test from Q-2 and Q-1 to Q0, Q+1, and Q+2, although not all quarters show significance. A clear pattern of significant changes surrounds the inclusion quarter, Q0, which is generally significant at the 1% level, up to and following the quarter of inclusion. Quarters that do not show any significance are primarily those before the inclusion, Q-2 to Q-1, and after the inclusion, Q0 to Q+1 and Q+1 to Q+2. These results confirm a significant change in index fund ownership following inclusion in the STOXX Europe 600 index. Similar findings are reported from the Mann-Whitney *U* test in Panel B, with significant results occurring from Q-2 and Q-1 to Q0, Q+1, and Q+2. The main difference is that the change from Q-1 to Q0 is now significant at the 1% level (compared to 5% with Welch's *t*-test), further confirming a significant increase in index fund ownership post-inclusion.

Table 4.2 shows results for active fund ownership changes, derived using both Welch's *t*-test and the Mann-Whitney *U* test, and reveals similar trends to index ownership, yet with some differences. From the results in Panel A, significant changes occur from Q-2 and Q-1 to later quarters, *i.e.*, Q0, Q+1, and Q+2. While not all quarters show significant changes, there is a noticeable pattern of significant changes surrounding the inclusion quarter, as in Table 4.1. These findings corroborate that there was a significant ownership change in active funds, following the stocks' inclusion in the STOXX Europe 600 index. Similar patterns are evident from the Mann-Whitney *U* test results, presented in Panel B, where significant shifts occur for similar transitions. The main distinction is that the transition from Q-1 to Q0 is now significant at the 1% level, compared to the 5% level with Welch's *t*-test, reinforcing the finding of a significant increase in active funds ownership post-inclusion.

Table 4.1

Welch's t-test & the Mann-Whitney U Test on Index Fund Ownership Change

Panel A: Welch's *t*-test						
Index		End				
		Q-2	Q-1	Q0	Q+1	Q+2
Start	Q-2	0	1.0330	3.6100***	4.7381***	5.6106***
	Q-1		0	2.5566**	3.6835***	4.5528***
	Q0			0	1.1445	2.0226**
	Q+1				0	0.8761
	Q+2					0

Panel B: Mann-Whitney *U* test						
Index		End				
		Q-2	Q-1	Q0	Q+1	Q+2
Start	Q-2	75660.5	79419.5	89166.5***	93122.0***	96024.0***
	Q-1		75660.5	85716.5***	89733.5***	92719.5***
	Q0			75660.5	79848.5	83257.0**
	Q+1				75660.5	79009.5
	Q+2					75660.5

Note. Welch's *t*-test in Panel A and the Mann-Whitney *U* test in Panel B assess significant changes in index fund ownership between quarters, denoted by "Q" and the respective quarter numbers. Q0 indicates the quarter of the stock's inclusion in the index, with the surrounding quarters (Q-2, Q-1, Q+1, Q+2) representing two quarters before, and after the inclusion event. Here, start represents the starting quarter, or baseline sample, which is compared against the end quarter, or comparison sample. The diagonal tests are each sample against itself, resulting in a null result. ***, **, and *, indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 4.2

Welch's t-test & the Mann-Whitney U Test on Active Fund Ownership Change

Panel A: Welch's *t*-test						
Index		End				
		Q-2	Q-1	Q0	Q+1	Q+2
Start	Q-2	0	0.7393	1.7378*	2.3817**	2.5589**
	Q-1		0	1.0110	1.6631*	1.8430*
	Q0			0	0.6557	0.8370
	Q+1				0	0.1816
	Q+2					0
Panel B: Mann-Whitney *U* test						
Index		End				
		Q-2	Q-1	Q0	Q+1	Q+2
Start	Q-2	75660.5	78668.0	82310.5**	84668.0***	85285.5***
	Q-1		75660.5	79272.5	81692.5*	82314.5**
	Q0			75660.5	78158.0	78799.0
	Q+1				75660.5	76347.5
	Q+2					75660.5

Note. Welch's *t*-test in Panel A and the Mann-Whitney *U* test in Panel B assess significant changes in active fund ownership between quarters, denoted by "Q" and the respective quarter numbers. Q0 indicates the quarter of the stock's inclusion in the index, with the surrounding quarters (Q-2, Q-1, Q+1, Q+2) representing two quarters before, and after the inclusion event. Here, the start represents the starting quarter, or baseline sample, which is compared against the end quarter, or comparison sample. The diagonal tests are each sample against itself, resulting in a null result. ***, **, and *, indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

The results from Welch's *t*-test and the Mann-Whitney *U* test both indicate that there was a significant change in index ownership from the two quarters before the index inclusion to the quarters after, as well as the inclusion quarter itself. These results are in line with our hypothesis and confirm that there was a significant increase in index ownership after being included in the index. The increase in index ownership is also found by Coles *et al.* (2022). The increase is in their research design, as they mention, needed to investigate other effects of increased index fund ownership. In other words, their methodology relies on index ownership only increasing (while active does not), to be able to see the effect it has on different variables. In contrast to Coles *et al.* (2022), we find that active fund ownership also increases after the index inclusion. The change is, however, not as significant as the index ownership. Opposite to us, Coles *et al.* (2022) find that as the passive ownership increases, that nets out against the active ownership decreasing. There is a slight difference in methodology between our studies, though, whereby it is necessary for Coles *et al.* (2022) to have a significant change, while it is not for our method. The rationale is that Coles *et al.* (2022) investigate specific variables (*e.g.*, variance ratio or market beta), before and after index inclusion, using a difference-in-difference regression, while we use fund ownership as an independent variable. For Coles *et al.* (2022), it is a necessity for index ownership to increase over the quarters, which differs from our design. Our methodology, differently, does not rely on a significant change, *i.e.*, an increase, in index ownership since we will still be able to investigate the effect.

Even though active ownership increases, which may seem illogical (given the inclusion to an index), there are several plausible reasons as to why. An important point to state is that both index fund ownership and active fund ownership do not make out 100% of the stock ownership, that we use. Rather, it represents roughly 30% (see Table B.1, Appendix B) of the total ownership of the shares. As such, it is reasonable that both ownership types can increase. One potential reason that active ownership increases after the inclusion is that some active funds may only be allowed to buy stocks of a certain size. As a stock gets included in an index, *i.e.*, the STOXX Europe 600, this is likely a result of the stock having increased in market capitalization. Hence, this provides a possible explanation for why both fund ownership types increase.

4.2 Hypothesis 1: Effect on Returns

Table 4.3 presents the results from the panel regression with two-way fixed effects on the daily stock returns of the included stocks, analyzing the impact of index and active fund ownership. As is shown, there are significant values for both index and active ownership. Most of the active parameters are significant (at a high significance level) except for the period 2013-2018, while only the pandemic years are significant for the index ownership (also at a high level). Interestingly, the sign change between active and index ownership, as active is positive, while the index parameter is negative. This implies that as active (index) ownership increases, then the return increases (decreases). The change of sign is notable, although it should not be viewed that decisively, since only one index parameter is significant – and during the shortest period, *i.e.*, the pandemic years.

Table 4.3

Panel Regression of Average Daily Stock Return & Fund Ownership

Panel Regression				
Dependent Variable:	Average Daily Stock Return (Arithmetic Mean)			
Period:	2013-2023	2013-2018	2018-2023	2020-2022
Active	0.0079***	0.0069	0.0107***	0.0108**
	(3.6624)	(1.5847)	(3.9602)	(2.0643)
Index	-0.0093	-0.0100	-0.0103	-0.0464***
	(-0.9537)	(-0.9081)	(-0.8719)	(-2.8281)
Fixed time effect	Yes	Yes	Yes	Yes
Fixed entity effect	Yes	Yes	Yes	Yes
Observations	1935	820	1115	735
Entities	387	164	223	147
R^2 (within)	0.0012	0.0003	0.0030	0.0307

Note. The table shows the results from the two-way fixed effects panel regression of the average daily stock return on active and passive fund ownership. The estimating equation is:

$$ArithmeticMeanReturn_{i,t} = \beta_1 Active_{i,t} + \beta_2 Index_{i,t} + u_i + v_t + \varepsilon_{i,t}$$

Standard errors are robust and clustered by entity and time. Entities reflect the number of stocks, while observations represent the number of entities times five quarters. In, total there are four regressions, performed at the four time periods seen above. For each result, the parameter value is recorded with the *t*-statistics shown in parentheses. ***, **, and *, indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 4.4 also shows the regression output for the stock return on fund ownership types using the geometric mean return. Visible from the table, all values for active funds are significant, at either the 1%, 5%, or 10% levels. However, for index fund ownership, only the period 2020-2022 is significant at the 1% level. Notably, the parameter signs are positive for all active parameters and negative for all index parameters, indicating that an increase in active fund ownership is associated with higher returns, whereas an increase in index ownership correlates with negative returns.

Table 4.4

Panel Regression of Daily Geometric Mean Stock Returns & Fund Ownership

Panel Regression				
Dependent Variable:	Average Daily Stock Return (Geometric Mean)			
Period:	2013-2023	2013-2018	2018-2023	2020-2022
Active	0.0078***	0.0071*	0.0100***	0.0102**
	(3.4887)	(1.6730)	(3.9961)	(2.0007)
Index	-0.0123	-0.0099	-0.0154	-0.0489***
	(-1.4624)	(-0.9401)	(-1.3505)	(-2.9851)
Fixed time effect	Yes	Yes	Yes	Yes
Fixed entity effect	Yes	Yes	Yes	Yes
Observations	1935	820	1115	735
Entities	387	164	223	147
R^2 (within)	0.0036	0.0002	0.0056	0.0324

Note. The table shows the results from the two-way fixed effects panel regression of the daily geometric mean stock return on active and passive fund ownership. The estimating equation is:

$$GeometricMeanReturn_{i,t} = \beta_1 Active_{i,t} + \beta_2 Index_{i,t} + u_i + v_t + \varepsilon_{i,t}$$

Standard errors are robust and clustered by entity and time. Entities reflect the number of stocks, while observations represent the number of entities times five quarters. In, total there are four regressions, performed at the four time periods seen above. For each result, the parameter value is recorded with the *t*-statistics shown in parentheses. ***, **, and *, indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

By comparing the regression outputs of the arithmetic mean stock return to the geometric mean, as the dependent variable, the results are nearly identical. This is no surprise given that it still investigates the return, only with a different way of calculating it. The only difference, in terms of significance, is that the geometric mean return is also significant for the first five years, 2013-2018, although only at the 10% level. The signs are still the same, *i.e.*, active is positive while the index is negative. These results contradict our initial hypothesis that index ownership should lead to an increase in return, due to the automatic buying of the individual stock constituents. Why active ownership is associated with positive returns is hard to explain. One possible explanation is that there will be an upward pressure on the price, which will translate into positive returns, since active funds can now buy certain stocks that were not possible before, given that they were too small. The reasoning is similar to active fund ownership increasing after the index inclusion, given that they are now allowed to hold the stock, which they were not before.

In relation to previous studies, it is hard to draw any strong conclusions since few previous studies separate between index and active fund ownership, as we do. Instead, they study institutional ownership as a whole (see *e.g.*, Chang *et al.* (2015) and Vijh and Wang (2022)), which may contain both active and index funds. Both Chang *et al.* (2015) and Vijh and Wang (2022) document a positive relationship between increased institutional ownership and returns. This is in line with our finding for active fund ownership, counting active funds as part of institutional ownership. Oppositely, index ownership shows a negative relationship for us, yet should not be viewed as credible due to the lack of significance. While our findings are, partly, in line with Vijh and Wang (2022) and Chang *et al.* (2015), our results differ from Chen *et al.* (2004) and Patel and Welch (2017). Both Chen *et al.* (2004) and Patel and Welch (2017) find a negative relationship between institutional ownership and stock returns.

4.3 Hypothesis 2: Effect on Risk

Table 4.5 presents the regression outcomes of daily stock volatility against active and index fund ownership. The regression of daily volatility on active and index fund ownership shows some significant results, although only valid for index ownership. Over the whole period, 2013-2023, there is a significant increase (5% level) in volatility as the index fund ownership increases. The same pattern holds for the latter half of the period, 2018-2023, though even more pronounced. An increase in volatility is also discernible over the pandemic years, however, it is lower in magnitude and less significant. No significant result is shown for the active funds ownership. The increase in volatility after index inclusion is different from our initial hypothesis. Our reason was that as index, or passive, ownership goes up, then the active ownership must go down. Given fewer active investors, that should trade more often than passive, the volatility should decrease. Reasons for the increase could be that index-tracking funds buy the shares at weights that differ, at regular intervals. This will, then, lead to increased price disturbance, and thus, volatility. This hypothesis is in line with the findings of Sullivan and Xiong (2012).

Table 4.5

Panel Regression of Daily Stock Volatility & Fund Ownership

Panel Regression				
Dependent Variable:	Daily Stock Volatility			
Period:	2013-2023	2013-2018	2018-2023	2020-2022
Active	-0.0024	-0.0131	0.0195	0.0200
	(-0.2739)	(-1.4338)	(1.0842)	(0.6536)
Index	0.1047**	0.0262	0.1564**	0.0902*
	(2.0620)	(0.6476)	(2.2794)	(1.6956)
Fixed time effect	Yes	Yes	Yes	Yes
Fixed entity effect	Yes	Yes	Yes	Yes
Observations	1935	820	1115	735
Entities	387	164	223	147
R^2 (within)	0.0081	0.0037	0.0437	0.0024

Note. The table shows the results from the two-way fixed effects panel regression of the average daily stock volatility on active and passive fund ownership. The estimating equation is:

$$Volatility_{i,t} = \beta_1 Active_{i,t} + \beta_2 Index_{i,t} + u_i + v_t + \varepsilon_{i,t}$$

Standard errors are robust and clustered by entity and time. Entities reflect the number of stocks, while observations represent the number of entities times five quarters. In, total there are four regressions, performed at the four time periods seen above. For each result, the parameter value is recorded with the *t*-statistics shown in parentheses. ***, **, and *, indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 4.6 details the regression results of stock betas against fund ownership. For index ownership, the significant periods for volatility exactly correspond to those for stock beta; however, the significance is higher. Both the 2013-2023 and 2018-2023 periods are now significant at the 1% level, while the pandemic period is significant at the 5% level. Interestingly, all parameter signs are positive (as with the volatility), indicating that an increase in index fund ownership is associated with an increase in the stocks' beta. Even though there are no significant values for active ownership, the parameter values for active ownership are also positive, in line with the index ownership. Regarding the parameter signs, index fund ownership shows an increasing trend over time, where the latter five years have a higher parameter value than the whole period. Notably, the pandemic years, 2020-2022, record the highest value, indicating that the beta was the most sensitive during this period, as stocks got included in the index.

Table 4.6

Panel Regression of Stock Beta & Fund Ownership

	Panel Regression			
Dependent Variable:	Quarterly Stock Beta			
Period:	2013-2023	2013-2018	2018-2023	2020-2022
Active	0.3994	0.6504	0.1362	0.6385
	(1.0683)	(1.2124)	(0.2087)	(0.4701)
Index	3.7532***	0.1960	6.7190***	7.9332**
	(3.7845)	(0.1913)	(4.1682)	(2.2182)
Fixed time effect	Yes	Yes	Yes	Yes
Fixed entity effect	Yes	Yes	Yes	Yes
Observations	1935	820	1115	735
Entities	387	164	223	147
R^2 (within)	0.0288	0.0124	0.0477	0.0471

Note. The table shows the results from the two-way fixed effects panel regression of the stock beta on active and passive fund ownership. The estimating equation is:

$$Beta_{i,t} = \beta_1 Active_{i,t} + \beta_2 Index_{i,t} + u_i + v_t + \varepsilon_{i,t}$$

Standard errors are robust and clustered by entity and time. Entities reflect the number of stocks, while observations represent the number of entities times five quarters. In, total there are four regressions, performed at the four time periods seen above. For each result, the parameter value is recorded with the *t*-statistics shown in parentheses. ***, **, and *, indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

The regression of stock beta on fund ownership results is, as expected, similar to the regression on volatility. The significant parameters for the volatility are the same as for beta, *i.e.*, only index ownership and all parameters except the full period: 2013-2018. The difference is that it is slightly more significant: the 10% level for the volatility is now at the 5% level, and the same for the 5% level before, which is now at the 1% level. All parameters are positive, again, similar to volatility. The values are, interestingly, highest for the pandemic years, and then, in second, the latter five years: 2018-2023. A possible explanation for the pandemic years might be that the volatility of the index increased and that the comovement also went up (for instance, at the start of the outbreak when practically every stock dropped substantially). This differs from a "normal" market condition, in which stocks should move more freely in relation to the market index.

With regards to previous research, we find the same result as Ben-David *et al.* (2018), Baltussen *et al.* (2019), Coles *et al.* (2022), and Wang and Xu (2019). Ben-David *et al.* (2018), that has a similar methodology to our study, although they investigate the U.S. market during an earlier time frame, find that increased passive investing results in higher volatility in individual stocks. Coles *et al.* (2022) first predict that the idiosyncratic risk should increase in their theoretical model and then verify it in their empirical tests. Chabakauri and Rytchkov (2021), and Ahn and Patathoukas (2022) find the opposite relationship; namely, that the volatility decreases instead. This highlights the differing relationship between stocks' risk and index ownership. Possible explanations are that different articles perform their tests on different markets and during different time frames. Especially, the latter explanation should have an effect, given the rapid increase in passive investing during the last years.

Our finding that the stock beta goes up as index investors increase is in line with Da and Shiva (2018). As most other previous research, the authors study the U.S. market, and they use ETFs rather than index fund ownership, as we do. Still, they find a positive relationship between excess comovement of the stocks in relation to the market and passive ownership. This result is further corroborated by Baruch and Zhang (2022) and Sullivan and Xiong (2012). A slight difference, though, is that Baruch and Zhang (2022) test directly for correlation to the market and not beta. Even more evidence of a positive relationship is provided by Staer and Sottile (2018), in their tests of return comovement in relation to passive investing. As such, our result of a positive relationship between beta and passive investing confirms the findings of previous research.

4.4 Hypothesis 3: Effect on Price Efficiency

Table 4.7 presents the results from the regression using the variance ratio test with a two-day lag (VR2) against the index and active fund ownership. As is visible, there is no clear pattern in significance or parameter values. Active fund ownership is significant for the 2018-2023 period, at the 10% level, while index ownership is significant for the 2013-2018 period, also at the 10% level. Beyond these findings, there are no other significant values. The sign for the significant index ownership parameter is positive, indicating that the VR2 increases as index fund ownership increases. For active ownership, the significant value is also positive. It is hard to make any inference from this given that it is only two parameters that are weakly significant.

Table 4.7

Panel Regression of VR2 & Fund Ownership

Panel Regression				
Dependent Variable:	Variance Ratio Lag 2			
Period:	2013-2023	2013-2018	2018-2023	2020-2022
Active	-0.1747	-0.1692	0.2778*	-0.1015
	(-1.0177)	(-0.7060)	(-1.7335)	(-0.4660)
Index	0.6188	1.0654*	0.2490	0.2441
	(1.2865)	(1.6981)	(0.6188)	(0.3311)
Fixed time effect	Yes	Yes	Yes	Yes
Fixed entity effect	Yes	Yes	Yes	Yes
Observations	1935	820	1115	735
Entities	387	164	223	147
R^2 (within)	0.0002	0.0056	0.0037	0.0011

Note. The table shows the results from the two-way fixed effects panel regression of the variance ratio with two lags (VR2) on active and passive fund ownership. The estimating equation is:

$$VR2_{i,t} = \beta_1 Active_{i,t} + \beta_2 Index_{i,t} + u_i + v_t + \varepsilon_{i,t}$$

Standard errors are robust and clustered by entity and time. Entities reflect the number of stocks, while observations represent the number of entities times five quarters. In, total there are four regressions, performed at the four time periods seen above. For each result, the parameter value is recorded with the *t*-statistics shown in parentheses. ***, **, and *, indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

In Table 4.7, both the significant parameters are positive, meaning that the VR value increases as the fund ownership increases. In itself, this is not that informative since we are interested if the market is moving towards efficiency, or inefficiency, as the ownership type increases. This can be seen if the VR test moves toward one, which is in line with a random walk, implying an efficient market. To test for this, we may compare the regression outputs with the timeline of the VR test in Figure 1.1, which shows how the VR2 and VR5 have developed over the sample period (not exactly the same sample period, as Figure 1.1 covers half a year later, *i.e.*, 2014-2023). Note that the VR test in Figure 1.1 does not use the STOXX Europe 600 index, but a broad ETF covering the whole European stock market. Still, it may provide a guideline of the trend of the VR test over the time period. During 2013-2018, when index ownership is significant, the VR2 in Figure 1.1 both increased and decreased, which makes it difficult to extrapolate any clear trend. For 2018-2023, the case is similar, as there is no apparent trend towards either efficiency or inefficiency. As such, it is hard to draw any conclusions from the VR2 regression.

Table 4.8

Panel Regression of VR5 & Fund Ownership

Panel Regression				
Dependent Variable:	Variance Ratio Lag 5			
Period:	2013-2023	2013-2018	2018-2023	2020-2022
Active	-0.1594	-0.1866	-0.1893	-0.1248
	(-0.6993)	(-0.5075)	(-0.7025)	(-0.2859)
Index	0.3804	1.2768	-0.4494	-0.8607
	(0.4496)	(1.2267)	(-0.4586)	(-0.4230)
Fixed time effect	Yes	Yes	Yes	Yes
Fixed entity effect	Yes	Yes	Yes	Yes
Observations	1935	820	1115	735
Entities	387	164	223	147
R^2 (within)	7.817e-06	3.32e-05	0.0030	0.0045

Note. The table shows the results from the two-way fixed effects panel regression of the variance ratio with five lags (VR5) on active and passive fund ownership. The estimating equation is:

$$VR5_{i,t} = \beta_1 Active_{i,t} + \beta_2 Index_{i,t} + u_i + v_t + \varepsilon_{i,t}$$

Standard errors are robust and clustered by entity and time. Entities reflect the number of stocks, while observations represent the number of entities times five quarters. In, total there are four regressions, performed at the four time periods seen above. For each result, the parameter value is recorded with the *t*-statistics shown in parentheses. ***, **, and *, indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 4.8 presents the results from the regression of the variance ratio test with a five-day lag (VR5) on fund ownership types. As is visible, there are no significant values at any significance level, for either index or active ownership. As such, the results deviate slightly from the VR2 regression, where two values are significant.

In relation to previous research, our results are in line with Coles *et al.* (2022). They also fail to find a connection between index ownership and pricing efficiency, as seen by the VR test, and, generally, receive insignificant results. Coles *et al.* (2022) even utilize more efficiency metrics, such as the mispricing measure of Stambaugh *et al.* (2015), yet still find no meaningful results. Israeli *et al.* (2017), Chen *et al.* (2024), and Glosten *et al.* (2021) are, however, able to show a significant connection between pricing efficiency and index investing. Israeli *et al.* (2017) and Chen *et al.* (2024) find a negative relationship, while Glosten *et al.* (2021), oppositely, find a positive one. Noteworthy, there are some differences between the methods and the market studied between their studies and ours. Still, it provides evidence that there might exist a link, even though we do not find it.

With regards to the articles that only rely on theoretical models, Jaquart *et al.* (2023), find, opposite to us, that the pricing efficiency increases as the active ownership increases. Bond and Garcia (2022) also differ from our findings, as they show that the relative efficiency of stocks will go up as index ownership increases. Baruch and Zhang (2022) also differ from our findings, since they show a negative relationship between increased index investing and the pricing efficiency of stocks. Moreover, Buss and Sandaresan (2023) find a positive relationship, although their efficiency metric is slightly different to ours. In total, this underlines the differing findings by previous studies. Furthermore, it shows that the majority of previous research is able to find a connection between passive investing and stocks' pricing efficiency, while we do not.

The absence of significant results for the VR2 and VR5 regressions contradicts our hypothesis that as the index ownership increases, the stocks' pricing mechanism should become more inefficient. The rationale for this is that with fewer active investors that (actively) exploit arbitrage opportunities, and, hence, add value to the price discovery process, the pricing should be less efficient. Possible reasons for our non-results are hard to specify, although a potential reason is that index fund ownership still constitutes a relatively small share of the total ownership of the stocks (see Appendix B, Table B.1). As index investing grows even larger, following the current long-term trend, it is possible that a more pronounced relationship to efficiency will materialize.

4.5 Limitations

This thesis focuses on how the rise of index fund ownership influences stocks' return, risk, and pricing efficiency within the European equity markets. It employs a two-way fixed effect panel regression and tests six different dependent variables against index and active fund ownership. However, it is recognized that other variables influence these dependent variables, meaning that index and active fund ownership only explain a small part of the variation in these dependent variables. This can be seen from the low R-squared (within) values we receive from all the regressions. Furthermore, it is acknowledged that other relevant variables could have been used as dependent variables, such as trading volume (used by *e.g.*, Coles *et al.*, 2022), or excess return defined by some asset pricing model (used by *e.g.*, Ben-David *et al.*, 2018).

Another limitation of our research is the omission of sample data due to various factors. Among them are expired data, invalid ticker symbols, and data that did not meet the requirements of having two full quarters of price data prior to the index inclusion (see Section 3.3 for a detailed discussion). The omission of this data may skew the results, potentially introducing an element of survivorship bias, where only stocks that did not expire in the first five years are represented for the majority of this period. This highlights the importance of performing regressions over different periods to examine the first, and second five-year periods in isolation, thereby gaining a deeper understanding of both the full period and the sub-periods.

5 Conclusion

Over the past decades, the amount of capital devoted to index investing worldwide has increased tremendously. This trend is expected to continue and originates from index funds' lower cost and ability to generate competitive returns, often beating those of active funds. This thesis investigates the European stock market, using the established STOXX Europe 600 index, to test for potential effects on stocks' return, risk, and pricing efficiency, as they are added to the index. We focus on the European market, which has been relatively underexplored compared to the larger American stock market in the previous literature.

We analyze data from 2013 to 2023, covering all additions to the index during this period. By gathering fund ownership data, classified as either index or active, we test for potential effects on the stocks. Our analysis highlights four different periods for comparison purposes, among them the pandemic years 2020-2022. After index inclusion, we establish that both index and active fund ownership increases, though the index is most pronounced. In relation to stock returns, using both the geometric and arithmetic mean, we show that active fund ownership has a positive relationship, while index ownership has a non-substantial connection. We find that increased index ownership is associated with higher risk of the stocks, as seen by the stock beta and volatility. As to pricing efficiency, we apply the well-established variance ratio test, with two different lags: two and five days. For both lags, we document no significant relationship to either index or active ownership. While not desirable, this is in line with previous research.

This thesis presents important results regarding the rapid growth in index investing. We conclude that increased index investing does not undermine the pricing efficiency of stocks, contrary to our initial hypothesis. Our results imply, however, that a side effect of increased index ownership might be an increased idiosyncratic risk. This is valuable information for any investor that hold stocks with any passive ownership. As the trend of increased index ownership is expected to continue, seeing if these conclusions still hold will be of particular interest. Not examined in this thesis, is the effect of other fund ownership types, such as hedge funds, which would be interesting to see if it has the same effect as index ownership. Moreover, we only observe stocks joining the index and its subsequent effect. An interesting idea would be to also study those stocks leaving the index and test if that results in an opposite effect to joining. We leave these important questions for future research to disentangle.

References

Ahad, N. A., & Yahaya, S. S. S. (2014). Sensitivity analysis of Welch's *t*-test. *AIP Conference Proceedings*, 1605(1), 888-893. https://doi.org/10.1063/1.4887707

Ahn, B.H., & Patatoukas, P.N. (2022). Identifying the Effect of Stock Indexing: Impetus or Impediment to Arbitrage and Price Discovery? *Journal of Financial and Quantitative Analysis, 57*(5), 2022-2062. https://doi.org/10.1017/S0022109021000235

Baltussen, G., van Bekkum, S., & Da, Z. (2019). Indexing and stock market serial dependence around the world. *Journal of Financial Economics, 132*(1), 26–48. https://doi.org/10.1016/j.jfineco.2018.07.016

Baruch, S., & Zhang, X. (2022). The Distortion in Prices Due to Passive Investing. *Management Science, 68*(8), 6219-6234. https://doi.org/10.1287/mnsc.2021.4114

Ben-David I., Franzoni F., & Moussawi R. (2018). Do ETFs increase volatility? *Journal of Finance, 73*(6), 2471–2535. https://doi.org/10.1111/jofi.12727

Bogle, J.C. (2016). The index mutual fund: 40 years of growth, change, and challenge. *Financial Analysts Journal, 72*(1), 9–13. https://doi.org/10.2469/faj.v72.n1.5

Bond, P., & Garcia, D. (2022). The Equilibrium Consequences of Indexing. *Review of Financial Studies, 35*(7), 3175–3230. https://doi.org/10.1093/rfs/hhab106

Breugem, M., & Buss, A. (2019). Institutional investors and information acquisition: Implications for asset prices and informational efficiency. *Review of Financial Studies, 32*(6), 2260–2301. https://doi.org/10.1093/rfs/hhy103

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

Buss, A., & Sundaresan, S. (2023). More Risk, More Information: How Passive Ownership Can Improve Informational Efficiency. *Review of Financial Studies, 36*(12), 4713-4758. http://dx.doi.org/10.2139/ssrn.3615173

Callaert, H. (1999). Nonparametric hypotheses for the two-sample location problem. *Journal of Statistics Education, 7*(2). https://doi.org/10.1080/10691898.1999.12131270

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance, 52*(1), 57–82. https://doi.org/10.1111/j.1540-6261.1997.tb03808.x

Chabakauri, G., & Rytchkov, O. (2020). Asset pricing with index investing. *Journal of Financial Economics*, *2021*(141), 195-216. https://doi.org/10.1016/j.jfineco.2020.06.023

Chang, Y. C., Hong, H., & Liskovich, I. (2015). Regression discontinuity and the price effects of stock market indexing. *Review of Financial Studies*, *28*(1), 212–246. https://doi.org/10.1093/rfs/hhu041

Chen, G., Liu, X., Liu, X., & Zhao, Z. (2024). ETF ownership and stock pricing efficiency: The role of ETF arbitrage. *Finance Research Letters, 62*(1). https://doi.org/10.1016/j.frl.2024.105108

Chen, H., Noronha, G., & Singal, V. (2004). The price response to S&P 500 index additions and deletions: Evidence of asymmetry and a new explanation. *Journal of Finance*, *59*(4), 1901–1929. https://doi.org/10.1111/j.1540-6261.2004.00683.x

Coles, J. L., Heath, D., & Ringgenberg, M. C. (2022). On index investing. *Journal of Financial Economics, 145*(3), 665-683. https://doi.org/10.1016/j.jfineco.2022.05.007

Da, Z., & Shive, S. (2018). Exchange traded funds and asset return correlations. *European Financial Management, 24*(1), 136–168. http://dx.doi.org/10.2139/ssrn.2158361

Divine, G. W., Norton, H. J., Barón, A. E., & Juárez-Colunga, E. (2018). The Wilcoxon–Mann–Whitney procedure fails as a test of medians. *The American Statistician*, *72*(3), 278-286. https://doi.org/10.1080/00031305.2017.1305291

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, *25*(2), 383-417. https://doi.org/10.2307/2325486

Fama, E. F., & French, K. R. (2010). Luck versus skill in the cross-section of mutual fund returns. *The Journal of Finance, 65*(5), 1915–1947. https://doi.org/10.1111/j.1540-6261.2010.01598.x

Glosten, L., Nallareddy, S., & Zou, Y. (2021). ETF activity and informational efficiency of underlying securities. *Management Science, 67*(1), 22–47. https://doi.org/10.1287/mnsc.2019.3427

Greene, W. H. (2020). *Econometric Analysis* (8th ed.). Pearson.

Grossman, S. J., & Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *American Economic Review*, *70*(3), 393-408. https://www.jstor.org/stable/1805228

Henningsen, A., & Hamann, J. D. (2007). systemfit: A package for estimating systems of simultaneous equations in R. *Journal of Statistical Software, 23*(4), 1-40. https://doi.org/10.18637/jss.v023.i04

Investment Company Institute. (2024). *Investment Company Fact Book – A Review of Trends and Activities in the Investment Company Industry*. https://www.icifactbook.org/pdf/2024-factbook.pdf

Israeli, D., Lee, C.M.C., & Sridharan, S.A. (2017). Is There a Dark Side to Exchange Traded Funds? An Information Perspective.. *Review of Accounting Studies, 22*(3), 1048-1083. https://ssrn.com/abstract=2625975

Jaquart, P., Motz, M., Köhler, L., & Weinhardt, C. (2023). The impact of active and passive investment on market efficiency: a simulation study. *Journal of Applied Economics, 26*(1). https://doi.org/10.1080/15140326.2023.2188634

Jiang, H., Vayanos, D., & Zheng, L. (2022). Passive Investing and the Rise of Mega-Firms. https://ssrn.com/abstract=3749534

Krause, T., Ehsani, S., & Lien, D. (2014). Exchange-Traded Funds, Liquidity, and Volatility. *Applied Financial Economics, 24*(24), 1617–1630. https://doi.org/10.1080/09603107.2014.941530

Lakonishok, J., & Shapiro, A. C (1984). Stock Returns, Beta, Variance and Size: An Empirical Analysis. *Financial Analyst Journal*, 40(4), 36-41. https://www.jstor.org/stable/4478758

Liebi, L.J. (2020). The impact of ETFs on financial markets: a literature review. *Financial Markets and Portfolio Management*, *34*, 165-178. https://doi.org/10.1007/s11408-020-00349-1

Lo, A. W., & MacKinlay, A. C. (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. *The Review of Financial Studies*, *1*(1), 41-66. https://doi.org/10.1093/rfs/1.1.41

Mann, H. B., & Whitney, D. R. (1947). On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *The Annals of Mathematical Statistics*, *18*(1), 50–60. https://doi.org/10.1214/aoms/1177730491

Merton, R. C. (1980). On Estimating the Expected Return on the Market: An Exploratory Investigation. *Journal of Financial Economics*, *8*(4), 323-361. https://doi.org/10.1016/0304-405X(80)90007-0

Morningstar. (2024). *Are ETFs Taking Over Europe Too?.* Retrieved 2024-05-18 from https://www.morningstar.co.uk/uk/news/246333/are-etfs-taking-over-europe-too.aspx

Nandal, S., & Kumar, A. (2021). The risks and returns of index funds and ETFs: A Review of Existing Literature. *Turkish Online Journal of Qualitative Inquiry (TOJQI)*, *12*(8), 4141-4148.

Patel, N., & Welch, I. (2017). Extended stock returns in response to S&P 500 index changes. *Review of Asset Pricing Studies*, *7*(2), 172–208. https://doi.org/10.1093/rapstu/rax012

Pesaran, M. H. (2015). *Time Series and Panel Data Econometrics*. Oxford University Press.

Petersen, M. A. (2009). Estimating standard errors in finance panel data sets: Comparing approaches. *The Review of Financial Studies*, 22(1), 435-480. https://doi.org/10.1093/rfs/hhn053

Staer, A., & Sottile, P. Equivalent volume and comovement. (2018). *Quarterly Review of Economics and Finance, 68*, 143–157. https://doi.org/10.1016/j.qref.2017.11.001

Stambaugh, R. F., Yu, J., & Yuan, Y. (2015). Arbitrage asymmetry and the idiosyncratic volatility puzzle. *Journal of Finance, 70*(5), 1903–1948. doi:10.1111/jofi.12286

STOXX. (2024). *STOXX index methodology guide (portfolio based indices)*. https://www.stoxx.com/document/Indices/Common/Indexguide/stoxx_index_guide.pdf

Sullivan, R. N., & Xiong, J. X. (2012). How index trading increases market vulnerability. *Financial Analysts Journal, 68*(2), 70–84. https://doi.org/10.2469/faj.v68.n2.7

Vijh, A. M., & Wang, J. (2022). Negative returns on addition to the S&P 500 index and positive returns on deletion? New evidence on the attractiveness of S&P 500 versus S&P 400 indexes. *Financial Management, 51*(4), 1127–1164. https://doi.org/10.1111/fima.12391

Wang, H., & Xu, L. (2019). Do exchange-traded fund flows increase the volatility of the underlying index? Evidence from the emerging market in China. *Accounting & Finance, 58*(5), 1525–1548. https://doi.org/10.1111/acfi.12437

Welch, B. L. (1947). The Generalization of `Student's' Problem when Several Different Population Variances are Involved. *Biometrika*, *34*(1/2), 28–35. https://doi.org/10.2307/2332510

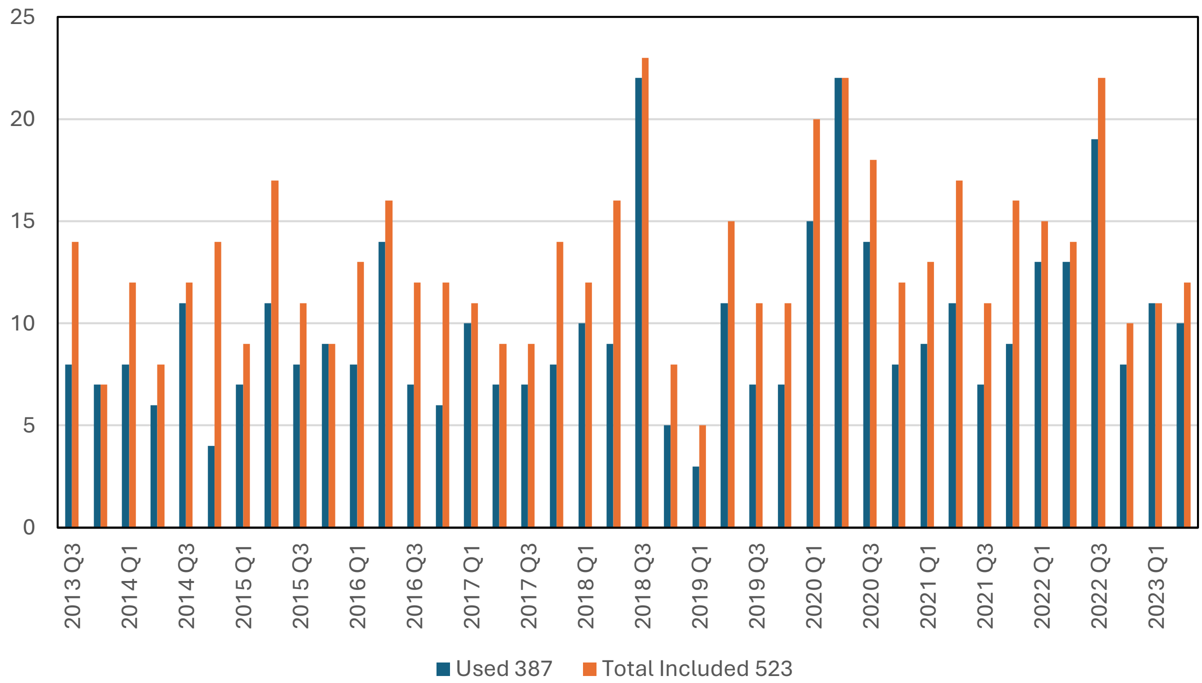
Wooldridge, J. M. (2001). *Econometric Analysis of Cross Section and Panel Data* (1st ed.). MIT Press.

Appendix

Appendix A

Figure A.1

All Included Stocks and Those Selected



Note. The figure shows all stocks that got included to the index during the sample period, in orange bars. The blue bars represent those stocks that were included in the sample. The difference between the bars is due to those stocks that were not suitable to use.

Appendix B

Table B.1

Summary Statistics on Independent Variables

Panel A: Index Fund Ownership (Decimal Form)					
Quarter	Q-2	Q-1	Q0	Q+1	Q+2
Count	387	387	387	387	387
Mean	0.0443	0.0468	0.0531	0.0559	0.0580
Std.	0.0345	0.0349	0.0344	0.0347	0.0348
Min	0.0000	0.0000	0.0000	0.0000	0.0000
Max	0.2137	0.2167	0.2199	0.2232	0.2267

Panel B: Active Fund Ownership (Decimal Form)					
Quarter	Q-2	Q-1	Q0	Q+1	Q+2
Count	387	387	387	387	387
Mean	0.2335	0.2408	0.2508	0.2573	0.2591
Std.	0.1455	0.1413	0.1387	0.1379	0.1378
Min	0.0000	0.0000	0.0000	0.0000	0.0000
Max	0.9131	0.8578	0.8109	0.7910	0.7662

Note. Panel A and B measure index and active funds ownership, respectively, in decimal form of the total shares outstanding of each stock. Q refers to "quarter", and the numbers of the specific quarter in relation to the inclusion quarter, *i.e.*, Q0. Count measures the number of stocks that were added to the index during the sample period, adjusted for those stocks that were removed. Mean, standard deviation, min, max, and median represent the summed values for all stock during the quarters.

Table B.2

Summary Statistics on Dependent Variables

Panel A: Average Daily Stock Return – Arithmetic Mean (Decimal Form)					
Quarter	Q-2	Q-1	Q0	Q+1	Q+2
Count	387	387	387	387	387
Mean	0.0021	0.0014	0.0003	0.0002	0.0017
Std.	0.0034	0.0035	0.0027	0.0029	0.0032
Min	-0.0133	-0.0234	-0.0100	-0.0105	-0.0134
Max	0.0172	0.0248	0.0117	0.0106	0.0151

Panel B: Average Daily Stock Return – Geometric Mean (Decimal Form)					
Quarter	Q-2	Q-1	Q0	Q+1	Q+2
Count	387	387	387	387	387
Mean	0.0011	0.0017	0.0014	0.0000	-0.0001
Std.	0.0036	0.0033	0.0032	0.0028	0.0029
Min	-0.0278	-0.0165	-0.0142	-0.0123	-0.0113
Max	0.0208	0.0141	0.0116	0.0110	0.0097

Panel C: Daily Stock Volatility (Decimal Form)					
Quarter	Q-2	Q-1	Q0	Q+1	Q+2
Count	387	387	387	387	387
Mean	0.0229	0.0221	0.0216	0.0214	0.0225
Std.	0.0129	0.0124	0.0104	0.0105	0.0123
Min	0.0014	0.0014	0.0051	0.0064	0.0055
Max	0.1094	0.1021	0.0753	0.0966	0.1216

Panel D: Quarterly Stock Beta					
Quarter	Q-2	Q-1	Q0	Q+1	Q+2
Count	387	387	387	387	387
Mean	0.8391	0.8608	0.8803	0.9536	0.9661
Std.	0.5172	0.5304	0.5486	0.5326	0.5293
Min	-1.4506	-1.3807	-0.8684	-0.3557	-1.0174
Max	3.4944	2.8003	3.1686	4.1670	2.7812

Panel E: Variance Ratio Lag 2 (VR2)					
Quarter	Q-2	Q-1	Q0	Q+1	Q+2
Count	387	387	387	387	387
Mean	1.0042	0.9996	0.9980	0.9870	0.9735
Std.	0.1487	0.1495	0.1461	0.1401	0.1391
Min	0.5469	0.6125	0.4348	0.5556	0.5041
Max	1.3804	1.4912	1.5001	1.3918	1.3376

Panel F: Variance Ratio Lag 5 (VR5)					
Quarter	Q-2	Q-1	Q0	Q+1	Q+2
Count	387	387	387	387	387
Mean	0.9645	0.9591	0.9650	0.9271	0.9022
Std.	0.2960	0.2924	0.3113	0.2854	0.2828
Min	0.1975	0.3055	0.2727	0.3365	0.3312
Max	1.8585	2.0360	2.3174	1.9327	1.9777

Note. Panel A, B, C, D, E, and F refer to all the used dependent variables in the panel regression. Q refers to "quarter", and the numbers of the specific quarter in relation to the inclusion quarter, *i.e.*, Q0. Count measures the number of stocks that were added to the index during the sample period, adjusted for those stocks that were removed. Mean, standard deviation, min, max, and median represent the summed values for all stock during the quarters.