ETFs and Volatility

An empirical study in the Swedish stock market

Lukas B. Bantle

Nikolai Doepel

Master Thesis

Stockholm School of Economics

2019

ETFs and Volatility – An empirical study in the Swedish stock market

Abstract:

ETFs have experienced tremendous growth and with it have gained increased relevance in the market over the last decade. Because of their special characteristics and close connection to their underlying securities, ETFs may propagate liquidity shocks through arbitrage channels into their underlying securities and thus increase the volatility of these securities. We estimate regression models to examine a connection between ETFs and the volatility of the underlying securities in the Swedish stock market. The results show, that increased ETF ownership in a stock is related to an increase in the volatility of the stock. Examining this relation in more detail on the OMXS30 index level, we find, that the ETF creation and redemption flows are unrelated to an increase in volatility. On the other hand, ETF trading volume is related positively to volatility and the relationship is magnified in periods of high ETF mispricing. This increased volatility constitutes both noise and fundamental volatility, which indicates, that ETFs may increase noise in the market, but also act as a price discovery channel.

Keywords:

ETF, volatility, arbitrage, noise, price discovery

Authors:

Lukas Bantle (41265)

Nikolai Doepel (41269)

Tutor:

Paolo Sodini, Professor, Department of Finance

Acknowledgement:

To Professor Paolo Sodini, we express our deepest gratitude for his continuous and unwavering support throughout the research and writing process.

Master Thesis

Master in Finance

Stockholm School of Economics

Table of Contents

List of Abbreviations

AP	Authorized Participant
AUM	Assets Under Management
BN	Beveridge Nelson
BtM	Book-to-Market
C/R	Creation and Redemption
Eikon	Thomson Reuters Eikon
IS	Information Share
IVV	iShares Core S&P 500 ETF
ETF	Exchange Traded Fund
MA	Moving Average
MOC	Market-on-Close
NAV	Net Asset Value
SPX	S&P 500 Cash Index
SPY	SPDR S&P500 ETF
U.S.	United States
VAR	Vector Autoregressive
VECM	Vector Error Correction Model
VMA	Vector Moving Average
VR	Variance Ratio
VOO	Vanguard S&P 500 ETF
VWAP	Volume Weighted Average Prices
Xact	XACT OMXS30 ETF
XLF	Financial Select Sector SPDR Fund

List of Tables

List of Figures

List of Appendices

1. Introduction

1.1. Motivation

Exchange Traded Funds (ETFs) have experienced tremendous growth ever since the introduction of the first and today's largest ETF, the SPDR S&P500 ETF (SPY), in 1993. ETFs have now surpassed $5Trn in global AUM and have thus overtaken hedge funds in total AUM (Blackrock, 2018). Even more impressive numbers can be seen in the trading volume: Holding around 10% of the market capitalization of all securities in the U.S., ETFs represented over 30% of the daily trading volume by the end of 2016 (Ben-David et al, 2018). In Sweden, the average ETF ownership in the largest companies has approximately tripled in the last 7 years (see *Figure 2*). As these numbers are expected to grow continuously at a considerable rate (EY, 2017), ETFs represent an area that has started to attract a lot of research in recent years, and calls for further important research, that is relevant for not only market makers, but market participants and regulators as well.

ETFs are generally seen as a useful financial innovation. Through their construction, they can offer easy exposure to a basket of assets at low cost regarding management fees and trading cost. Furthermore, they allow investors to access a basket of securities to which they otherwise may have no, or only costly, access to. This may include corporate bonds or stocks in emerging markets. Moreover, they offer tax advantages over other financial products, as the realized capital gains throughout the holding period tend to be lower than for products with a similar investment purpose like mutual funds due to the reinvestment of dividends. While these characteristics seem to make ETFs particularly interesting for long-term investors (Bhattacharya and O'Hara, 2018), recent research has shown that the average holding period for ETFs is shorter than for other funds (Ben-David et al, 2018) and for their underlying stocks (Broman and Shum, 2018). Unlike other funds, ETFs can be traded throughout the day, just like a stock. As such, these instruments are often used by market participants to hedge their positions, make directional bets, or to profit from arbitrage opportunities. Hence, these highly traded and liquid products, which are often only supposed to passively track an index (Broman and Shum, 2018), are suspected to also impact their underlying stocks in regard to liquidity, volatility, co-movement, prices, and general price-efficiency.

Prior research has mainly focused on the U.S. ETF market, with little regard for the European market. To the best of our knowledge, there has been little research conducted surrounding ETFs in the European market (except for Winne et al, 2014; Xu and Yin, 2017b), and none in the Swedish market. While the U.S. market for ETFs is evidently larger than the European market, the European ETF market also enjoys considerable growth (EY, 2017). A striking difference between the two markets lies within their market concentration. The European market is more fragmented, and there are usually several market participants that set up ETFs for the same index. Whereas there are three ETFs tracking the S&P 500 in the U.S., there are 35 different ETFs tracking the Euro STOXX 50 (Hill et al., 2015). Furthermore, in the U.S. the ratio of OTC to secondary market ETF trading is assumed to be 30:70, while in Europe this ratio is assumed to be reversed. However, there is limited certainty about these numbers, as until the beginning of 2018 OTC and private bilateral ETF trades did not have to be reported in Europe. With the introduction of MiFID II on January 3rd of 2018 however, these trades are now to be disclosed, which will help unveil the true liquidity and the depth of the European ETF market. This may thus affect the interest in ETFs, and thus, their effect on the underlying market (Skypala, 2018). For these reasons, the way ETFs affect the underlying securities in the European market may be different to the influence they have in the American market.

As such, with our study we want to test for potential impacts of ETFs on the volatility of the underlying securities in the Swedish stock market. We conduct our analysis on both a stock level and an index level. The analysis of the latter enables us to distinguish between the source of the ETF-volatility relation, i.e. primary or secondary market trading, and the nature of the volatility, i.e. fundamental volatility or noise.

We start by regressing the volatility of some of the largest Swedish stocks on the ETF ownership, other control variables, and fixed effects from the period 2012-2018. Our results indicate a positive link between ETF ownership and stock returns volatility which is in line with findings of previous literature. We find, that a one unit increase of standardized ETF ownership is associated with an increase in volatility to the degree of 8.1% of a standard deviation of volatility. These findings motivate us to move our analysis to the index level, where we distinguish between an ETF primary market trading effect, which is measured by the creation and redemption flows, and an ETF secondary market trading effect, which is measured by the daily trading volume of the Xact OMXS30 ETF (Xact). We find, that the ETF’s secondary market trading volume is related to the volatility in the market whereas we are unable to establish such a link between the ETF’s creation and redemptions flows (C/R flows) and volatility. In an extension of this analysis, we identify a moderating relation of the mispricing of an ETF (the difference between its net asset value and price) on the effect that trading volume has on volatility. That is, our regression shows, that the higher the mispricing of the ETF at the close of a trading day is, the higher is the estimated effect of the next day's trading volume on the next day's volatility.

To then identify the nature of the volatility, we make use of the Variance Ratio (VR) after Lo and MacKinlay (1988) to estimate the noise that is present in the market. We then regress the VR on the trading volume and C/R flows. Our results indicate, that the trading volume of the Xact is related to increased noise in the market while no relationship seems to exist between C/R flows and noise. In a next step, we calculate the fundamental volatility in the OMXS30 through the Beveridge Nelson decomposition (Beveridge and Nelson, 1981). Regressing on the resulting fundamental volatility we can establish a link with the ETF trading volume but not with the flows. In a final step, we quantify the price discovery happening in the Xact compared to the underlying cash index by calculating the Information Share after Hasbrouck (1995). While we find a negligible Information Share for the ETF in the whole sample period between 2012-2019, analysing more recent data for the year 2018 assigns an Information Share of about 11% to the Xact. All these findings, that indicate an influence of trading volume on both the noise and fundamental volatility in their underlying securities, are generally in line with results of previous literature. However, not being able to find a significant link between the ETF C/R flows and any kind of volatility in the market generally contradicts the findings of previous research. This may be, among other reasons, due to the relative infancy of the Swedish ETF market.

Hence, our empirical study contributes to the current literature in the following ways: First, we contribute to the general debate surrounding the possible (negative) impacts that ETFs have on their underlying securities and the resulting implications for market participants and regulators. Secondly, to the best of our knowledge, it is one of the few studies that analyses the effect of ETFs in a European market setting on volatility, and the first one to do so in the Swedish market. Thus, the study analyses a less advanced market than that in the U.S., which also shows structural differences such as in the concentration of the market. Thirdly, we are the first study to differentiate between the ETF impact that stems from the Authorized Participants' activity in the primary market, and the general trading activity of all market participants in the secondary ETF market. Hence, we lay the grounds for a more precise identification of the ETF mechanisms that may have negative implications for market stability. Finally, our study extensively analyses the relation between ETF ownership and volatility at multiple levels. It examines both the individual stock level and the index level as well as the total volatility, noise and fundamental volatility in the market.

The rest of our thesis is structured in the following way: The remainder of section 1 provides a brief introduction into the mechanisms of ETFs. Section 2 discusses the related literature. In section 3, we explain the ideas and concepts behind our hypotheses. Section 4 explains the data collection and data cleaning process while highlighting some key characteristics of important data. In Section 5, we describe the methodology for the stock level analysis and interpret its results. Section 6 constitutes the main part of our analysis, where we describe the methodologies used and the results of our index level analysis of total volatility, noise, fundamental volatility, and information share. This is followed by section 7 where we discuss the findings and limitations of our study. Section 8 concludes this thesis.

1.2. How ETFs work

An Exchange Traded Fund is a pooled investment product that presents a basket of other securities. It combines features of open-ended and closed-ended mutual funds in that new shares can be created or redeemed and that the shares can be traded throughout the day on public exchanges. ETFs are issued and set up by a sponsor, who determines the purpose of the fund and decides on the securities to be contained in the fund. The issuer then allows selected market makers, called Authorized Participants (APs), to create or redeem shares of the ETF in exchange for the proper basket of underlying securities (or sometimes cash).

ETFs often track a certain index (e.g. SPY tracks the S&P 500) or the performance of a certain industry (e.g. XLF tracks the financial industry). ETFs can either physically own the underlying securities or present a promise to deliver the underlying assets using derivates. The latter enables the issuing of special ETFs called leveraged/inverse ETFs. These ETFs try to replicate the performance of an index 2x/3x or negatively -1x/-2x/-3x by using other financial products such as swaps.

The APs' role for ETFs is twofold. First, they act as liquidity providers as they have access to the ETF's underlying assets. APs react to market demands for ETFs and create or redeem ETF shares accordingly. Market participants can also place an order with the AP to create or redeem shares through it. Moreover, APs help to minimize the tracking error of the ETF by arbitraging away differences in the ETF's net asset value (NAV) and its market price. When the market price of an ETF is higher (lower) than its NAV, the AP will create (redeem) ETF shares with the sponsor in exchange for the underlying securities. Through the positive price pressure as the AP buys (sells) the underlying securities and sells (buys) the ETF on the market, the NAV and the price of the ETF will converge in a normal market setting. The creation or redemption process happens at the end of the trading day and generally in unit sizes of 50,000 or 100,000. Furthermore, any market participant can take advantage of pricing differences between the NAV and ETF market price in the secondary market as ETFs are tradeable throughout the day. To reduce the tracking error, sponsors usually facilitate the arbitrage process by publishing the NAV of an ETF every 15 seconds. Through these arbitrage channels and the law of one price, an ETF's NAV and its price should always converge.

For further details about the mechanics of ETFs, the article by Antoniewicz and Jane (2014) offers a more in-depth discussion about the workings and regulatory frameworks for ETFs. Hill et al. (2015) provide an extensive description of the ETF landscape ranging from how to evaluate ETFs to the different types of ETFs present in the market. Additional literature is covered in the next section.

2. Literature Review

This thesis can be categorized as fitting into three main streams of literature: The effect of non-fundamental demand shocks from institutions or funds onto securities, the effect of indexing on the market, and the influence of ETFs on their underlying securities and market quality in general. Particularly the latter strand is the most relevant for our research. Hence, this literature review will focus on the research on ETFs and our findings will be contextualized throughout the thesis.

The literature on the effect of institutional investors and funds onto stocks reveals a correlation between them. Basak and Pavlova (2013) show, that institutional investors amplify stock market volatility by tilting their portfolio towards the benchmark and more risky stocks. Greenwood and Thesmar (2011) argue, that the concentrated ownership of funds and resulting correlated liquidity shocks amplify stock fragility and thus price volatility. Furthermore, Lou (2012) finds a relation between flow shocks into mutual funds and temporary price shocks into individual stocks.

Furthermore, our research fits into the greater theme of the rise of (passive) index investing and its consequences on the market (see Wurgler, 2010; Sullivan and Xiong, 2012; Appel et al, 2016; Baltussen et al. 2016; Cong and Xu, 2016; Bond and García, 2018). Wurgler (2010) and Sulivan and Xiong (2012) describe the negative consequences of index investing for the market because of increased market fragility and higher correlations among stocks. Cong and Xu (2016), while agreeing on these effects, explain part of their origin lies in higher price efficiency regarding systematic risks in single stocks.

A similar debate about price efficiency can be found in the ETF literature. Several recent studies indicate, that ETFs now lead the price discovery, after futures have long been recognized as the price discovery leader (Wermers and Xue, 2015; Xu and Yin, 2017a; Bhattacharya and O'Hara, 2018; Buckle et al, 2018; Xu et al, 2018). Xu and Yin (2017a) argue, that higher flows into ETFs increase the relative information share of these and that the price discovery regularly occurs in the largest ETF that tracks an index. When looking at flows from APs, Xu et al. (2018) claim, that the reasons for APs to create or redeem shares are threefold: While they respond to market demands for the ETFs and arbitrage differences between the NAV and the ETF price away, APs also use ETFs to make their own bets on the movement of the market as related to upcoming news events (informed trades). Similarly, Pan and Zeng (2019) find, that APs may use their unique position, as both an AP and a market maker, to unwind their bond inventory imbalances in the bond ETF market by creating and redeeming ETF shares.

Furthermore, they find that a high liquidity mismatch propagates price discrepancy between the ETF and the NAV of its underlying bonds. This effect is exacerbated in times of high volatility, low bond liquidity and higher trading costs, thus adding fragility into the market. Petäjistö (2017) confirms these findings for illiquid asset markets in general. Bhattacharya and O'Hara (2018) argue, that in illiquid markets, rather than the stocks determining the price of the ETF, the "tail is wagging the dog" and demand shocks into the ETF determine the price of the underlying assets through arbitrage rather than vice versa. As such, capital may not be efficiently allocated anymore.

Da and Shive (2018) find higher co-movement among stocks that are included in ETFs compared to stocks that are not. This effect is exacerbated for ETFs with higher turnover and tends to be stronger for smaller stocks and stocks with lower turnover. Furthermore, they see an increased Beta and a decreased lagged Beta correlated to higher ETF turnover, which they argue indicates a decrease in price efficiency. Broman (2016) also finds excess co-movement among stocks caused by demand shocks into ETFs. By finding strong price reversals after these shocks, he argues the found co-movement presents non-fundamental demand and is thus harmful for pricing efficiency. The model developed by Leippold et al. (2015) predicts, that even non-indexed stocks are affected by the demand shocks into ETFs and thus display excess co-movement.

Malamud's (2015) dynamic model for ETF flows predicts more mixed results. While he agrees, that ETFs in general increase correlation among stocks, he argues that the introduction of a new ETF may decrease the co-movement. Staer (2017) analyses price reversals after return shocks caused by ETF flows in more detail. His study finds that 38% of the return shocks reverse after 5 days, while the remaining 62% of the movement can be interpreted as permanent. Hence, he argues, that ETFs increase pricing efficiency to a certain extent, as ETFs act as a price discovery channel. Analogously, Glosten et al. (2016) decompose stock earnings into a systematic and non-systematic component. They conclude, that ETFs enhance the incorporation of earnings announcements into the prices. This improved price efficiency is particularly observable for small and hard-to-access stocks. Furthermore, they find that their earnings decomposition can explain parts of the additional co-movement found by Da and Shive in 2018 (see also Cong and Xu, 2016).

Conversely, Israeli et al. (2017) argue that adverse selection makes noise traders invest into ETFs instead of individual stocks and thus decreases the advantages for analysts to research a company. As such, less information is processed, and pricing efficiency decreases. They confirm Glosten et al.'s (2016) finding for increased contemporaneous price efficiency, however, they show that the lagged earnings announcement's incorporation decreases, and stock synchronicity increases. Additionally, Hamm (2014) argues, that adverse selection also decreases liquidity of individual stocks.

When looking at the effect of ETFs on market volatility, the Flash Crash in May 2010 sparked interest in the negative influence of ETFs and particularly that of leveraged ETFs (Madhavan, 2012). At the end of each trading day, these funds need to actively re-leverage their portfolio to obtain the adequate exposure for the next day. As there is greater rebalancing need when the market moved more, leveraged ETFs may exacerbate already existing volatility in the market. Trainor (2010) could not identify any increased volatility due to leveraged ETFs by comparing volatility between opening hours and closing hours across different years. However, the paper fails to incorporate any flow or asset data from leveraged ETFs and solely investigates the S&P 500 where leveraged ETFs used to make up a fairly small amount of the whole market capitalization back in 2010.

A more sophisticated study conducted by Shum et al. (2016) concludes, that leveraged ETF rebalancing does indeed have a statistically significant effect on market volatility at market-on-close (MOC). This effect is especially visible in days with high market volatility and opens up the potential for predatory trading. Cheng and Madhavan (2009) estimate, that if the market moves by 1% during a trading day, leveraged ETFs could account for up to 16.8% of the trading volume at MOC. In general, Humphries (2010) argues that leveraged ETFs carry the problem of being margin products without the protection of margin rules. They may thus be dangerous for investors and may ultimately decrease market stability.

Looking at plain equity ETFs, several studies could identify a correlation between ETFs and stock volatility. Stratmann and Welborn (2012) find, that APs' creation and redemption process to avoid fails to deliver granger causes higher MOC volatility in the underlying stocks. Xu and Yin (2017b) identify a contemporaneous and lagged granger causality between trading volume of ETFs and volatility in the underlying stocks. Krause et al. (2012) utilize a volatility spillover framework to uncover bi-directional volatility spillovers between ETFs and their largest component stocks. They show, that the spillover from ETFs to their component stock is larger than the reverse relationship in 42 out of 50 cases. Furthermore, they can identify a general upward trend in the spillover effect and a negative relationship of illiquidity on the spillover effect. However, these papers fail to address whether this volume-volatility relation may be due to improved price discovery or merely presents additional noise.

Wermers and Xue (2015) attempt to distinguish between the price discovery function of an ETF and added noise by analysing the lead-lag relationship between the ETF price and the underlying cash index. They focus their analysis on the S&P 500 and thus the SPY and the SPX. Whenever the SPY leads the SPX, they argue that the ETF acts as a price discovery channel. However, when the SPX leads the SPY, the additional regression value of SPY volume on SPX volatility reflects the added noise. With this setup, they find that noise trades have a significant impact on the SPX volatility. However, the influence of noise trades is considerably smaller than the influence of 'informed trades' and loses its predictive power after 300 seconds

Xu et al. (2016) categorize ETF flows by the motivation of the trade. They demonstrate, that trades based on private information and on belief dispersion are correlated to the variance in efficient price innovations (fundamental volatility), while liquidity trades have little correlation with the fundamental volatility. Again, this suggests that ETFs act as a price discovery channel. However, the correlation of liquidity-based trading and belief-dispersion-based trading with the total market volatility suggests, that ETFs negatively impact the market quality by increasing non-fundamental volatility. Furthermore, the authors illustrate that it is mainly an index' largest ETF that dominates the influence on both types of volatility.

Wang and Xu (2019) focus on the influence of creation/redemption ETF flows onto the market's total and fundamental volatility. They categorize flows as either backward looking (APs response to market demand) or forward looking (APs act on private information on market movement). Furthermore, by separating fundamental volatility from total volatility using Beveridge Nelson decomposition and Information Share after Hasbrouck, they can analyse the predictive power of the different flows on the volatilities. This way, they find that forward looking flows have significant predictive power on both total volatility and fundamental volatility. Moreover, higher arbitrage opportunities in ETFs increase the effect on total volatility but does not have an impact on fundamental volatility. On the other hand, backward looking flows show no significant impact on market volatility. This study indicates, that the supposedly passive funds are not so passive after all, as the AP's information-based activity significantly impacts the fund and underlying market price.

Ben-David et al. (2018) investigate ETFs' effect on volatility from a different perspective. After proving that ETFs attract short term and high frequency investors, they find a correlation between a stock's ETF ownership and volatility. By demonstrating that liquidity shocks impounded into stocks through ETFs have a mean reverting effect they prove that the volatility increase is non-fundamental. Furthermore, the authors argue, that the systematic, non-diversifiable risk for stocks increases with higher ETF ownership. It is noteworthy, that all aforementioned research finds a stronger impact from ETFs on the underlying stocks in recent years, indicating an ever-growing influence of ETFs.

On the other hand, while Agapova and Volkov (2018) find higher volatility for bonds included in bond ETFs relative to those that are not, they cannot identify mean reversion for bond prices following liquidity shocks into ETFs. This could mean, that ETFs significantly improve the price discovery process in the bond market. However, it could also be due to a smaller size of the bond ETF market or be a result of the illiquidity of the underlying (Krause et al, 2015; Ben-David et al, 2018).

Based on this previous research, we explain the rationale behind our hypotheses in the following section.

3. Hypothesis Development

Our main testable hypothesis is whether ETFs have an influence on the volatilities of the underlying securities in Sweden. In more detail, we want to decompose this effect into a primary market trading effect, that is caused by the liquidity buffer provided by APs through the special mechanisms of an ETF, and a secondary market trading effect, that is caused by the frequent arbitrage opportunities ETFs provide, as well as the quick incorporation of new information into the prices that is facilitated by ETFs. Furthermore, we will want to distinguish between fundamental volatility – volatility that is inherent in the market and consistent with the efficient market hypothesis – and market noise – volatility that is additional to the fundamental volatility and reduces market efficiency.

There are two ways in which ETFs could influence the volatility of its underlying securities, which are both based on the fact, that an ETF's price and the NAV of its underlying securities must necessarily converge. Firstly, volatility in the market may be increased by the C/R process from APs. As ETF shares are created/redeemed in large baskets of 50,000/100,000 shares in exchange for the underlying securities as defined in the creation baskets, this leads to a large demand for/supply of shares in the underlying stocks by the APs, which should consequently influence the price of the stocks and thus the volatility. To exemplify this, we can assume that a market participant (rightly) wants to bet on the Swedish market going up and thus wants to buy ETFs that track the OMXS30 (*Figure 1.b.a*). As this increases the price of the ETF while the NAV of the underlying securities stays the same, the APs see the opportunity to profit from selling (short) the ETF (*Figure 1.b.b*). To fulfill their obligation to deliver the ETF shares to the buyer, APs want to create additional ETF shares, as this is cheaper for them than buying the ETF, given that the NAV of the underlying securities is below the price of the ETF. As (physical) ETF shares are created in exchange for the securities defined in the creation basket, the AP must buy the required securities in the market which therefore increases the price for the underlying securities (*Figure 1.b.c.*). In a case where the original demand for the ETF presented noise, the liquidity provided by the AP helped implement wrong views on the market into the price and thus increased the noise in the market. Prices will thus revert to the fundamental value after some time (*Figure 1.a.c.*). These processes work analogously with a market participant wanting to sell the ETF and the AP buying it to redeem it for the underlying securities which they will have shorted after buying the ETF. Hence, this additional liquidity buffer provided by APs, who will always be able to exchange ETF shares for underlying securities (and vice versa) and thus profit from a mispricing, may add additional volatility to the market. Hence, a higher C/R flow of ETF shares may be related to a higher volatility in the market. Such a relation is indicated by Wang and Xu (2019) and we will test for such an effect under the name *primary market trading hypothesis*.

Secondly, as a discrepancy between the price and the NAV of an ETF creates arbitrage opportunities for all market participants, liquidity shocks into the ETF could propagate through arbitrage channels onto their underlying stocks and thus cause a price movement in them. To exemplify this, we can assume that unusually high amounts of buy orders for an ETF would lead to a sudden price increase in the ETF (*Figure 1.a.a/1.b.b*). As described above, arbitrageurs will try to profit from the difference in NAV and market price and will short the ETF while going long in the underlying assets. As such, the prices of the underlying assets will increase until the arbitrage opportunities disappear (*Figure 1.a.b/1.b.c*). From this state, there are two possible future states depending on the origin of the liquidity shock. If the liquidity shock reflected fundamental demand (*Figure 1.b.a*), both the ETF and the underlying assets are efficiently priced and will stay at that level (*Figure 1.b.c*); the ETF therefore functioned as a price discovery channel. However, if the liquidity shock presented noise, the prices will revert to their prior level (*Figure 1.a.c*); the ETF channel consequently added noise to the underlying stocks. Given this framework, such liquidity shocks should be more common and have a more significant impact on a stock's price when its ETF ownership is larger. Furthermore, as this effect occurs when the ETF is traded, the effect should also be related to the trading volume of the ETF. This idea is the ground for our second hypothesis and will be called *secondary market trading hypothesis* in the following.

However, as described above, an increased volatility may stem from the fact that the ETFs can act as an additional price discovery channel that is used by market participants to easily incorporate their information about the macroeconomic circumstances into the market and make trades. This is because ETFs allow an investor to buy a whole basket of securities that have a common systematic Beta in one highly liquid financial product. Hence, we want to analyse for both the primary market trading hypothesis and secondary market trading hypothesis whether an increase in volatility in the underlying stocks is related to an increase in the fundamental volatility, the noise, or both. A relation to fundamental volatility indicates, that the ETF acts as a price discovery tool. Consequently, we have four different hypotheses which we want to analyse[1]:

I.a	*Noise primary market trading hypothesis*
I.b	*Price discovery primary market trading hypothesis*
II.a	*Noise secondary market trading hypothesis*
II.b	*Price discovery secondary market trading hypothesis*

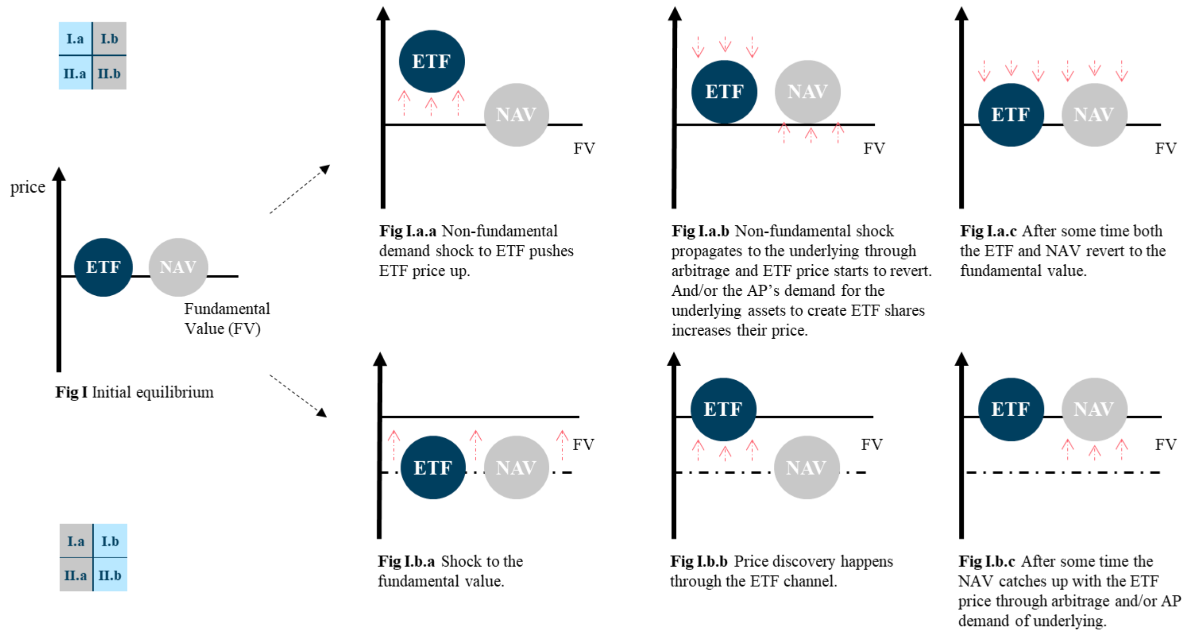

Figure 1. Liquidity shock propagation through ETFs

[1] Note, that these hypotheses are partly based on the idea, that ETFs add an additional demand layer into the market. This is proven by Ben-David et al. (2018).

4. Data

The following will explain the data collection and cleaning process for both the stock level and index level data. Furthermore, we will briefly discuss key characteristics of the collected data.

As we combine analyses on the stock and index level in order to answer our research question, we leverage two separate datasets for our analysis. Firstly, in order to lay the foundation for our subsequent index level analysis, we use a stock level dataset in order to establish the link between the volatility of stock market returns and the emergence of ETFs. Our second dataset contains the data for the analysis on the index level. Further, both our datasets cover overlapping periods (01/2012-12/2018 for the stock level, and 03/2012- 03/2019 for the index level, respectively).

4.1. Stock level data

Our stock level dataset initially contained historical data of the 50 largest stocks on the Swedish stock exchange measured by market capitalization, as of February 2019. In order to perform our analysis and obtain the data necessary, we primarily utilize the Thomson Reuters Eikon (Eikon) and Yahoo Finance databases to collect both trading data as well as shareholder data for the stocks. Our trading data consists of daily closing and volume weighted average prices (VWAP), share-denominated daily trading volume, market capitalization and Price-to-Book metrics, whereas our shareholding data consists of end-of-quarter holdings in the stocks by different funds.

The shareholding ownership obtained from Eikon contains the number of shares held by different funds at the end of each quarter. In order to separate between ETF funds and non-ETF funds, we utilize Mitre Media's ETFdb.com's list of ETFs with exposure to the Swedish market and cross-reference it against the Eikon list of funds holding stocks. For completeness, we complement the list of ETFs by including all additional funds with shareholdings in the individual stocks that contain "ETF" in their names. Due to data source access issues, our data collection methods differ from previous literature. For example, Ben-David et al. (2018), in their analysis of the US stock market, utilize Center for Research in Security Prices (CRSP) data to identify ETFs based on system specific classifications, and complement the list by using Compustat and OptionMetrics databases. Following the consolidation of the ETF list, our sample contains 688 ETFs.

4.1.1. Data issues

When collecting data for the 50 largest stocks, several issues arise that limit our dataset. Firstly, some of the companies are recent listings. For example, Essity was spun out of SCA in 2017, and the short time frame of the available data limits their usefulness in our analysis. Secondly, other companies had various missing data in the Thomson Reuters Eikon database, preventing the inclusion in our sample. Finally, due to listing on multiple exchanges, some other companies contained spurious data, including Astra Zeneca, rendering them invalid for inclusion. These limitations resulted in a dataset of 38 of the 50 largest stocks on the Stockholm Stock Exchange being used in our study.

4.1.2. Sample characteristics

For our sample of 38 stocks, the average ETF ownership, measured as the average of the shares owned by ETFs relative to the total shares outstanding for the individual companies, has increased steadily over time, as shown in *Table 2*. The trend seen in *Table 2* is in line with the global trend of growing ETFs: the funds have grown from having less than USD 700 billion in AUM preceding the Great Financial Crisis to USD 5 trillion globally in 2018 (Evans and Wilson, 2018). The data displays accelerated growth from the end of 2016 onwards. This is in line with global trends, and the growth is attributed mainly to increased demand from global institutional investors (Ståhl, 2018).

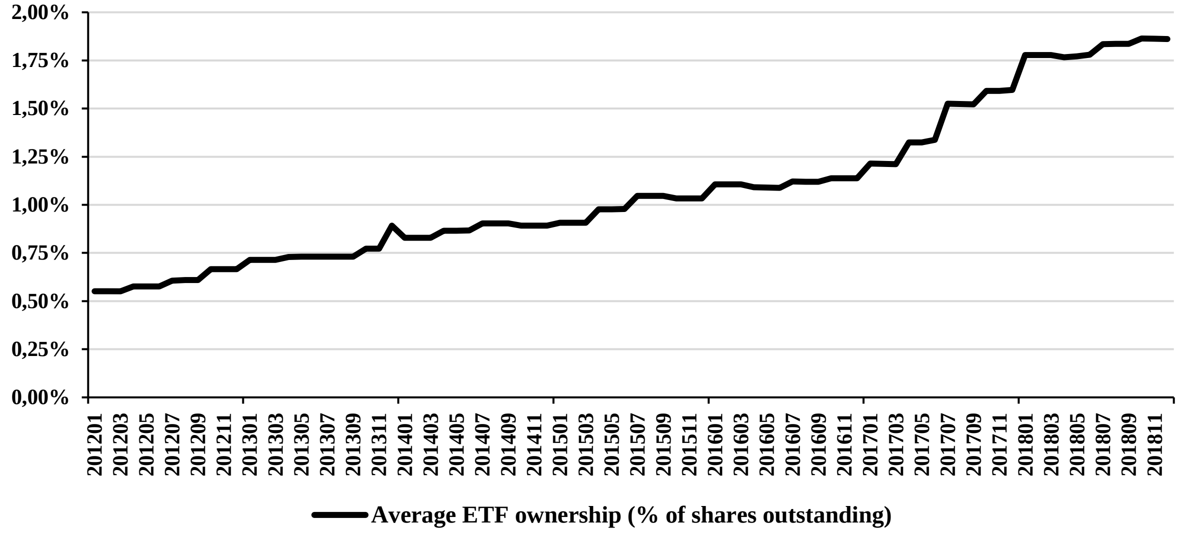

Figure 2. ETF ownership over time
The figure illustrates the average proportion of shares held as a percentage of the total shares outstanding for the 38 stocks used in the sample

Comparing the development in ownership to previous literature reveals a difference between the Swedish and the American markets. In our sample, we find that the proportion of shares held by ETFs has increased from 0.55% in January 2012 to 1.86% in September 2018. Thus, the relative infancy of the Swedish market is notable: 10 years earlier, in 2002, the ETF ownership of the stocks in the S&P 500 already amounted to the same portion as in Sweden in 2012 (Ben-David et al, 2018). By 2012 the ETF ownership had grown to 5.58% in the S&P 500.

Further, the development of the individual stocks in our sample has varied significantly. As an example, the ETF ownership in the stock that has experienced the least growth, Latour, in fact decreased from 0.22% to 0.02%. This differs significantly to the change in the ETF ownership of Swedish Match, which more than quadrupled, from 1.07% to 4.51% over the period. The number of funds invested in the stocks largely explains this. Over the period, the average number of ETFs with positions in Swedish Match increased from 36 funds in 2012 to 154 funds in 2018, whereas the number of ETFs invested in Latour increased from 1 to 20 funds between 2012 and 2018, but the largest fund invested in 2012, Invesco Global Listed Private Equity ETF, liquidated its position, and the funds with positions in the stock at the end of 2018 had an average position that is less than 1% of the size of Invesco's .

Overall, as shown in *Table 9,* the median number of ETFs invested in the stocks increased by 247%, from 34 funds per stock in 2012 to 118 ETFs per stock in 2018. Compared to the increase in the ETF ownership as a proportion of the shares outstanding, this reveals that while the number of ETFs invested in the stocks has increased substantially, their average holdings have stayed relatively constant.

4.2. Index level data

For our index level data, we focus on the OMXS30 index, given the existence of an index-tracking ETF. While initially wanting to expand our analysis and see whether there exist differences in the effects of large and small ETFs tracking the same index (Xu and Yin, 2017a), unfortunately this is not possible due to the retirement of the other ETF that has historically tracked the OMXS30 index, namely the SpotR OMXS30 UCITS ETF. The SpotR operated over the period 2011-2017. For this reason, our analysis, with the aim of exploring the impact on the Swedish market specifically, is limited to the Xact OMXS30 ETF, over the period 2012-2019. As the Xact OMXS30 is reinvesting dividends, we use the OMXS30 total return index as the basis for our analysis.

4.2.1. Data retrieval

We obtain the high frequency data for the OMXS30 index and the related ETF prices of the Xact OMXS30 ETF through Bloomberg. Due to data availability restrictions before 2012, our data ranges from March 2012 to March 2019 to encompass a sample period of 7 years. The frequency of the data is on a minute to minute basis. Further, we are able to leverage Bloomberg to obtain the number of ETF shares outstanding for the Xact. For additional ETF data, we use the Thomson Reuters Eikon database. This data includes the daily NAV of the Xact, as well as its trading volume and VWAP.

Whereas the index price displays continuous price changes, the ETF price shows time periods of consistent prices (on average a price change every 6.1 minutes). As Bloomberg only reports prices for timepoints when the price has changed, prices in the ETF of single timepoints have been assumed for subsequent timepoints until a new price was reported. For any calculations, the "Close" price as reported by Bloomberg has been used. To calculate the daily volatility of the index, the first-of-the-day returns have been removed in order to not distort the actual intraday volatility through after-hours trading.

4.2.2. Data cleaning and sample characteristics

When analysing the volatility over the period 2012-2019, we find that it has been relatively stable, with the exception of two periods: 2012 and 2015-2016 (see *Figure 3*). Significant macroeconomic events happened in these periods. The start of 2012 encompassed the latter period of the Eurozone crisis, whereas several events occurred in 2015 that moved the market. A negative repo-rate materialized for the first time in Sweden's history in February 2015, which was followed by the Greek bailout referendum in July 2015. Further, the Federal Reserve in America raised the Federal Funds target rate for the first time following the Global Financial Crisis in December 2015. We believe these events go a long way in explaining the periods of raised volatility over our sample period

From *Figure 3*, we can see the close relationship between the relative trading volume (to the number of shares outstanding) of the Xact OMXS30 ETF and the OMXS30 index intraday volatility. The 60-day trailing average measures of the two show a correlation of 0.497 over our sample period.

Next, we observe the activity levels of APs in C/R activities. *Figure 4* shows the number of days on which a change in the number of Xact shares outstanding occured. A change in ETF shares only happens through AP C/R flows, so-called ETF primary market trading activities. From *Figure 4*, we can see that the activity has increased relatively consistently over time.

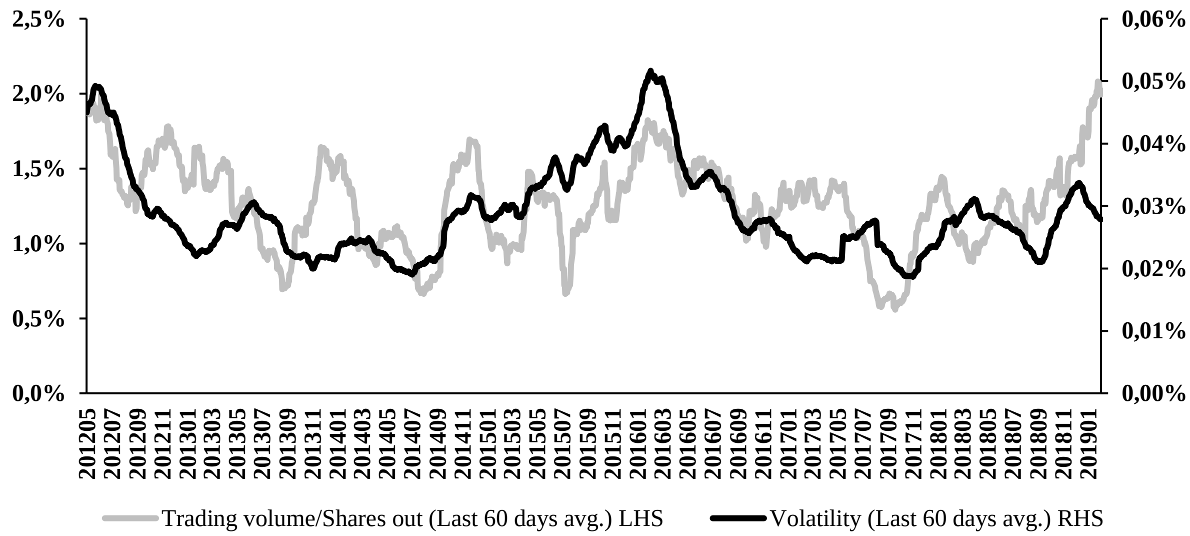

Figure 3. OMXS30 volatility and Xact OMXS30 relative trading volume
The figure shows the development of 60-day average trailing volatility for the OMXS30 index, calculated based on minute-to-minute returns, and the 60-day average trailing trading volume of the Xact OMXS30 ETF as a proportion of the number of shares outstanding.

Over the period, the average number of days per quarter with AP C/R flows was 10.1 days, implying 3.4 days of C/R flows per month. This observation, and the implied low activity of APs, indicates that we may not find support for our primary market hypotheses I.a and I.b.

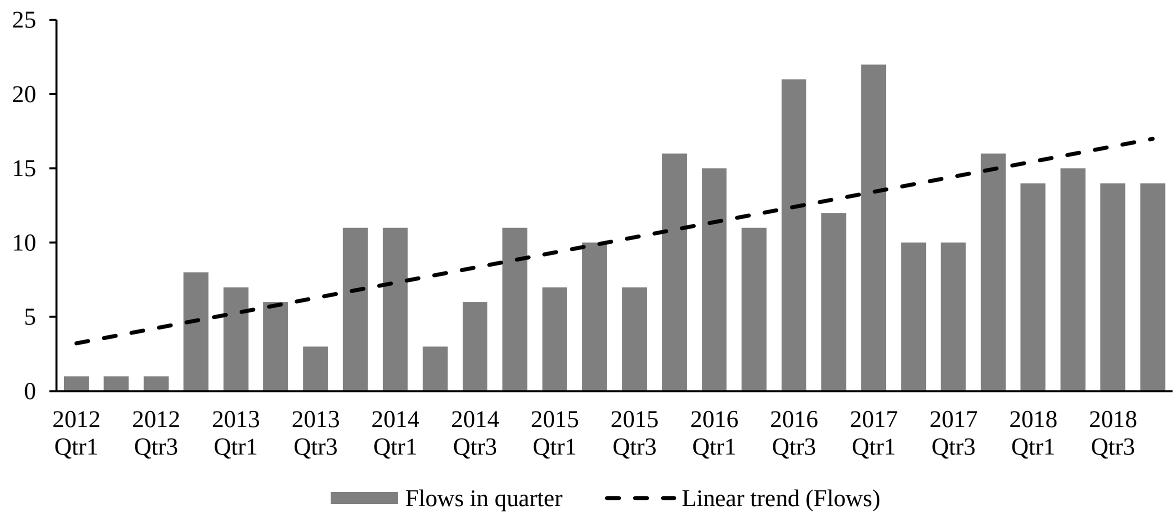

Figure 4**. **Number of days per quarter with AP C/R flows
The figure shows the number of days within each quarter between 2012-2018 in which a net change in the number of ETF shares outstanding occurred.

5. Stock Level Analysis

5.1. Methodology

In order to identify the impact of the increase in ETF ownership on the volatility of returns in the Swedish market, we initially set out to prove the effect on the individual stock level. Given the limited maturity of the ETF market in Sweden, we focus on the largest stocks on the Swedish stock exchange, as these stocks are the likeliest to attract ETF investment. In our initial regression, we focus on the methodology established by Ben-David et al. (2018), which focuses on the American stock market, and apply it to the largest stocks in Sweden.

We perform a panel regression, regressing daily volatility on ETF ownership, and limit the potential omitted variable bias by including lagged control variables. The control variables used follow the example of Ben-David et al. (2018), and encompass past 12-month return, inverse share price, Book-to-Market, Amihud (2002) illiquidity measure, market capitalization of the stock, and other funds' ownership of the stock. Additionally, we add three lagged factors of the dependent variable.

The rationale for including the control variables relates to the potentially spurious correlation between ETF ownership and stock volatility caused by variation in ETF ownership that is not correctly captured in the data, as updated ETF ownership data is available only every three months, as noted by Ben-David et al. (2018). In their paper, they mention three potential sources of spurious correlation. Firstly, for equal-weighted ETFs, when weights of different stocks in the ETF portfolios do not develop proportionally to the market capitalization of the stock, a spurious link could exist between the volatility of returns and ETF ownership, due to the correlation between volatility and stock size. In order to control for the potential issue this causes, Ben-David et al. use logged market capitalization as a control variable. However, Ben-David et al. have defined the ETF ownership through a assets-under-management (AUM) measure, which enables the potential divergence in the ownership measure. Due to differing data providers, with different data availability, we define ETF ownership on the basis of share position rather than AUM. As such, our definition of ETF ownership prevents the issue faced by Ben-David et al., yet we have included logged market capitalization as a control variable for completeness. Secondly, Ben-David et al. note the potential existence of endogeneity issues, arising from the number of ETFs covering a stock. The authors note that demand for ETFs may depend on the underlying popularity and attractiveness of the underlying sector and/or asset, which in turn may impact the trading intensity, and thus, the volatility of the underlying securities. Through this mechanism, the causal effect of ETF ownership on underlying stock volatility may be impacted. Including the inverse stock price and the Amihud (2002) illiquidity variable (see equation (2)) allows us to control for stock size and liquidity. For completeness, indicators of stock returns that may be related to volatility are included. Hence, we include the Book-to-Market (BtM) ratio and past 12-month return. The rationale for including the BtM is justified by results found by Fama and French (1998). In their study, they find that large and high BtM stocks underperform large and low BtM stocks due to higher return volatilities for large and high BtM stocks. By including other fund ownership (similarly defined as ETF ownership, see equation (1)), we are able to distinguish between the ETF ownership's unique effect on volatility relative to other fund ownership. Further, in line with Ben-David et al. (2018), we also include the lagged volatility variables to address the potential issue related to how persistence in volatility may bring about reverse causality. Finally, for robustness, date and stock-level fixed effects are included.

Volatility is calculated using the daily log returns based on closing prices within a month. The ETF (and other) ownership variables for company i at time t are defined as the sum of the number of shares owned by different funds j relative to the total shares outstanding of that company:

$$ETF\ ownership_{(i,t)} = \sum_{j=1}^{n} \left(\frac{ETF\ owned\ shares_{(i,j,t)}}{Shares\ outstanding_{(i,t)}} \right) \tag{1}$$

Since the ETF ownership is measured in number of shares, stock splits are not an issue, as the numerator and denominator are affected in the same proportion. The shares outstanding are calculated using market capitalization and daily closing share price data. Past 12-month return is calculated as the absolute return based on the share price in the period and the share price in the period 12 months prior. For the monthly share price used in the return calculation, the average daily share price within the month is used. The inverse share price, Book-to-Market, market capitalization, and Amihud variables also represent the average of the daily measures within the month. The daily Amihud illiquidity measure is calculated using the absolute returns based on closing prices, the volume of shares traded, and the value-weighted average share price.

$$Amihud_{(i,t)} = \left| \frac{Closing\ price_{(i,t)} / Closing\ price_{(i,t-1)}}{Trading\ volume\ (shares)_{(i,t)} * VWAP_{(i,t)}} \right| \tag{2}$$

The dependent variable volatility (and its lagged explanatory variables), and the explanatory variables ETF ownership and other ownership, are standardized over the whole period in order to aid economic interpretation. The variables are standardized by subtracting the mean and dividing by the standard deviation of the observations. The panel regression we perform is defined as follows:

$$Volatility_{(i,t)} = \beta_1 ETF\ Ownership_{(i,t)} + \beta_i controls + fixed\ effects + \ \varepsilon_{(i,t)} \qquad (3)$$

Company and time fixed effects have been included in the regression to control for company specific and time-dependent omitted variables. Furthermore, the standard errors have been clustered at both the stock and date level.

5.2. Results

We now continue by reporting the results of the regression lined out in 5.1 (*Table 1)*. The analysis for the impact of ETF ownership of stocks on the volatility of daily returns in the Swedish setting reveals similar findings to the relationship found in other markets. In line with Ben-David et al. (2018), we find a statistically significant positive relationship between ETF ownership and the volatility of daily returns on the stock level. We continue by discussing the results and comparing them in detail to the findings by Ben-David et al. (2018) of the American stock market.

When regressing volatility only on ETF ownership and controlling for stock returns, size and illiquidity characteristics (column 1), we find a weakly significant relationship between ETF ownership and volatility. Although the relationship is positive, it differs in significance to findings reported in the American stock market. Ben-David et al. (2018) report a highly statistically significant (1% level) positive relationship. However, this may be due to the usage of a smaller sample size. Whereas our sample contains only 38 stocks, the S&P 500 contains 500, including companies with more broadly different characteristics. Further, when looking at the magnitude, we find that a one standard deviation unit increase in ETF ownership is associated with an increase of 7.2% of a standard deviation in daily volatility. The authors find, on the other hand, that a one-unit increase in standardized ETF ownership of a stock is associated with an increase in standard deviation of daily volatility of 16.4% for stocks in the S&P 500. Hence, we observe a much more muted effect in our sample.

Table 1: Regression Table: ETF ownership on individual stocks' return volatility

The table reports estimates from the panel regression of volatility on ETF ownership and controls. The frequency of the observations is monthly, and volatility is calculated using the log daily returns within the month. The control variables in columns (1) to (3) are lagged past 12-month return, lagged inverse stock price, lagged Book-to-Market, lagged Amihud (2002) illiquidity measure, and lagged logged market capitalization. In column (2), other fund ownership of the stock is added as a control variable. Column (3) shows results when further controlling for lagged volatility. Amihud is scaled by 10^6. The dependent variable and the ownership variables are standardized by subtracting the mean and dividing by the standard deviation. Standard errors are double-clustered at the company (stock) and monthly level, and the legend for the statistical significance of observations is explained at the bottom of the table. The sample covers the period April 2012 until December 2018 (January-March 2012 is excluded due to the absence of lagged dependent variables).

	Dependent variable:		
	Volatility		
	(1)	(2)	(3)
ETF ownership	0.072^{*}	0.078^{*}	0.081^{*}
	(0.044)	(0.044)	(0.043)
Amihud (t-1)	6.395^{**}	6.311^{**}	2.863
	(2.830)	(2.864)	(2.571)
Book-to-Market (t-1)	-0.205	-0.208	-0.203
	(0.208)	(0.208)	(0.156)
Inverse price (t-1)	10.064^{*}	9.818^{*}	7.108
	(5.745)	(5.712)	(4.601)
log(Market cap (t-1)	-0.640^{**}	-0.664^{**}	-0.399^{*}
	(0.321)	(0.321)	(0.225)
Past 12M returns (t-1)	-0.004	-0.005	-0.013
	(0.043)	(0.043)	(0.035)
Other fund ownership		-0.022	-0.018
		(0.034)	(0.025)
Volatility (t-1)			0.119^{***}
			(0.035)
Volatility (t-2)			0.047^{*}
			(0.025)
Volatility (t-3)			0.159^{***}
			(0.037)
Observations	3,075	3,075	3,075
R^2	0.460	0.460	0.486
Adjusted R^2	0.437	0.438	0.464
Residual Std. Error	0.751 (df = 2951)	0.751 (df = 2950)	0.733 (df = 2947)
Note:			*p<0.1; **p<0.05; ***p<0.01

We want to verify the validity of our findings, by including additional control variables. In column (2) we report the effect of including other fund ownership, whereas column (3) displays the results when adding lagged dependent variables. We note that the significance of the ETF ownership-volatility relationship persists, at a slightly increased level. When controlling for other fund ownership, the coefficient increases by almost 10%, and a standard deviation increase in ETF ownership is associated with an increase in daily volatility of 7.8%

of a standard deviation. The corresponding change when including all control variables increases further to 8.1%. This effect is of similar magnitude to the one found by Ben-David et al. (2018) in the S&P 500 when controlling for all variables. They report a corresponding coefficient of 7.7%, which implies that the largest stocks in each market experience similar volatility effects from increased ETF ownership.

We proceed to discussing the observations of the control variables. When observing the control variables' impact on volatility, our findings indicate statistically significant predictive power on volatility for Amihud illiquidity (positive coefficient), inverse price (positive coefficient), market capitalization (negative coefficient), as well as lagged dependent variable factors (positive coefficients). These findings are consistent with the findings of Ben-David et al. (2018) in both direction and significance for the S&P 500 and Russell 3000 stocks. Hence, given that all the significant coefficients in this regression are in line with the coefficients estimated by Ben-David et al, we are confident in the validity of this regression.

However, there is a significant difference in our other fund ownership measures, due to infeasibility of obtaining a more granular fund ownership classification in our dataset. Ben-David et al. (2018) are able to granularize other ownership into index-, active- and hedge funds. While our findings fail to identify any predictive power for the other ownership variable, Ben-David et al. find predictive power for active fund ownership and hedge fund ownership. The statistical significances and directions of the coefficients differ both in direction and magnitude within Ben-David et al.'s sample. Whereas index fund (statistically insignificant) and active fund ownership (statistically significant at 1% level) are associated positively with volatility, the relationship between volatility and hedge fund ownership (statistically significant at 1% level) is negative. However, given that our analysis clusters other ownership, not separating the ownership of index-, active- and hedge funds, we cannot reject the hypothesis that other funds' ownership has predictive power on volatility in our sample. Through not granularizing the type of ownership in our analysis, the comparison to Ben-David et al. is convoluted, and may explain the negative coefficient and lack of statistical significance of the aggregated measure of other ownership in our sample.

Henceforth, we conclude this analysis with the result, that ETF ownership does have a positive, and statistically weakly significant impact on volatility on the stock level, in line with previous research. As we find the statistically significant link between ETF ownership and daily volatility of the stocks, we so far find support for our main testable hypothesis and move on to the index level to test for our four sub hypotheses.

6. Index level analysis

6.1. Total Volatility

Having established the link between ETF ownership and volatility on the stock level, we aim to replicate the analysis on the index level. We do this for three reasons.

Firstly, on the stock level, the source of the volatility is convoluted. Evidently, the effect of the primary market activities is captured by the ETF ownership measure, as an increase in the ETF ownership of a stock is the result of C/R flows into the ETF. ETF shares are created in exchange for the underlying shares – as defined in the creation basket – and subsequently the ownership of the underlying securities is transferred to the ETF, thus increasing the position of the ETF in that stock. However, APs usually create new ETF shares in response to a high market demand for the ETF. Furthermore, APs often take the opposite side in an ETF trade, as they can easily liquidate their ETF positions. As such, a change in ETF ownership is not only linked to primary market trading activity, but also to higher trading volume (Wang and Xu, 2019). Consequently, our explanatory variable of ETF ownership used in the stock level regression captures both the primary and secondary market trading effects as described in the hypothesis development. However, on an index level, we can separate these effects, by using both C/R flows and trading volume as independent explanatory variables, as we have a direct match between the ETF and the underlying securities. This requirement for a direct match between data for the ETF and the underlying securities prevents us from separating the primary and secondary market trading effects on the stock level. This is because the effect of both the C/R flows and trading volume of broader ETFs like an EAFE ETF (such as iShares MSCI EAFE ETF) are not directly attributable to only parts of their underlying stocks that are included in our sample (such as Volvo), which would lead to spurious results.

Secondly, we want to be able to decompose the volatility into its fundamental part (related to an efficient price and efficient market) and its noise part (disturbances in the market that make the market less efficient) and rerun the regression. To identify the noise, we use the Variance Ratio (Lo and MacKinlay, 1988) and for the fundamental volatility we use the Beveridge Nelson decomposition (Beveridge and Nelson, 1981). Both require high frequency data that we only have available for the OMXS30 Index and not for the individual stocks.

Thirdly, to reinforce our findings, we want to analyse the price discovery that happens in the ETF compared to the index itself (i.e. the underlying stocks) by calculating the Information

Share (Hasbrouck, 1995). For this, we need cointegrated high frequency data sets, which we have available for the Xact and the OMXS30 Index.

6.1.1. Methodology

We aim to establish a link between volatility and the primary and secondary market activities related to ETFs in order to identify the cause of the increased volatility attributable to ETFs.

Firstly, to establish the link between the ETF ownership variable used at the stock level to capture the effect of ETFs, and an analogous measure on the index level, we mathematically decompose the ownership into its share component and value component. Recall from (1) that we define ETF ownership on the stock level as the number of shares held by ETFs divided by the number of total shares outstanding. Whereas this measure omits the pricing component, on the stock level, it is irrelevant, as the shares have one common price. On the other hand, given that an ETF is a separate security to the index, differences in the relationship between index capitalization and the NAV of the ETF may occur. Thus, we can define ETF ownership on the index level, analogously to the stock level, while including the pricing factor, as:

$$ETF\ Ownership_t = \left(ETF\ shares\ outstanding_t * \frac{ETF\ Net\ Asset\ Value_t}{OMXS30\ Index_t}\right) \quad (4)$$

We can rewrite it for ease of interpretation as:

$$Own_t = \left(Shares_t * \frac{NAV_t}{Index_t}\right) = (Shares_t * X_t) \quad (5)$$

The use of company date fixed effects on the stock level effectively results in the ETF ownership variable being a measure of change. This can be defined as:

$$\Delta Own_t = Own_t - Own_{t-1} \quad (6)$$

From (5) follows

$$\Delta Own_t = \Delta Shares_t * X_{t-1} + Shares_{t-1} * \Delta X_t + \Delta Shares_t * \Delta X_t \quad (7)$$

By making the assumption that ΔX_t is negligible, that is, the change in the relationship of ETF NAV to Index capitalization is almost 0 (which seems justified, as the ETF is supposed to closely track the index (additionally, in numbers ΔX_{t+1} is on average only 0.00176% of X_t)), we obtain:

$$\Delta Own_t = \Delta Shares_t * X_{t-1} \tag{8}$$

Where

$$\Delta Shares_t = Shares_t - Shares_{t-1} \tag{9}$$

As a change in shares outstanding only occurs due to C/R flows, this captures the primary market trading activities of APs. As such, we define our primary market trading variable as:

$$Flow_{(t)} = \left| ETF\ shares\ outstanding_{(t+1)} - ETF\ shares\ outstanding_{(t)} \right| \tag{10}$$

By using flow, we have obtatin a variable that allows us to observe the change in ETF ownership, which is in line with our stock level regression, and simultaneously allows us to capture primary market activity. We take the absolutes of flow, because both the creation and redemption processes present shocks to the underlying securities that may influence volatility in the same direction (i.e. increase it).

As the settlement of ETF C/R flows occur after trading hours, this activity is only observable in the next trading day's disclosed number of ETF shares outstanding. However, what the flow captures are actions taken by the AP during the trading day which impact volatility. To exemplify this, consider an event where an AP accepts an ETF buy-order (i.e. sells the ETF). In the case where an AP does not hold the ETF on their books, they must create the security, by exchanging the basket of the underlying with the financial sponsor for a share in the ETF. This requires the AP to acquire the basket of the underlying securities, thus constituting a shock to these, which in turn impacts the volatility. Hence, we utilize flow as a contemporaneous variable in the regression.

There are several reasons market participants trade ETFs, such as incorporating their belief on the market movement or liquidity trades. Regardless, according to basic market principles, the ETF price will move in response to these trades. In turn, this may cause a pricing discrepancy between the ETF price and its NAV and create arbitrage opportunities for other market participants. As these arbitrage opportunities require a trading of the underlying securities, the ETF propagates demand shocks into the underlying stocks, and thus influences their price and consequently the volatility. Hence, a higher trading volume should capture the ETF's impact on volatility through secondary market trading activities. As such, we define:

$$Trading\ Volume_{(t)} = Number\ of\ ETF\ shares\ traded_{(t)} \tag{11}$$

Since these trades in the ETF impact volatility, a higher trading volume is expected to result in higher volatility of the index. Further, the trading volume is determined contemporaneously to volatility, as it is the trading volume in the ETF during the day which is expected to contribute to the volatility of the index on the given day.

Given the stock and time fixed effects used for robustness in our analysis on the stock level, we effectively evaluate the changes in the explanatory variables relative to the changes in the dependent variable. This also serves the purpose of detrending the variables, preventing spurious results. For consistency, since we do not control for fixed effects on the index level analysis (due to the usage of a single ETF), we must detrend the variables explicitly on the index level. Our index level regression, used to identify the source of the volatility, is defined as:

$$\Delta Volatility_{(t)} = \alpha + \beta_1 Flow_{(t)} + \beta_2 \Delta Trading\ Volume_{(t)} + \beta_{(i)} \Delta Controls_{(i,t)} + \varepsilon_{(t)} \qquad (12)$$

As explained above, flow is in itself a difference-measure, measuring the change in ETF shares outstanding. Due to this, it suffices to include it as a level measure. The first of our control variables is past 12-month return of the index. As this is a general predictor of returns, and thus related to volatility, we want to keep it exogenous, and thus we include the lagged variable. Secondly, we control for autocorrelation by including three lagged dependent variables. Additionally, including the lagged dependent variables allows us to control for potential issues related to reverse causality, in line with the analysis on the stock level.

6.1.2. Results

We start by reporting the results for our index level regression of OMXS30 index intraday volatility on flow, Xact ETF trading volume, and controls in *Table 2* column (1).

When controlling only for lagged past 12-month return of the index, we obtain a positive, statistically significant impact of trading volume on intraday volatility, implying that higher trading volume during the day is related to higher volatility. This result supports our hypotheses II.a and II.b that higher activity in the ETF secondary market is associated with an increase in volatility, without taking a stand on the nature of the volatilities. Empirical findings from previous research lend their support to our findings. Xu and Yin (2017b), when studying the American market, find that the contemporaneous trading volumes of the ETFs that track the S&P 500 index are a key determinant in the volatility of the index.

Table 2. Regression Table: Index Intraday Volatility

The table reports estimates from the time series regression of the change in intraday volatility of the OMXS30 index on ETF creation/redemption flows, the change in ETF trading volume, and controls. The frequency of the observations is daily, and volatility is calculated using the minute-by-minute log daily returns within the day where first of the day returns have been removed. In column (1) we control for lagged past 12-month return, and to account for potential reverse causality, column (2) contains three lagged dependent variables as explanatory variables. Due to the usage of a single ETF, standard errors are unclustered. The intraday volatility variable and its lagged variables have been scaled by a factor of 10^3, whereas the change in trading volume is expressed in millions (scaled by 10^-6). The sample covers the period March 2012 until March 2019. The legend for the statistical significance of observations is explained at the bottom of the table.

	Dependent variable:	
	ΔIntraday volatility	
	(1)	(2)
Flow (t)	0.004	0.001
	(0.005)	(0.004)
ΔTrading volume (t)	0.008^{***}	0.006^{***}
	(0.002)	(0.002)
ΔPast 12M return (t-1)	0.098	-0.452^{***}
	(0.148)	(0.127)
ΔIntraday volatility (t-1)		-0.625^{***}
		(0.024)
ΔIntraday volatility (t-2)		-0.362^{***}
		(0.026)
ΔIntraday volatility (t-3)		-0.151^{***}
		(0.024)
Constant	-0.001	-0.000
	(0.003)	(0.002)
Observations	1,746	1,746
R^2	0.007	0.287
Adjusted R^2	0.005	0.285
Residual Std. Error	0.108 (df = 1742)	0.092 (df = 1739)
F Statistic	3.919^{***} (df = 3; 1742)	116.762^{***} (df = 6; 1739)
Note:		*p<0.1; **p<0.05; ***p<0.01

On the other hand, we fail to obtain the statistically significant relationship between flow and intraday volatility observed by Wang and Xu (2019). Our three main assumptions for the different observations revolve around the difference in data set characteristics. Firstly, they look at a significantly higher number of ETFs, utilizing 70 Chinese ETFs, whereas we are limited to the only existing OMXS30 ETF given our need for a one-to-one mapping with the underlying index. This also limits the amount of observable flows, as C/R flows for the Xact only occur on average every 6th trading day. Secondly, their study covers the period January 2015 to

December 2017, whereas our study spans a significantly longer period (2012-2019). Given our observation of increased volatility in 2015-2016 in the Swedish market, looking at a limited period may yield results that are invalid over a longer period. Thirdly, there may be differences in the relative maturities and workings of the Swedish and Chinese ETF markets, affecting the relative importance of ETF flows. Overall, given the lack of the relationship in our data, we obtain our first indications against our hypotheses I.a and I.b that AP redemption and creation flows contribute to the volatility of the underlying index.

In column (2) we report the results when additionally including three lags of the dependent variable. In line with the preceding analysis, we find continued support for the ETF secondary market trading hypotheses, as trading volume's predictive power remains positive and highly statistically significant. An increase in trading volume of a thousand shares is associated with an increase in intraday volatility of 0.6%, 5.0%[2] of a standard deviation of absolute intraday volatility. Perhaps unsurprisingly given the results in column (1), the effects of the ETF primary market trading activities (flow) remain statistically insignificant predictors of intraday volatility.

When lagged dependent variables are included, lagged past 12-month returns obtain predictive power on the intraday index volatility. The statistical significance emerges only after inclusion of the lagged variables, with a change in sign of the coefficient. The findings are partially in line with previous literature (i.e. Ben-David et al. (2018)). The predictive power of the past 12-month returns seems to be a market-specific phenomenon, as the measure possesses predictive power on volatility for Ben-David et al.'s sample of Russell 3000 stocks, but not for S&P 500 stocks, over the period 2009-2015. Thus, we think the finding is not indicative of spurious results. Finally, the 0-value on the regression constant and its associated lack of statistical significance increases our confidence in the successful detrending of the variables, and thus, in our results.

We now summarize our findings from the index level analysis of the source of the volatility. Our findings imply that the increased volatility of the returns we observe on the individual stock level in the previous regression, attributable to a change in ETF ownership, seems to stem from the supply and demand shocks in the secondary ETF trading market, as higher trading in this market is related to a higher volatility of the underlying. We do, however,

[2] Coeff(ΔTrading volume) /10^3 / St. Dev (ΔIntraday volatility(abs.)) = St. Dev Δ (Intraday volatility(abs.)) = (0.006)/10^3 / (1.21*10^-4) = 0.0413
Note: Coefficient in *Table 2* represents trading volume in millions of shares, whereas Intraday volatility has been scaled by a factor of 10^3

need to be cautious in establishing causality from these results, as the relationship might be bidirectional. This means, that while the trading volume of an ETF influences the volatility of the index it aims to track, the volatility itself may also influence the trading volumes of said ETF. This is implied by the findings of Xu and Yin (2017b), who find evidence for a two-way Granger causality between ETF trading volumes and the underlying index volatility for the S&P 500 index, and for indices in Australia, France, Japan and the UK. This, they claim, demonstrates the robustness of the finding in the S&P 500, and the finding's applicability to other indices (such as the OMXS30 analysed in our study). Finally, we find no evidence of the impact of AP C/R flows on the underlying securities' volatility. Thus, we have found support for the ETF secondary market trading hypotheses (hypotheses II.a and II.b), while our findings thus far indicate lack of support for the primary market trading hypotheses I.a and I.b.

6.1.3. Mispricing Impact

In an extension of this analysis we want to analyse the effect that the relative mispricing of the ETF at market close has on the volatility of the market on the next day. This comes as a natural expansion of our previous analysis, given that our hypotheses are based on the idea, that volatility increases through arbitrage channels. Consequently, when there is a higher mispricing between the NAV and the price of the ETF, there is more opportunity for arbitrage. As such, we define Mispricing as

$$Mispricing_t = \left| \frac{NAV_t}{ETF\ Price_t} - 1 \right| \tag{13}$$

where the NAV and ETF price are taken from the end of the trading day. We want to take an absolute measure of this value, so that both a higher NAV than price and lower NAV than price increase the mispricing factor. Thus, if Mispricing is 0, the NAV of the ETF perfectly matches its market price and the market for the ETF is efficient. On the other hand, if there is a high Mispricing value, the market can be assumed to have been less efficient during the trading day (or the last minutes), and there is necessarily more arbitrage potential on the following day. Continuing this logic, a higher Mispricing should have a positive moderating effect on the effect that trading volume has on the next day's volatility. As such, we will multiply contemporaneous trading volume with the Mispricing factor of the previous day and use it as an explanatory variable for volatility. In a previous step we add 1 to Mispricing, to prevent spurious results (as the whole term would become 0 when Mispricing is 0) and ease interpretation. Results are shown in

The results show, that while Mispricing itself does not have a significant effect on the next day's volatility, it does have a positive moderating effect on the relation trading volume has towards volatility. The regression estimates, that the effect of a change in trading volume on change in volatility is magnified by a higher Mispricing factor. This can be seen when looking at the coefficients. The regression estimates the following model:

$$\Delta Vol_t = \cdots + \beta \Delta TV_t + \gamma Mispricing_{t-1} + \delta(\Delta TV_t * Mispricing_{t-1}) + \cdots \tag{14}$$

$$\Delta Vol_t = \cdots + \Delta TV_t(\beta + \delta Mispricing_{t-1}) + \gamma Mispricing_{t-1} + \cdots \tag{15}$$

We can just focus on the interaction term for the trading volume. The interaction term $(\beta + \delta Mispricing_{t-1})$ for ΔTV_t is always positive, because $\delta > -\beta$[3] and $Mispricing_{t-1} \geq 1$. Consequently, the higher the Mispricing is, the higher is the impact of change in trading volume on volatility. Related interactions were identified by Wang and Xu (2019). They find, that higher mispricing stimulates the effect that the consequent trading volume of newly created ETF shares has on the volatility of the underlying market's volatility. As such, our findings are in line with the findings of previous research conducted in the Chinese market.

Given these results, we now proceed from identifying the source of the volatility increase from ETF ownership to the analysis of the nature of the volatility (noise contribution versus fundamental volatility), as measured by the Variance Ratio and Beveridge Nelson decomposition, respectively, in order to distinguish between hypotheses II.a and II.b. What follows is an examination of the theoretical frameworks of each measure, before we report our findings.

[3] $1.083 > -(-1.079)$

Table 3. Regression Table: Index Intraday Volatility with Mispricing

The table reports estimates from the time series regression of the change in intraday volatility of the OMXS30 index on ETF creation/redemption flows, the change in ETF trading volume, and controls. The frequency of the observations is daily, and volatility is calculated using the minute-by-minute log daily returns within the day where the first of the day returns have been removed. In column (1) we control for lagged past 12-month return, Mispricing and lagged dependent variables. Column (2) includes an interaction term, showing the magnified effect of trading volume when an arbitrage existed the previous day. The intraday volatility variable and its lagged variables have are scaled by a factor of 10^3, whereas the change in trading volume is expressed in millions (scaled by 10^-6). The sample covers the period March 2012 until March 2019. The legend for the statistical significance of observations is explained at the bottom of the table.

	Dependent variable:	
	ΔIntraday volatility	
	(1)	(2)
Mispricing (t-1)	0.042	3.207
	(0.637)	(1.995)
ΔTrading volume (t)	0.006***	-1.079*
	(0.002)	(0.648)
Flow (t)	0.001	0.001
	(0.004)	(0.004)
ΔPast 12M return (t-1)	-0.452***	-0.444***
	(0.127)	(0.127)
ΔIntraday volatility (t-2)	-0.625***	-0.626***
	(0.024)	(0.024)
ΔIntraday volatility (t-3)	-0.362***	-0.364***
	(0.026)	(0.026)
ΔIntraday volatility (t-1)	-0.151***	-0.151***
	(0.024)	(0.024)
Mispr.(t-1) * ΔTrad. vol.(t)		1.083*
		(0.647)
Constant	-0.042	-3.211
	(0.637)	(1.997)
Observations	1,746	1,746
R^2	0.287	0.288
Adjusted R^2	0.284	0.285
Residual Std. Error	0.092 (df = 1738)	0.092 (df = 1737)
F Statistic	100.025*** (df = 7; 1738)	87.963*** (df = 8; 1737)
Note:		*p<0.1; **p<0.05; ***p<0.01

6.2. Variance Ratio

6.2.1. Methodology

Although we have shown the relationship between ETFs and the volatility, we have not analysed this volatility in terms of its market efficiency. As previous research suggests, such as Wermers and Xue (2015), ETFs function as a price discovery channel and may thus not necessarily add additional noise to the market. Hence, we need a way to distinguish between the additional noise stemming from ETFs and the price discovery function of the ETFs. Although it is inherently difficult to observe the noise, the approach of using the Variance Ratio as proposed by Lo and MacKinlay (1988) offers a solution. The idea behind this methodology is, that if the stock price follows a random walk, then the variance of q-interval log returns should be q times as large as the one-interval log returns. As such, let us define P_t as the stock price, and p_t as its logarithmic value:

$$p_t = \ln P_t \tag{16}$$

We can then estimate the mean of the one-interval return as

$$\hat{\mu} = \frac{1}{n}\sum_{j=1}^{n}(p_j - p_{j-1}) \tag{17}$$

and consequently, estimate the one-interval variance as

$$\hat{\sigma}_1^2 = \frac{1}{n-1}\sum_{j=1}^{n}(p_j - p_{j-1} - \hat{\mu})^2 \tag{18}$$

The estimator of the variance of a q interval is

$$\hat{\sigma}_2^2 = \frac{1}{\left(\frac{n}{q}\right) - 1}\sum_{j=q}^{n}(p_j - p_{j-q} - q\hat{\mu})^2 \tag{19}$$

Now we can define the Variance Ratio (VR) as

$$VR = \left|\frac{\hat{\sigma}_2^2}{q * \hat{\sigma}_1^2} - 1\right| \tag{20}$$

which should be 0 if the market follows a random walk and is efficient. Therefore, according to O’Hara and Ye (2011) we can use the VR to measure the noise in the market: The further the

VR deviates from 0, the less efficient is the market and the higher is the noise contained in the market.

To measure the effect that ETFs have on the VR and thus the noise in the market, we calculate the VR with intraday data, as the noise could increase through arbitrage activities from various market players throughout the day. Thus, we will use 1-minute intervals and select q as 3, such that multi-period intervals reflect 3-minute intervals, based on Ben-David et al. (2015) choosing q as 3 to measure intraday noise. The first returns of the day – that is the return from closing price in t-1 to 9am opening price in t – have been removed to avoid spoiling the actual intraday VR with after-hours trading activities. The sampling frequency is daily and the sample period ranges from March 2012 until March 2019. To then analyse the effect of ETF flows and trading volume on noise, we run the index level regression with VR as the new dependent variable. Summary statistics for the VR are shown in *Table 12.*.

6.2.2. Results

We report the results of our analysis of the relationship of ETF primary and secondary market trading activities with noise in the market in *Table 4*.

In line with our previous regression on the total volatility, we obtain further evidence against the hypothesis of the noise primary market trading hypothesis. Our analysis reports no statistically significant relationship between AP C/R flows and the change in the VR.

On the other hand, the link between ETF trading volume and intraday volatility captured earlier seems to be in part related to additional noise in the market. Our regression returns a positive, statistically significant relationship between the change in the Variance Ratio and the change in the ETF trading volume relative to the prior trading day. An increase in the change in daily trading volume of a million shares is associated with an increase in the change of the VR of 0.5%, representing of 5.7% of a standard deviation of the VR[4]. However, the VR itself does not have a creamy economic interpretation, and thus the magnitude of the coefficient is of secondary importance to the establishment of the relationship. The relationship is significant at the 5% level when controlling for the change in lagged past 12-month returns. Further, when adding lagged dependent variables, the positive relationship persists, however significant only

[4] Note: In ***Table 4***, trading volume represents trading volume in millions of shares. The calculation uses the level, and not the change, of the standard deviation of the VR. See footnote 2 for calculation.

at the 10% level. The associated increase in the change in daily trading volume of a million shares in the change of VR is 0.4%, corresponding to 4.6% of a standard deviation of VR.

Relating these findings to previous literature, Ben-David et al. (2018) find that the VR increases in conjunction with increased ETF ownership. As we conjecture that higher ETF ownership leads to higher ETF trading volume, this is consistent with our findings.

Table 4. Regression Table: Index Level Noise

The table reports estimates from the time series regression of the change in the market noise of the OMXS30 index on ETF creation/redemption flows, the change in ETF trading volume, and controls. The frequency of the observations is daily, and the noise is measured by the Variance Ratio as proposed by Lo & MacKinlay (1988) with minute-to-minute data and q = 3. In column (1) we control for lagged past 12-month return, and to account for potential reverse causality, column (2) contains three lagged dependent variables as explanatory variables. Due to the usage of a single ETF, standard errors are unclustered. The change in trading volume is in millions (scaled by 10^-6). The sample covers the period March 2012 until March 2019. The legend for the statistical significance of observations is explained at the bottom of the table.

	Dependent variable:	
	ΔVariance Ratio	
	(1)	(2)
Flow (t)	-0.000	0.002
	(0.005)	(0.004)
ΔTrading volume (t)	0.005^{**}	0.004^{*}
	(0.003)	(0.002)
ΔPast 12M return (t-1)	0.317^{*}	0.051
	(0.163)	(0.132)
ΔVariance Ratio (t-1)		-0.702^{***}
		(0.023)
ΔVariance Ratio (t-2)		-0.462^{***}
		(0.026)
ΔVariance Ratio (t-3)		-0.247^{***}
		(0.023)
Constant	-0.000	-0.000
	(0.003)	(0.002)
Observations	1,746	1,746
R^2	0.004	0.347
Adjusted R^2	0.003	0.345
Residual Std. Error	0.119 (df = 1742)	0.097 (df = 1739)
F Statistic	2.506^{*} (df = 3; 1742)	154.005^{***} (df = 6; 1739)
Note:		*p<0.1; **p<0.05; ***p<0.01

Further, Wermers and Xue (2015) classify ETF trades into informed trades and noise trades. In their analysis, they find that the predictive power of noise trades on the index disappears within three minutes, suggesting that this subset of trades contributes to noise in the market. In light of this, it should not be surprising that we find a statistically significant relationship between noise in the market and ETF trading volume. The lack of indications for an effect of AP C/R flows on volatility, and specifically noise, in our findings continues to be somewhat puzzling, given widespread evidence of this in the previous literature. Staer (2017) finds, that 38% of the price changes in the underlying securities that are associated with C/R flows revert after 5 days. In essence, this constitutes noise, as an initial movement away from and subsequent reversal to the original value does not entail a fundamental price shock. Further, Wang and Xu (2019), in analysing the Chinese market, find that some C/R flows can significantly predict the total volatility of the underlying index, but not the fundamental volatility, which suggests, that these flows add noise to the market.

While this section provides indications that ETF trading volume is related to noise, this finding does not rule out that the Xact may also act as a price discovery channel. As such, we want to analyse the fundamental volatility of the market, as a relation to the fundamental volatility indicates a price discovery function (Xu et al, 2016; Wang and Xu, 2019). Below, we introduce the concept of the Beveridge Nelson (BN) fundamental volatility decomposition, before reporting the results of our analysis of the relationship between fundamental volatility and ETF trading activities.

6.3. Beveridge Nelson Decomposition

6.3.1. Methodology

The findings of Glosten et al. (2016) and Xu et al. (2016) show, that ETFs function as a price discovery tool for the underlying stocks. Hence, we use a method to identify the fundamental volatility in the underlying market. Explanatory power for trading volume or C/R flows would thus indicate that the processes are related to underlying securities adjusting to their fundamental values and thus the ETF functioning as a price discovery channel. The approach developed by Beveridge and Nelson (1981) offers a widely accepted solution to estimating the fundamental volatility (Morley 2011)[5].

The model is based on the idea, that if the first differences of a non-stationary time series are stationary, they can be described through a vector autoregressive (VAR) model, where the uncorrelated random disturbances present the innovations in the market. To be specific we can define the first differences (i.e. the returns) of the log prices p_t (i.e. log prices of the Index) as

[5] As we have a univariate setting, the disagreements of the literature in the calculation and consequent interpretation of the Beveridge Nelson decomposition can be disregarded. See Morley (2011) for more.

$r_t = p_t - p_{t-1}$ and portray it as a univariate AR model, as we are only interested in the random-walk variance (Wang and Xu, 2019). Hence, we get the following representation for r_t:

$$r_t = \sum_{k=1}^{\Delta} \lambda_k r_{t-k} + \varepsilon_t \tag{21}$$

with Δ as the number of lags, ε_t as residuals with mean zero and variance σ^2, and λ_k as the estimators for r_{t-k}. According to Wang and Xu (2019), we can thus generate estimates for $\Phi(1) = 1 - \sum_{k=1}^{\Delta} \hat{\lambda}_k$ and residuals $\hat{\varepsilon}_t$. Then, $\Phi(1)^{-1}\hat{\varepsilon}_t$ represents the permanent component of the price shock or innovation $\hat{\varepsilon}_t$ and its variance $\Phi(1)^{-2}Var(\hat{\varepsilon}_t)$ can be interpreted as the fundamental volatility of the market[6].

Applying this concept to the Swedish market, we first want to prove, that the first differences in the prices are stationary. That is, we take the log returns of the OMXS30 total return index and run an Augmented Dickey-Fuller test. The result[7] lets us dismiss the null-hypothesis that the return series has a unit root and thus it implies stationarity in the log returns. As such, we calculate the fundamental variance of the index with the BN decomposition, using 30 lags[8], in line with Wang and Xu (2019). We estimate the AR-model coefficients over the whole sampling period and calculate the BN fundamental variance on a daily basis. To be congruent with the previous regression, we take the square root of the fundamental variance to obtain the fundamental volatility of the index. We then regress the daily fundamental volatility on the daily flows and daily trading volume of the Xact OMXS30 ETF to investigate their relation towards the fundamental volatility of the OMXS30 total return index. Summary statistics for the fundamental volatility are shown in *Table 12*.

6.3.2. Results

We report the results of our analysis of the relationship between fundamental volatility and ETF primary and secondary market trading activities in *Table 5*.

Given the lack of significance when analysing the relationship of AP C/R flows with both the total volatility, as well as noise in the market, it is unsurprising that we obtain no significant link between fundamental volatility and AP flows. However, again, the observations

[6] Note that the fundamental volatility may be higher than the actual volatility of the market if there is positive autocorrelation. However, this does not influence the results of our analysis, since we are only interested in the level of the fundamental volatility and refrain from making interpretations from an economical perspective.
[7] From R: Dickey-Fuller = -95.103, Lag order = 95, p-value = 0.01
[8] We also calculate the fundamental volatility based on 10, 20, and 40 lags. See *6.5 Robustness Tests*.

and the implied lack of a relationship contradicts the findings in previous literature. Wang and Xu (2019) find that there is a two-way Granger causality between ETF C/R flows and fundamental volatility. Given the two-way relationship, the lack of observation of the relationship in our sample seems all the more peculiar. We note again the differences in our samples and assume that the lack of relationship between ETF C/R flows and fundamental volatility comes both from the limited degree of AP C/R flows, as well as the relative infancy of the Swedish ETF market.

Table 5. Regression Table: Index Level Fundament Volatility

The table reports estimates from the time series regression of the change in the fundamental volatility of the OMXS30 index on ETF creation/redemption flows, the change in ETF trading volume, and controls. The frequency of the observations is daily, and the fundamental volatility is obtained through Beveridge Nelson decomposition, calculated using 30 lags. In column (1) we control for lagged past 12-month return, and to account for potential reverse causality, column (2) contains three lagged dependent variables as explanatory variables. Due to the usage of a single ETF, standard errors are unclustered. The dependent variable and its lags have been scaled by a factor of 10^3, whereas the change in trading volume is expressed in millions (scaled by 10^-6). The sample covers the period March 2012 until March 2019. The legend for the statistical significance of observations is explained at the bottom of the table.

	Dependent variable:	
	ΔBN fund. volatility	
	(1)	(2)
Flow (t)	0.004	0.001
	(0.005)	(0.004)
ΔTrading volume (t)	0.008^{***}	0.006^{***}
	(0.002)	(0.002)
ΔPast 12M return (t-1)	0.099	-0.481^{***}
	(0.158)	(0.136)
ΔBN fund. volatility (t-1)		-0.624^{***}
		(0.024)
ΔBN fund. volatility (t-2)		-0.363^{***}
		(0.026)
ΔBN fund. volatility (t-3)		-0.152^{***}
		(0.024)
Constant	-0.001	-0.000
	(0.003)	(0.002)
Observations	1,746	1,746
R^2	0.007	0.287
Adjusted R^2	0.005	0.285
Residual Std. Error	0.115 (df = 1742)	0.098 (df = 1739)
F Statistic	3.879^{***} (df = 3; 1742)	116.662^{***} (df = 6; 1739)
Note:		*p<0.1; **p<0.05; ***p<0.01

As such, the relative and total size of an ETF and its resulting relevance for the market environment seems to be important for its impact on the market environment.

Based on the results, it seems that the ETF's trading volume in the secondary market is related to both the fundamental and noise component of the volatility. We find a highly statistically significant, positive relationship between the change in BN fundamental volatility and the change in ETF trading volume. This result persists when controlling for lagged dependent variables. Then, an increase in the change in trading volume (in thousands of shares) is associated with an increase of 0.6% increase in the change in fundamental volatility[9], representing 4.7% of a standard deviation of BN fundamental volatility. As with VR, the economic interpretation of an increase in the BN fundamental volatility is unclear, and again, secondary in importance to the identification of the relationship. Our findings are in line with those of Wermers and Xue (2015), who find predictive power of the SPY ETF on the development of the S&P 500 Index. Moreover, they find that the predictive power of the ETF can be traced back to informed trades happening in the SPY, suggesting that ETFs are traded to incorporate new fundamental information into the market, and are thus related to the fundamental volatility.

This identified link between ETF trading volume and fundamental volatility, which is supported by previous literature on the predictive power of ETFs on their underlying indices, gives us an indication, that ETFs may serve as a price discovery channel in the market. Hence, this section has provided additional evidence against hypothesis I.b and for hypotheses II.b.

To further substantiate our evidence for the *price discovery secondary market trading hypothesis (II.b)*, we conclude our analysis in the next section by analysing the relative Information Shares of the Xact OMXS30 ETF and the OMXS30 index. This will allow us to determine whether the Xact, and by extension other ETFs, indeed constitute price discovery channels in the Swedish market.

[9] Note: In *Table 5*, the BN fundamental volatility measure has been scaled by a factor of 10^3, whereas trading volume is expressed in millions. Please refer to footnote 2 for calculation methodology.

6.4. Information Share

6.4.1. Methodology

In the last section, we were able to find evidence for the *price discovery secondary market trading hypothesis*, but our methodology did not allow us to draw meaningful economically quantifiable conclusions from our results. As such, in a next step we want to quantify the price discovery happening in the Xact OMXS30 ETF compared to the OMXS30 cash index. For this, we utilize the Information Share (IS) after Hasbrouck (1995), which offers a widely accepted solution to this issue (Yan and Zivot, 2007).

Hasbrouck (1995) demonstrates, that the (log) price p_t of a security consists of an efficient price m_t– its fundamental value – and a temporary pricing error u_t. Even though the efficient price is not directly observable in the market, it is shared with other financial products representing the same underlying, i.e. the OMXS 30 index with an ETF tracking the index. As these prices cannot diverge without bound in the long run, we can say they are cointegrated (Engle and Granger 1987). If the efficient price then follows a random walk, it is integrated of order 1 and can be represented by

$$m_t = m_{t-1} + u_t \tag{22}$$

where u_t can be interpreted as the new information about the efficient price. Hence, its variance represents the variance in efficient price innovation (Xu et al, 2016). As the prices p_t are of order 1, the first-order difference Δp_t has an order of 0 and can be represented by a Vector Moving Average (VMA) representation

$$\Delta p_t = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \cdots \tag{23}$$

Looking at our case, $p_t$is a 2x1 vector in the form of $p_t = (p_t^{OMXS30}, p_t^{Xact})'$ and ψ_i represents a 2x2 MA coefficient matrix in lag i. Additionally, ε_t is a 2x1 vector that represents the innovations revealed in the respective market $\varepsilon_t = (\varepsilon_t^{OMXS30}, \varepsilon_t^{Xact})'$ with $E(\varepsilon_t) = 0$ and $Var(\varepsilon_t) = \Omega$. Next, we can sum the MA coefficient matrices to obtain

$$\psi(1) = I + \psi_1 + \psi_2 + \cdots \tag{24}$$

where I is a 2x2 identity matrix. Thus, $\psi(1)$ contains the information about the permanent impact of the innovations ε_t in the long run price dynamics and the efficient price. As Hasbrouck (1995) proves, that $\psi(1)$ has two identical rows, we can take either row and denominate $\Psi = (\Psi^{OMXS30}, \Psi^{Xact})$. According to Hasbrouck, $\Psi\varepsilon_t$ is thus the permanent component of the price change and its variance is $\Psi\Omega\Psi'$. Hence, one can calculate the contribution of innovations in the security j to the common (fundamental) price with:

$$IS_j = \frac{\psi_j \Omega_{jj}}{\Psi\Omega\Psi'} \tag{25}$$

To calculate this, we transform the VMA to a bivariate vector error correction model (VECM), processing intraday minute-to-minute price data of the index and the Xact OMXS30 ETF. From the VECM we calculate the long-term impacts of innovations Ψ^{OMXS30} and $\Psi^{Xact}$and calculate the covariance-variance matrix. Furthermore, to ensure the validity of the method, we confirm in a first step that the OMXS30 total return index and Xact OMXS30 ETF price are cointegrated of order 1 with the Johansen test for maximum Eigenvalue[10].

Deciding on the sampling frequency, Hasbrouck (1995) notes, that it is advisable to sample at high frequencies to reduce the contemporaneous correlation in the residuals (as the correlation between market innovations needs to be negligible to have Ω diagonal). Yan and Zivot (2007) find, that the necessary sampling frequency appear to be context specific: while some studies require a 10-second frequency, other studies obtain the same valid results using a 5-minute frequency. On the other hand, they argue, that sampling too frequently may distort the data with transitory microstructure noise. As such, we follow studies conducted in the context of ETFs that use the concept of ISs. Xu et al. (2016), Xu and Yin (2017a), and Wang and Xu (2019) all use 1-minute frequency data with 30 lags to calculate the IS and find reasonable results.

6.4.2. Results

Calculating the IS of the Xact OMXS30 ETF gives quite interesting results compared to the IS found in the U.S. market:

Table 6. Information Shares for the Nasdaq OMXS30 index

	Information Shares in the Swedish OMXS30	
Time period	2012-2019	2018
$IS_{\text{OMXS30 index}}$ -mean	99.998%	88.795%
$IS_{\text{Xact ETF}}$ -lower	0.000%	0.061%
$IS_{\text{Xact ETF}}$ -upper	0.004%	22.350%
$IS_{\text{Xact ETF}}$ -mean	0.002%	11.205%

[10] For r =0: 10 percent confidence value = 13.75, whereas our test statistic gives us 4.698. For r <=1 1 percent confidence value = 12.97, whereas out test statistic is 18004.78. Thus, we can confirm cointegration of order 1 with high statistical confidence.

Table 7. Information Shares for the Standard & Poor's 500 index.

	Information Shares in the American S&P 500		
Time period	1996-2000	2000-2010	2010-2014
$IS_{S\&P\ 500}$ -mean	51.40%	34.53%	23.01%
IS_{SPY} -mean	48.60%	56.69%	70.77%
IS_{IVV} -mean		8.78%	4.08%
IS_{VOO} -mean			2.14%

Source: Xu and Yin (2017b)

As of these results, the actual price discovery for the common efficient price is negligible for the total period of 2012–2019 (see *Table 6*). For comparison, the SPY has an IS of almost 71% in the S&P500 (see *Table 7*), which can be practically interpreted as the SPY being the price leader on almost 71% of the trading days over the period 2010-2014. Given our results in the previous regression, where we were able to link higher ETF trading activities to the fundamental volatility of the OMXS30 index, we would have expected a higher IS in the ETF. However, comparing some key figures of the Xact to the SPY (*Table 8*) gives some potential reasons for the indications as to why there is such a significant difference in ISs between the Swedish and American ETF market:

Table 8. Comparison of Xact's and SPY's key characteristics

ETF Name	**Xact OMXS 30 ETF**	**SPDR S&P 500 ETF**
Assets under Management (AUM)	~$1.1B	~ $271B
AUM to underlying market cap	~0.15%	~1.15%
Daily trading volume in USD	~$21m	~ $18B
Trading volume compared to AUM	2%	7%
	Source: Thomson Eikon	Source: Etf.com

The SPY holds a significant portion of the underlying securities (~1.15%) whereas the Xact has a comparatively low AUM of only about 0.15% of the total OMXS30 market cap. Furthermore, the average trading volume compared to the total AUM is around 2% for the Xact (as calculated by the daily average for 01/2019-03/2019) whereas the daily trading volume compared to its AUM is around 7% (ETF.com, 2019). And finally, the SPY contains 500 underlying securities, whereas the Xact ETF contains 30 stocks, which self-evidently makes the SPY the financial product of choice when wanting to incorporate information about the macroeconomic circumstances into the securities, as it is more cumbersome (and expensive) to trade 500 single stocks rather than 30 single stocks. While the first two reasons are due to the relative infancy of the Swedish ETF market compared to the U.S. ETF market and may adjust in the future, the difference in "accessibility" advantage of buying the SPY over all its underlying securities presents a structural difference, that will most likely keep the SPY's IS at higher levels than the Xact's IS even in the future.

In addition to these quantitative facts, the results should be taken with caution, as we are not completely confident in the Xact ETF price data that is used in this analysis. Retracting the data from Bloomberg seemed to sometimes cluster data points, because the system did not extract the price data for each minute separately but timestamped the moment a price change occurred and only extracted those points in time. Given the sheer size of our dataset, we were unable to examine each datapoint for validity but rather sample checked the data. Supporting this argument for potentially erroneous data is the fact, that dividing our dataset into subsets, does indeed often give significantly higher results for the IS. For example, we are more confident in our dataset for the year 2018 where we calculate an Information Share of 11.2%[11]. Analysing the respective data, gives indications as to why this is the case. Whereas over the whole time period from 2012-2019 the quoted ETF price changes on average every 8.46 minutes, the price changes every 5.75 minutes in 2018, which shows that either the Xact is traded more frequently or that the data is more accurate, or both. Related to that, the introduction of MiFID II in 2018 requires all ETF trades to be reported. This should on the one hand lead to more accurate data and, on the other hand, increase the popularity of ETFs in general, as a more accurate liquidity of these instruments is revealed. As such, while the Xact's IS of 11.2% is still considerably lower than the IS of 70.8% of its American counterpart, which can be explained by the reasons mentioned above, the results give an indication of a trend in the data towards higher future relevance and related higher IS of the Xact.

Consequently, these results for the IS should not be interpreted as disproving the *price discovery secondary market hypothesis* but can act as a further guidance on potential future research.

6.5. Robustness Tests

In this thesis, we apply a series of robustness tests throughout our analysis. Given that we replicate the stock level analysis from a published paper and apply it in a different geographical setting, we include only robustness factors used in that study. We include time- and stock-specific fixed effects, as well as clustering the standard errors. Additionally, to control for reverse causality, we include lagged dependent variables in our analysis.

Moving onto the index level, we perform several tests. For the intraday volatility, VR and fundamental volatility, we use changes in the variables to detrend them. Additionally, due to

[11] We follow Baillie et al. (2002) and calculate the Information Share as the average of the lower and upper bounds of the Information Share.

suspected autocorrelation, we include lagged dependent variables to account for this. Further, in order to ensure robustness, we apply Newey-West error-terms, and find that the significance of our variables is unaffected (*Table 16*). In addition, we winsorize the intraday volatility, fundamental volatility, trading volume and mispricing at the 1% and 99% levels (see *Figure 6* ff.*)*. None of the winsorization processes affect our results (see *Table 17* and *Table 19*).

Given the requirement of MiFID II to disclose ETF trades, and expected better data quality, we perform the analysis of the intraday volatility for 2018 onwards, and find that our results persist (see *Table 18*).

In order to verify the results from our fundamental volatility analysis, we calculate additional measures of the fundamental volatility using 10, 20, and 40 lags. Regardless of the number of lags used to calculate the fundamental volatility, our findings are robust (see *Table 20)*.

7. Discussion and Limitations

7.1. Discussion

In this section, we will provide a discussion of our results and give an overview of how our analyses relate to each other. A visualization of these advancements through our four hypotheses is provided in *Figure 5*.

This thesis finds a significant relationship between ETF ownership and volatility, in that the trading of ETF shares in the secondary market increases the noise in the market and the ETF functions as a price discovery channel. We start out by finding this relationship between ETF ownership and volatility on the stock level in the Swedish market, which motivates a further analysis on an index level, i.e. the OMXS30. This shift is necessary in order to test our four different sub hypotheses that distinguish between the source (primary or secondary market trading) and the nature (noise or fundamental volatility) of the volatility increase. On the index level, this analysis is enabled by having a direct match between the ETF and its underlying security, i.e. the OMXS30 index which tracks the price of the underlying securities. Hence, the effect of our ETF related variables that capture the primary and secondary market trading effects, namely the AP creation/redemption flows and the ETF trading volume respectively, are directly attributable to the measured underlying security. Moreover, to decompose volatility into its noise and fundamental volatility components, requires access to high frequency data that we could obtain for the OMXS30.

On the index level, we regress the changes in volatility on the AP creation/redemption flows and the changes in ETF trading volume with additional control variables. The model estimates a significant positive coefficient for the trading volume but not for the flow factor. These findings provide evidence for the secondary market trading hypotheses and against the primary market trading hypotheses. Moreover, we identify mispricing of the ETF as a positively modifying variable for the effect that trading volume has on volatility, which supports our hypothesis that the increase in volatility stems from the arbitrage channels enabled by ETFs.

Calculating the noise in the OMXS30 through the Variance Ratio and the fundamental volatility of the OMXS30 through the Beveridge Nelson decomposition allows us to analyse the nature of the volatility increase. In line with our previous findings, we identify a correlation between secondary market trading activity and noise and fundamental volatility but find none for the primary market trading activity. Finally, estimating the price discovery function of the

Xact OMXS30 ETF with regards to the OMXS30 index through the Information Share leads to notable results for the year 2018, for which we have the most accurate data. This supplies evidence, that an ETF's relation to an increase in volatility is partially due to the fact, that it acts a price discovery channel.

Our findings for the stock level regression are in line with that of previous literature (Ben-David et al, 2018). On the index level we find contrasting evidence to previous literature in not being able to identify a significant relation between the AP creation/redemption flows and the volatility of the underlying index (Xu et al, 2016; Ben-David et al, 2018; Wang and Xu, 2019). Possible explanations are extensively covered in section *6.1.2 Results* and the following limitations. On the other hand, our results for the secondary market trading hypothesis are consistent with that of previous research conducted in the U.S. market (Krause et al, 2012; Wermers and Xue, 2015; Glosten et al, 2016; Xu and Yin, 2017b).

This thesis adds to the growing body of literature about the possible negative consequences of ETFs in the relatively unexplored European ETF market. We find further evidence for the previously, in the U.S. market, discovered relationship between ETFs and market volatility in the Swedish market, while being the first study to distinguish between primary and secondary market trading activity within one analysis (to the best of our knowledge). The findings suggest, that the special characteristics of ETFs, in that they are price transparent and tradable throughout the day, make them increase the noise in the market while also being useful as price discovery tools. Given that the ETF market is growing continuously at an impressive rate (EY, 2017), market participants and regulators should be wary of the impacts that ETFs have on the underlying securities when deciding on actions and potential new regulations.

7.2. Limitations

Although this thesis attempts to solidify its findings as much as possible through various robustness tests, there are some limitations to our study. We will elaborate on the most important ones in the following.

7.2.1. Stock Level Analysis

The analysis on the stock level did not allow us to identify the mechanisms through which an increase in ETF ownership seems to increase the volatility of the market. While the ETF ownership factor was assumed to act as a proxy for all the arbitrage activity, liquidity trading, and AP trading flows, it may also capture other ETF effects such as a volatility spillover effects from other markets into the Swedish market. As such, one must be careful in drawing any causal inferences from these results. Furthermore, this thesis failed to granularize other ownership into its subcomponents of active, index, and hedge funds, which may work in opposite directions according to Ben-David et al. (2018). Hence, the results for the other ownership factor are spurious and require more in-depth analysis. However, making such an analysis was not critical for our research question and requires extensive effort in categorizing the other funds. Generally, a future study may want to include a larger sample size into this analysis to ensure more statistically significant results. However, given the purposes of the stock level analysis for this thesis, in that it served as the motivation for the subsequent analysis that focuses on the OMXS30 and thus a similar set of companies, these limitations are not crucial to the validity of the overall implications for our testable hypotheses.

7.2.2. Index Level Analysis

Moving on to the index level there is an overarching limitation emerging from the fact that we analysed the impact of only one ETF, and thus one must be careful in extrapolating these findings onto other ETFs. However, as previous studies have shown, it is an index' largest ETF that mainly impacts the underlying securities (Xu and Yin, 2017a), which is why having a smaller second OMXS30 ETF in the market would probably not have changed our results significantly.

When looking at the results, one must be cautious with the interpretation of the relation between AP creation/redemption flows and volatility. First, in the Xact OMXS30 ETF there are few flows compared to the U.S. or Chinese markets, which effectively reduces the sample size for flows to around 300 (number of days with flows). However, this still constitutes a considerable sample size if a truly strong effect were present. Second, it is difficult to pinpoint the exact timing at which AP flows may affect the underlying securities, as the settlement of the ETF creation/redemption process may be as late as in t+3 after initiation. To resolve this, we experimented with the time relationship between flows to changes in volatility but were not able to identify significance for any lagged variable of flow. Third, additional analysis showed that AP creation flows are correlated with trading volume (i.e. APs responding to high ETF market demand), which might distort our results. However, overall correlation between flows and trading volume was close to 0 and including the explanatory variables individually in the regressions did not significantly change the results either.

Another issue is related to the measure of trading volume. As not all trades had to be reported in Europe before MiFID II, the trading volume in our data likely does not represent the true trading volume. However, we can generally assume, that on average the same percentage of actual trades were reported each day. As such, this problem should not impact the qualitative results of the regression. Moreover, running a separate regression for the year 2018 for robustness, substantiates the findings obtained for the entire period.

Regarding our measures for noise and fundamental volatility, it is important to note, that these are in general imperfect measures of the variable to be measured (for VR, see O'Hara and Ye, 2011; for BN, see Morley, 2011). Furthermore, the correlation between total volatility and fundamental volatility is self-evidently quite high, as they generally move in unison, which makes the total volatility results akin to the fundamental volatility results. However, given that these methodologies have been used by other researchers in the same context and our results for the noise and fundamental volatility are in line with the expectations our initial index level regression sets, the findings should still be valid. The possible limitations of the Information Share and its underlying data has been extensively covered in *6.4.2. Results*.

Generally, it is inherently difficult to draw any causal inferences from our results, especially regarding the direction of causality. Naturally, the contemporaneous relationship of trading volume and volatility is bidirectional, as ETFs do not only influence their underlying securities, but are naturally also influenced by their underlying securities. Nonetheless, this thesis identifies significant correlations between ETFs and the volatility of their underlying, which gives crucial implications that ETFs may indeed actively impact the underlying market dynamics.

8. Conclusion

The objective of this thesis was to examine whether ETF ownership impacts the volatility of its underlying securities in Sweden. For this, we have analysed the Swedish stock market both on the stock and index level for the period between 2012 and 2019. We find, that increased ETF ownership in a stock is related to an increase in the volatility of the stock. Then, shifting our analysis to the OMXS30 index enabled us to determine the origin and nature of the volatility increase. The results show, that primary market trading activity, i.e. the creation and redemption of ETF shares by Authorized Participants, does not increase the volatility in the market. On the other hand, the secondary market trading activity, as captured by the trading volume of ETFs, is related to higher noise and fundamental volatility in the market and modified by a higher mispricing of the ETF. Furthermore, our findings indicate, that the Xact OMXS30 ETF acts as a price discovery channel in the OMXS30.

Based on our findings, a logical extension of our study would be to analyse the motives behind the different trades in the secondary market (akin to Xu et al. 2016), to further pinpoint the exact mechanism that makes ETFs increase volatility. Furthermore, the contradicting results regarding the impact of ETF share creation and redemption flows in comparison to previous literature, prompts for an analysis of the effect of these flows in the context of other European markets. More generally, there is an absence of ETF literature that differentiates between the impacts of categorically different ETFs on their underlying securities. This categorization could for example be based on the geographical scope, industrial scope, the size of the ETF, or synthetic ETFs. Such findings may give crucial insights into the mechanisms through which ETFs impact the stock market.

Given the ongoing growth of the worldwide ETF market and its increasing importance and impact on the overall market, the area warrants the need for continuous research. This thesis gives indications of an increase in market volatility through ETF secondary market trading activity, highlights the potential differences of the effects that ETFs have in the European markets, and emphasizes the necessity for a more in-depth analysis of the causal relationship between ETFs and their underlying securities.

References

Agapova, A., & Volkov, N. (2018). ETFs and Price Volatility of Underlying Bonds. Journal of Financial Markets. Available at: https://www.aeaweb.org/conference/2019/preliminary/paper/ZyyD7R8t

Amihud, Y. (2002). ILLIQUIDITY AND STOCK RETURNS: Cross-Section and Time-Series Effects. *Journal of Financial Markets*, *5*(1), 31–56. https://doi.org/10.1016/S1386-4181(01)00024-6

Antoniewicz, R., & Heinrichs, J. (2014). Understanding Exchange-Traded Funds: How ETFs Work. *SSRN Electronic Journal.* Advance online publication. https://doi.org/10.2139/ssrn.2523540

Appel, I. R., Gormley, T.A., & Keim, D. B. (2016). Passive Investors, Not Passive Owners. *Journal of Financial Economics*, *121*(1), 111–141. https://doi.org/10.1016/j.jfineco.2016.03.003

Baillie, R. T., Geoffrey Booth, G., Tse, Y., & Zabotina, T. (2002). Price discovery and common factor models. *Journal of Financial Markets*, *5*(3), 309–321. https://doi.org/10.1016/S1386-4181(02)00027-7

Baltussen, G., van Bekkum, S., & Da, Z. (2018). Indexing and stock market serial dependence around the world. *Journal of Financial Economics.* Advance online publication. https://doi.org/10.1016/j.jfineco.2018.07.016

Basak, S., & Pavlova, A. (2013). Asset Prices and Institutional Investors. *The American Economic Review*, *103*(5), 1728–1758. https://doi.org/10.1257/aer.103.5.1728

Ben-David, I., Francesco, F., & Moussawi, R. (2015). Do ETFs Increase Volatility? *NBER Working Paper 20071*. https://doi.org/10.3386/w20071

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

Beveridge, S., & Nelson, C. R. (1981). A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the ‘business cycle’. *Journal of Monetary Economics*, *7*(2), 151–174. https://doi.org/10.1016/0304-3932(81)90040-4

Bhattacharya, A., & O’Hara, M. (2018). Can ETFs increase market fragility? Effect of information linkages in ETF markets. *Working Paper, Cornell University*. https://doi.org/10.2139/ssrn.2740699

BlackRock. (2018). *BlackRock Global ETP Landscape September 2018*. BlackRock Inc.

Bond, P., & García, D. (2018). The equilibrium consequences of indexing. *Work. Pap., Foster Sch. Bus., Univ. Wash., Seattle Google Scholar Article Location*. Available at: http://w4.stern.nyu.edu/finance/docs/pdfs/Seminars/1801/1801w-bond.pdf

Broman, M. S. (2016). Liquidity, style investing and excess comovement of exchange-traded fund returns. *Journal of Financial Markets*, *30*, 27–53. https://doi.org/10.1016/j.finmar.2016.05.002

Broman, M. S., & Shum, P. (2018). Relative Liquidity, Fund Flows and Short-Term Demand: Evidence from Exchange-Traded Funds. *Financial Review*, *53*(1), 87–115. https://doi.org/10.1111/fire.12159

Buckle, M., Chen, J., Guo, Q., & Tong, C. (2018). Do ETFs lead the price moves? Evidence from the major US markets. *International Review of Financial Analysis*, *58*, 91–103. https://doi.org/10.1016/j.irfa.2017.12.005

Cheng, M., & Madhavan, A. (2009). The dynamics of leveraged and inverse exchange-traded funds. *Journal of Investment Management (JOIM)*. (Fourth Quarter). Available at SSRN: https://ssrn.com/abstract=1539120

Clifford, C. P., Fulkerson, J. A., & Jordan, B. D. (2014). What Drives ETF Flows? *Financial Review*, *49*(3), 619–642. https://doi.org/10.1111/fire.12049.

Cong, L. W., & Xu, D. X. (2016). Rise of Factor Investing Asset prices, informational efficiency, and security design. *29th Australasian Finance and Banking Conference*. https://doi.org/10.2139/ssrn.2800590

Da, Z., & Shive, S. (2018). Exchange traded funds and asset return correlations. *European Financial Management*, *24*(1), 136–168. https://doi.org/10.1111/eufm.12137

Engle, R. F., & Granger, C. W. J. (1987). Co-Integration and Error Correction: Representation, Estimation, and Testing. *Econometrica: Journal of the Econometric Society*, *55*(2), 251–276. https://doi.org/10.2307/1913236

ETF.com. (2019). SPY – SPDR S&P500 ETF TRUST – Overview. Retrieved 2019 April 19 from https://www.etf.com/SPY#overview

Evans, R., & Wilson, C. (2018, September 13). How ETFs Became the Market. *Bloomberg*. Retrieved from: https://www.bloomberg.com/graphics/2018-growing-etf-market/?srnd=etfs

EY. (2017). *Reshaping around the investor - Global ETF Research 2017*. EYGM Limited

Fama, E. F., & French, K. R. (1998). *Value versus growth: The international evidence*. The journal of finance, 53(6), 1975-1999. https://doi.org/10.1111/0022-1082.00080

Glosten, L. R., Suresh, N., & Yuan, Z. ETF trading and informational efficiency of underlying securities. *Research Paper No. 16–71, Columbia Business School*. Available at: https://www.rhsmith.umd.edu/files/Documents/Departments/Finance/fall2015/glosten.pdf

Greenwood, R., & Thesmar, D. (2011). Stock price fragility. *Journal of Financial Economics, 102*(3), 471–490. https://doi.org/10.1016/j.jfineco.2011.06.003

Hamm, S. J. W. (2014). The Effect of ETFs on Stock Liquidity. *Working Paper, Ohio State University.* Advance online publication. https://doi.org/10.2139/ssrn.1687914

Hasbrouck, J. (1995). One Security, Many Markets: Determining the Contributions to Price Discovery. *The Journal of Finance*, *50*(4), 1175–1199. https://doi.org/10.1111/jofi.12623

Hill, J. M., Nadig, D., & Hougan, M. (2015). *A comprehensive guide to exchange-traded funds (ETFs)*. Charlottesville (Va.): CFA Institute Research Foundation.

Humphries, W. M. (2010). Leveraged ETFs: The Trojan Horse Has Passed the Margin-Rule Gates. *Seattle UL Rev*, *34*, 299–323. Available at: https://digitalcommons.law.seattleu.edu/cgi/viewcontent.cgi?article=1966&context=sulr

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://doi.org/10.1007/s11142-017-9400-8

Krause, T. A., Ehsani, S., & Lien, D. D. (2012). Exchange Traded Funds, Liquidity, and Market Volatility. *Applied Financial Economics*, *24*(24), 1617–1630. https://doi.org/10.2139/ssrn.2153903

Leippold, M., Su, L., & Ziegler, A. (2015). How Index futures and ETFs affect stock return correlations. *Working Paper*, *University of Zurich*. https://doi.org/10.2139/ssrn.2620955

Lo, A. W., & MacKinlay, C, A. (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

Lou, D. (2012). A Flow-Based Explanation for Return Predictability. *Review of Financial Studies*, *25*(12), 3457–3489. https://doi.org/10.1093/rfs/hhs103

Madhavan, A. (2012). Exchange-Traded Funds, Market Structure and the Flash Crash. *Financial Analysts Journal*, *68*(4), 20–35. https://doi.org/10.2139/ssrn.1932925

Malamud, S. (2015). A Dynamic Equilibrium Model of ETFs. *Working Paper, Swiss Finance.* Advance online publication. https://doi.org/10.2139/ssrn.2662433

Morley, J. C. (2011). The two interpretations of the Beveridge–Nelson decomposition. *Macroeconomic Dynamics*, *15*(3), 419-439. doi: 10.1017/S1365100510000118

O'Hara, M., & Ye, M. (2011). Is Market Fragmentation Harming Market Quality? *Journal of Financial Economics*, *100*(3), 459–474. https://doi.org/10.2139/ssrn.1356839

Pan, K., & Zeng, Y. (2019). ETF Arbitrage Under Liquidity Mismatch. *Fourth Annual Conference on Financial Market Regulation*. http://dx.doi.org/10.2139/ssrn.2895478

Petäjistö, A. (2017). Inefficiencies in the Pricing of Exchange-Traded Funds. *Financial Analysts Journal*, *73*(1), 24–54. https://doi.org/10.2469/faj.v73.n1.7

Skypala, P. (2018, September 10), EU rule changes deliver mixed results so far. *The Financial Times.* Retrieved from: https://www.ft.com/content/46641e26-93f2-11e8-95f8-8640db9060a7

Shum, P., Hejazi, W., Haryanto, E., & Rodier, A. (2016). Intraday Share Price Volatility and Leveraged ETF Rebalancing. *Review of Finance, 20*(6), 2379–2409. https://doi.org/10.1093/rof/rfv061

Staer, A. (2017). Fund Flows and Underlying Returns: The Case of ETFs. *International Journal of Business*, *22*(4), 275–304. http://dx.doi.org/10.2139/ssrn.2158468

Stratmann, T., & Welborn, J. W. (2012). Exchange-Traded Funds, Fails-to-Deliver, and Market Volatility. *Working Paper, George Mason University.* Advance online publication. https://doi.org/10.2139/ssrn.2183251

Ståhl, P. (2018, January 9), Nya ETF-rekord, och de stora blir allt större. *Avanza.* Retrieved from: https://www.avanza.se/placera/redaktionellt/2018/01/09/rekord-for-borshandlade-produkter.html

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

Trainor, W. J. (2010). Do Leveraged ETFs Increase Volatility. *Technology and Investment, 01*(03), 215–220. https://doi.org/10.4236/ti.2010.13026

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*, *5*, 31–55. https://doi.org/10.1111/acfi.12437

Wermers, R., & Xue, J. (2015). Intraday ETF trading and the volatility of the underlying. *Working Paper, University of Maryland*. Available at: https://www.lyxoretf.at/pdfDocuments/5THE-ROLE-OF-ETFs--IN-INTRADAY-PRICE-DISCOVERYWP-5549243049350179714.pdf

Winne, R. de, Gresse, C., & Platten, I. (2014). Liquidity and risk sharing benefits from opening an ETF market with liquidity providers: Evidence from the CAC 40 index. *International Review of Financial Analysis*, *34*, 31–43. https://doi.org/10.1016/j.irfa.2014.04.003

Wurgler, J. (2010). On the economic consequences of index investing. *No. W16376. National Bureau of Economic Research*. https://doi.org/10.3386/w16376

Xu, L., & Yin, X. (2017a). Does ETF trading affect the efficiency of the underlying index? *International Review of Financial Analysis*, *51*, 82–101. https://doi.org/10.1016/j.irfa.2017.02.009

Xu, L., & Yin, X. (2017b). Exchange Traded Funds and Stock Market Volatility. *International Review of Finance*, *17*(4), 525–560. https://doi.org/10.1111/irfi.12121

Xu, L., Yin, X., & Zhao, J. (2016). Differently Motivated ETF Trading Activities and the Volatility of the Underlying Index. *FIRN Research Paper No. 2805276*. https://doi.org/10.2139/ssrn.2805276

Xu, L., Yin, X., & Zhao, J. (2018). Are Authorized Participants of Exchange-Traded Funds Informed Traders? *Working Paper. Available at: http://ssrn.Com/abstract=3221852.*

Yan, B., & Zivot, E. W. (2007). A Structural Analysis of Price Discovery Measures. *SSRN Electronic Journal.* Advance online publication. https://doi.org/10.2139/ssrn.979364

Appendix

Appendix 1: Overview of hypothesis development

	Hypothesis Advancement Overview	Hypothesis I.a *Noise primary market trading*	Hypothesis I.b *Price discovery primary market trading*	Hypothesis II.a *Noise secondary market trading*	Hypothesis II.b *Price discovery secondary market trading*
Stock Level	5. Daily Volatility	(✓)	(✓)	(✓)	(✓)
Index Level	6.1 Intraday Volatility	(×)	(×)	(✓)	(✓)
	6.1.3 Mispricing	~	~	(✓)	(✓)
	6.2 Variance Ratio	×	~	✓	~
	6.3 Fundamental Vol.	~	×	~	✓
	6.4 Information Share	~	~	~	✓
	Summary	**✗**	**✗**	**✓**	**✓**

✓ = evidence for hypothesis (✓) = indications for hypothesis × = evidence against hypothesis (×) = indications against hypothesis ~ = not tested for

Figure 5. Advancement of hypothesis testing

Appendix 2: Summary of ETF ownership for 38 Swedish stocks

Table 9. Number of ETFs and relative ownership of companies, 2012 and 2018

	31/12/2011		30/09/2018		Change	
Company	**# of ETFs(1)**	**Relative ETF own.(2)**	**# of ETFs(1)**	**Relative ETF own.(2)**	**# of ETFs(1)**	**Relative ETF own.(2)**
AAK.ST	3	0.02%	29	1.48%	+26	+1.46%
ABB.ST	3	0.02%	5	0.02%	+2	-0.00%
ALFA.ST	36	0.93%	118	2.03%	+82	+1.11%
ASSA-B.ST	37	0.91%	140	2.15%	+103	+1.24%
ATCO-A.ST	44	0.44%	146	1.36%	+102	+0.92%
ATCO-B.ST	32	0.21%	117	0.67%	+85	+0.46%
SOBI.ST	0	0.00%	64	1.71%	+64	+1.71%
BOL.ST	36	1.14%	176	4.55%	+140	+3.41%
CAST.ST	8	0.25%	49	2.03%	+41	+1.78%
ELUX-B.ST	40	0.98%	140	2.19%	+100	+1.22%
EKTA-B.ST	6	0.15%	55	1.91%	+49	+1.77%
ERIC-B.ST	62	1.03%	136	2.24%	+74	+1.20%
FABG.ST	11	0.35%	38	2.00%	+27	+1.65%
BALD-B.ST	2	0.03%	51	1.39%	+49	+1.35%
GETI-B.ST	25	0.71%	35	1.55%	+10	+0.84%
HEXA-B.ST	25	0.48%	118	1.79%	+93	+1.31%
HM-B.ST	46	0.82%	152	1.98%	+106	+1.15%
HUFV-A.ST	13	0.12%	26	1.12%	+13	+1.01%
INDU-C.ST	23	0.18%	88	0.78%	+65	+0.60%
INVE-B.ST	36	0.59%	121	1.43%	+85	+0.83%
KINV-B.ST	30	0.20%	105	2.35%	+75	+2.15%
LATO-B.ST	1	0.22%	20	0.02%	+19	-0.19%
LUND-B.ST	7	0.03%	82	1.16%	+75	+1.13%
LUPE.ST	37	0.80%	116	1.31%	+79	+0.52%
NIBE-B.ST	4	0.26%	52	1.19%	+48	+0.93%
SAAB-B.ST	8	0.09%	40	0.70%	+32	+0.60%
SAND.ST	46	0.88%	137	2.18%	+91	+1.31%
SCA-B.ST	39	0.21%	93	3.47%	+54	+3.26%
SEB-A.ST	40	0.78%	156	2.36%	+116	+1.57%
SECU-B.ST	30	0.91%	122	2.40%	+92	+1.50%
SHB-A.ST	46	0.92%	139	1.98%	+93	+1.06%
SKA-B.ST	36	1.05%	132	2.25%	+96	+1.20%
SKF-B.ST	47	0.89%	128	2.04%	+81	+1.14%
SWED-A.ST	44	0.77%	150	2.61%	+106	+1.84%
SWMA.ST	36	1.07%	154	4.51%	+118	+3.45%
TELIA.ST	68	0.79%	155	2.42%	+87	+1.63%
TREL-B.ST	15	0.45%	49	1.41%	+34	+0.96%
VOLV-B.ST	50	0.72%	165	1.94%	+115	+1.22%
Average	28	0.54%	100	1.86%	+72	+1.32%
Median	34	0.53%	118	1.96%	+80	+1.21%
Min	0	0.00%	5	0.02%	+2	-0.19%
Max	68	1.14%	176	4.55%	+140	+3.45%

(1) The number of ETFs with holdings in the stock
(2) Shares owned cumulatively by ETFs relative to the total amount of shares outstanding

Appendix 3: Summary statistics for stock level variables

Table 10. Summary statistics for stock level variables
Note that Daily Volatility, ETF Ownership and Other Ownership have been standardized by subtracting the mean and dividing by the standard deviation.

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Daily Volatility	3,192	0.000	1.000	-2.619	-0.689	0.491	6.501
ETF Ownership	3,192	0.000	1.000	-2.080	-0.735	0.562	3.148
Other Ownership	3,192	0.000	1.000	-3.407	-0.676	0.725	3.634
Past 12M Returns	3,153	0.211	0.596	-0.875	0.008	0.325	17.698
Log(Market Cap)	3,154	10.817	0.418	9.626	10.517	11.235	11.710
Inverse Price	3,154	0.011	0.010	0.001	0.006	0.013	0.105
Book-to-Market	3,154	0.543	0.445	-0.089	0.260	0.728	4.146
Amihud (x10^6)	3,152	0.001	0.007	0.000	0.000	0.000	0.294

Table 11. Correlation of stock level variables

	Daily Volatility	ETF Ownership	Other Ownership	Past 12M Returns	Log(Market Cap)	Inverse Price	Book-to-Market	Amihud
Daily Volatility	1.000							
ETF Ownership	-0.018	1.000						
Other Ownership	-0.050	0.484	1.000					
Past 12M Returns	-0.005	-0.053	-0.029	1.000				
Log(Market Cap)	-0.034	0.186	0.111	-0.108	1.000			
Inverse Price	0.060	-0.307	-0.250	0.074	-0.230	1.000		
Book-to-Market	0.021	-0.109	-0.109	0.070	-0.096	0.459	1.000	
Amihud (x10^6)	0.055	0.035	-0.029	-0.001	-0.094	0.019	0.016	1.000

Appendix 4: Summary statistics for index level variables

Table 12. Summary statistics for index level analysis variables (absolutes)

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Intraday volatility (x10^3)	1,747	0.292	0.121	0.103	0.213	0.340	2.369
Flow	1,747	0.157	0.541	0.000	0.000	0.000	7.000
ETF trading volume (m)	1,747	0.866	0.888	0.021	0.303	1.120	8.698
Index P12M returns	1,747	0.109	0.123	-0.194	0.015	0.201	0.353
Variance Ratio	1,747	0.115	0.087	0.000	0.046	0.164	0.537
Mispricing	1,747	1.000	0.004	0.865	0.999	1.001	1.016
BN FV (30 lags) (x10^3)	1,747	0.311	0.129	0.111	0.227	0.363	2.525

Table 13. Summary statistics for index level analysis variables (changes in variables)

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Intraday volatility (x10^3)	1,747	0.000	0.108	-2.202	-0.040	0.037	2.112
Flow	1,747	0.157	0.541	0.000	0.000	0.000	7.000
ETF trading volume (m)	1,747	0.000	1.121	-7.228	-0.453	0.439	8.234
Index P12M returns	1,747	0.000	0.018	-0.081	-0.010	0.010	0.103
Variance Ratio	1,747	0.000	0.119	-0.445	-0.074	0.071	0.420
Mispricing	1,747	1.000	0.004	0.865	0.999	1.001	1.016
BN FV (30 lags) (10^3)	1,747	0.000	0.116	-2.346	-0.042	0.040	2.252

Table 14. Correlation of index level variables (absolutes)

	Intraday Volatility	Flow	ETF trading volume (m)	Index P12M returns	Variance Ratio	Mispricing	BN fund. vol (30 lags)
Intraday Volatility	1.000						
Flow	-0.002	1.000					
ETF trading volume (m)	0.180	0.068	1.000				
Index P12M returns	-0.401	-0.014	-0.039	1.000			
Variance Ratio	-0.031	-0.001	0.038	0.022	1.000		
Mispricing	-0.084	0.004	-0.124	-0.010	0.007	1.000	
BN fund. vol (30 lags)	1.000	-0.002	0.180	-0.402	-0.036	-0.084	1.000

Table 15. Correlation of index level variables (changes in variables)

	Intraday Volatility	Flow	ETF trading volume (m)	Index P12M returns	Variance Ratio	Mispricing	BN fund. vol (30 lags)
Intraday Volatility	1.000						
Flow	0.011	1.000					
ETF trading volume (m)	0.078	-0.093	1.000				
Index P12M returns	0.013	0.004	-0.035	1.000			
Variance Ratio	-0.008	-0.005	0.046	0.045	1.000		
Mispricing	0.004	-0.088	0.061	0.087	0.015	1.000	
BN fund. vol (30 lags)	1.000	0.011	0.078	0.012	-0.014	0.003	1.000

Appendix 5: Distribution plots for index level variables

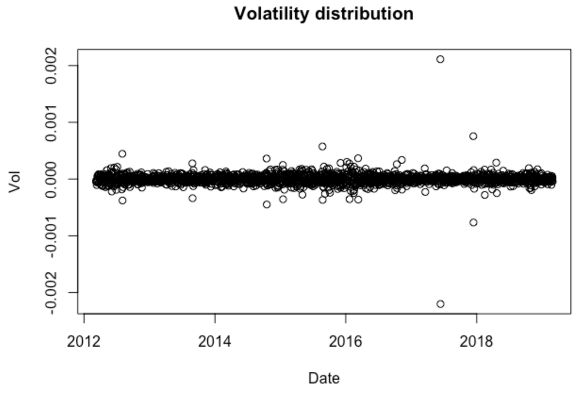

Figure 6. Intraday Volatility distribution

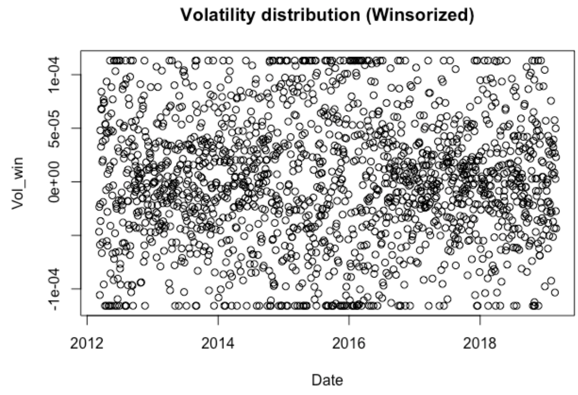

Figure 7. Intraday Volatility distribution winsorized

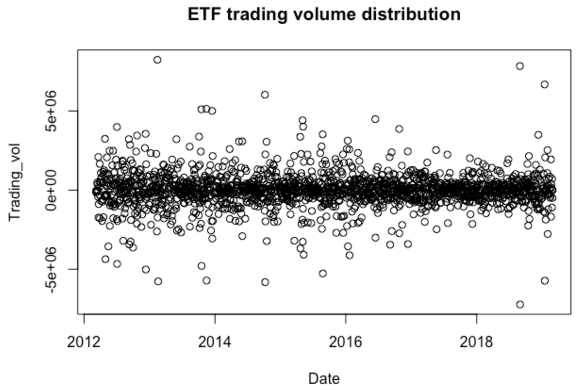

Figure 8. ETF trading volume distribution

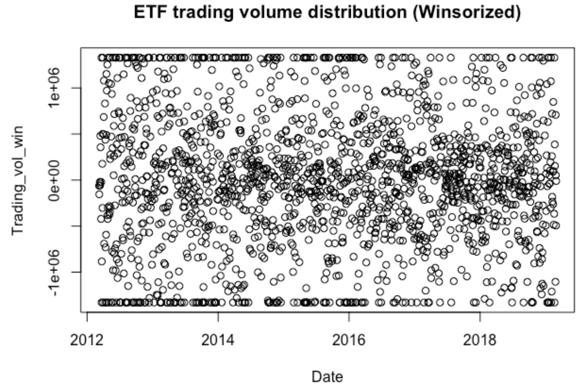

Figure 9. ETF trading volume distribution winsorized

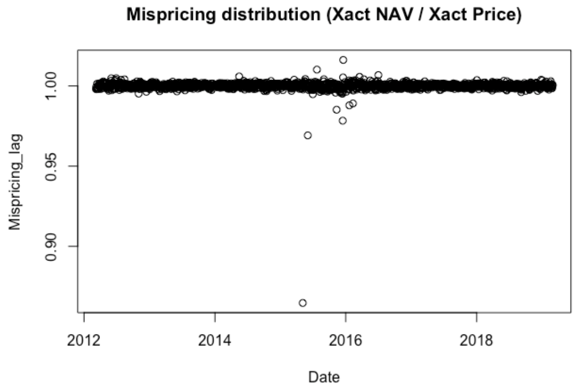

Figure 10. Mispricing distribution

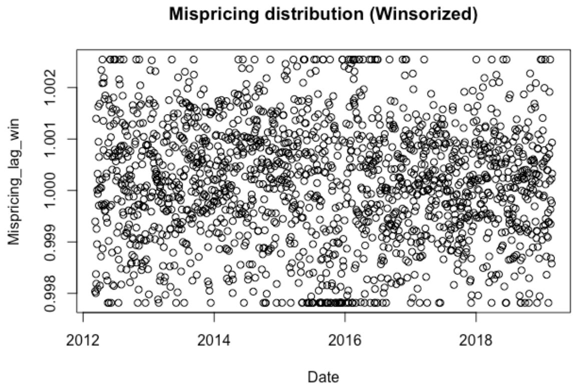

Figure 11. Mispricing distribution

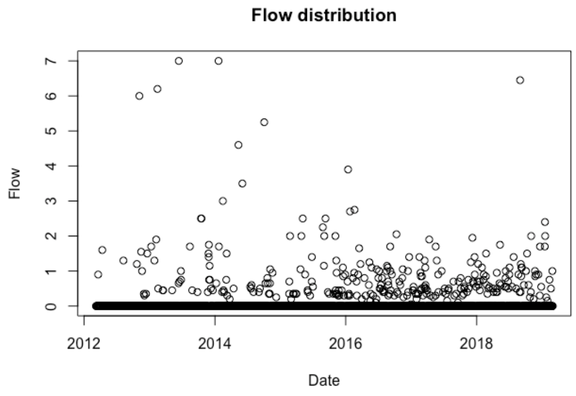

Figure 12. Flow distribution

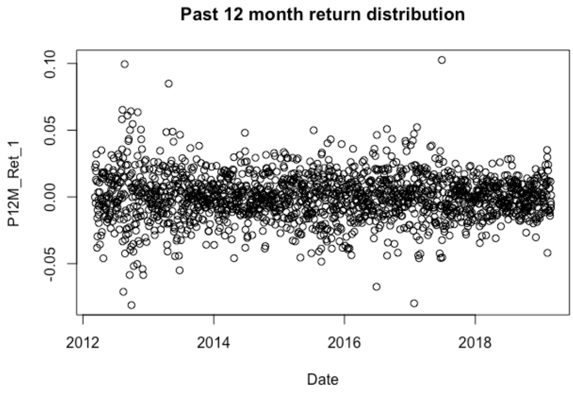

Figure 13. Past 12 month return distribution

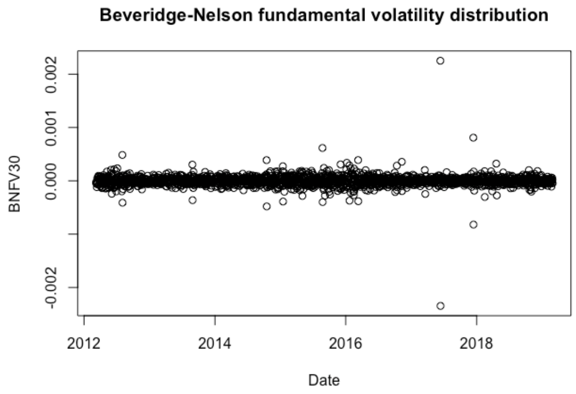

Figure 14. Fundamental volatility with 30 lags distribution

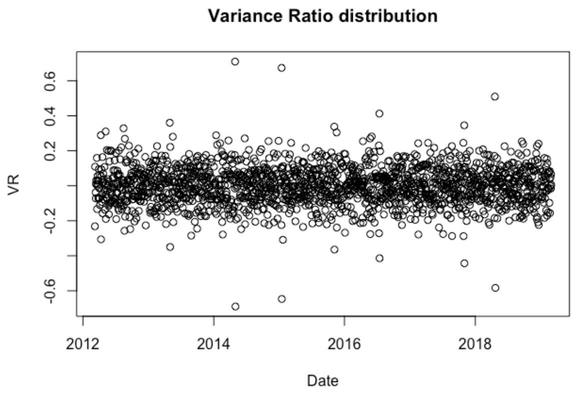

Figure 15. Variance Ratio distribution

Appendix 6: Intraday volatility robustness tests – Newey-West error terms

T*able 16. Robustness: Regression Table: Newey-West standard errors.*
The table shows the results presented in section 6 (see Table 2, Table 4, Table 5) for intraday volatility (IV), variance ratio (VR), and Beveridge Nelson fundamental volatility (BN), and their adjacent results when Newey-West standard errors have been computed.

	Dependent variable:					
	ΔIV	ΔIV(NW)	ΔVR	ΔVR(NW)	ΔBN	ΔBN(FV)
Flow (t)	0.001	0.001	0.002	0.002	0.001	0.001
	(0.004)	(0.002)	(0.004)	(0.004)	(0.004)	(0.003)
ΔTrading volume (t)	0.006***	0.006***	0.004*	0.004*	0.006***	0.006***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
ΔPast 12M return (t-1)	-0.452***	-0.452***	0.051	0.051	-0.481***	-0.481***
	(0.127)	(0.108)	(0.132)	(0.135)	(0.136)	(0.113)
ΔIV(t-1)	-0.625***	-0.625***				
	(0.024)	(0.074)				
ΔIV(t-2)	-0.362***	-0.362***				
	(0.026)	(0.077)				
ΔIV(t-3)	-0.151***	-0.151***				
	(0.024)	(0.051)				
ΔVR(t-1)			-0.702***	-0.702***		
			(0.023)	(0.024)		
ΔVR(t-1)			-0.462***	-0.462***		
			(0.026)	(0.023)		
ΔVR(t-1)			-0.247***	-0.247***		
			(0.023)	(0.021)		
ΔBN(t-1)					-0.624***	-0.624***
					(0.024)	(0.072)
ΔBN(t-2)					-0.363***	-0.363***
					(0.026)	(0.075)
ΔBN(t-3)					-0.152***	-0.152***
					(0.024)	(0.049)
Constant	-0.0004	-0.0004	-0.0004	-0.0004	-0.0004	-0.0004
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
Note:						*p<0.1; **p<0.05; ***p<0.01

Appendix 7: Intraday volatility robustness tests - winsorization

Table 17. Robustness: Regression Table: Winsorized intraday volatility

The table reports estimates from the time series regression of the change in intraday volatility of the OMXS30 index on ETF creation/redemption flows, the change in ETF trading volume, and controls. In column (1) we repeat the findings from the main text with all controls for reference. Column (2) shows the effect of winsorizing only Intraday volatility, whereas column (3) reports results when both intraday volatility and trading volume are winsorized (1st and 99th percentile). The intraday volatility variable and its lagged variables have been scaled by a factor of 10^3, whereas the change in trading volume is expressed in millions (scaled by 10^-6). The sample covers March 2012 until March 2019. The legend for the statistical significance of observations is explained at the bottom of the table.

	Dependent variable:		
	ΔIntraday volatility	ΔIntraday vol (win)	
	(1)	(2)	(3)
Flow (t)	0.001	0.002	0.002
	(0.004)	(0.002)	(0.002)
ΔTrading volume (t)	0.006^{***}	0.005^{***}	
	(0.002)	(0.001)	
ΔTrading volume (t,win)			0.008^{***}
			(0.002)
ΔPast 12M return (t-1)	-0.452^{***}	-0.262^{***}	-0.264^{***}
	(0.127)	(0.075)	(0.075)
ΔIntraday volatility (t-1)	-0.625^{***}		
	(0.024)		
ΔIntraday volatility (t-2)	-0.362^{***}		
	(0.026)		
ΔIntraday volatility (t-3)	-0.151^{***}		
	(0.024)		
ΔIntraday vol (t-1,win)		-0.481^{***}	-0.481^{***}
		(0.024)	(0.024)
ΔIntraday vol (t-2,win)		-0.269^{***}	-0.267^{***}
		(0.025)	(0.025)
ΔIntraday vol (t-3,win)		-0.085^{***}	-0.085^{***}
		(0.024)	(0.024)
Constant	-0.0004	-0.001	-0.001
	(0.002)	(0.001)	(0.001)
Observations	1,746	1,746	1,746
R^2	0.287	0.201	0.201
Adjusted R^2	0.285	0.199	0.198
Residual Std. Error (df = 1739)	0.092	0.053	0.053
F Statistic (df = 6; 1739)	116.762^{***}	73.100^{***}	72.885^{***}
Note:			*p<0.1; **p<0.05; ***p<0.01

Appendix 8: Intraday volatility robustness tests - period 2018 onwards

Table 18. Robustness: Regression Table: Index Intraday Volatility 2018

The table reports estimates from the time series regression of the change in intraday volatility of the OMXS30 index on ETF creation/redemption flows, the change in ETF trading volume, and controls. The frequency of the observations is daily, and volatility is calculated using the minute-by-minute log daily returns within the day. We control for lagged past 12-month return, and to account for potential reverse causality, we include three lagged dependent variables as explanatory variables. Due to the usage of a single ETF, standard errors are unclustered. The intraday volatility variable and its lagged variables have been scaled by a factor of 10^3, whereas the change in trading volume is expressed in millions (scaled by 10^-6). The sample covers January 2018 until March 2019. The legend for the statistical significance of observations is explained at the bottom of the table.

	Dependent variable:
	ΔIntraday volatility (2018)
Flow (t)	0.009
	(0.006)
ΔTrading volume (t)	0.009***
	(0.003)
ΔPast 12M return (t-1)	-1.021***
	(0.282)
ΔIntraday volatility (t-1)	-0.598***
	(0.059)
ΔIntraday volatility (t-2)	-0.331***
	(0.064)
ΔIntraday volatility (t-3)	-0.158***
	(0.057)
Constant	-0.002
	(0.003)
Observations	291
R^2	0.302
Adjusted R^2	0.287
Residual Std. Error	0.054 (df = 284)
F Statistic	20.437*** (df = 6; 284)
Note:	*p<0.1; **p<0.05; ***p<0.01

Appendix 9: BN fundamental volatility robustness tests - winsorization

Table 19. Robustness: Regression Table: Winsorized fundamental volatility

The table reports estimates from the time series regression of the change in fundamental volatility of the OMXS30 index on ETF creation/redemption flows, the change in ETF trading volume, and controls. In column (1) we repeat the findings from the main text with all controls for reference. Column (2) shows the effect of winsorizing only Beveridge Nelson fundamental volatility, whereas column (3) reports results when both fundamental volatility and trading volume are winsorized (1st and 99th percentile). The fundamental volatility variable and its lagged variables have been scaled by a factor of 10^3, whereas the change in trading volume is expressed in millions (scaled by 10^-6). The sample covers March 2012 until March 2019. The legend for the statistical significance of observations is explained at the bottom of the table.

	Dependent variable:		
	ΔBN fund. volatility		ΔBN fund. vol (win)
	(1)	(2)	(3)
Flow (t)	0.001	0.003	0.002
	(0.004)	(0.003)	(0.003)
ΔTrading volume (t)	0.006^{***}	0.006^{***}	
	(0.002)	(0.001)	
ΔTrading volume (t,win)			0.009^{***}
			(0.002)
ΔPast 12M return (t-1)	-0.481^{***}	-0.279^{***}	-0.281^{***}
	(0.136)	(0.080)	(0.080)
ΔBN fund. vol (t-1)	-0.624^{***}		
	(0.024)		
ΔBN fund. vol (t-2)	-0.363^{***}		
	(0.026)		
ΔBN fund. vol (t-3)	-0.152^{***}		
	(0.024)		
ΔBN fund. vol (t-1,win)		-0.481^{***}	-0.481^{***}
		(0.024)	(0.024)
ΔBN fund. vol (t-2,win)		-0.270^{***}	-0.269^{***}
		(0.025)	(0.025)
ΔBN fund. vol (t-3,win)		-0.086^{***}	-0.086^{***}
		(0.024)	(0.024)
Constant	-0.0004	-0.001	-0.001
	(0.002)	(0.001)	(0.001)
Observations	1,746	1,746	1,746
R^2	0.287	0.201	0.200
Adjusted R^2	0.285	0.198	0.198
Residual Std. Error (df = 17	0.098	0.057	0.057
F Statistic (df = 6; 1739)	116.662^{***}	72.943^{***}	72.666^{***}
Note:			$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Appendix 10: BN fundamental volatility robustness tests - different lags

Table 20. Robustness: Regression Table: BN calculated using 10, 20, and 40 lags

The table reports estimates from the time series regression of the change in the fundamental volatility of the OMXS30 index on ETF creation/redemption flows, the change in ETF trading volume, and controls (see Table 5). The Beveridge Nelson fundamental volatility is calculated with 10, 20, and 40 lags, reported in columns (1)-(3) respectively. The legend for the statistical significance of observations is explained at the bottom of the table. The dependent variable and its lags have been scaled by a factor of 10^3. The sample covers March 2012 until March 2019.

	Dependent variable:		
	ΔBNFV w. 10 lags	ΔBNFV w. 20 lags	ΔBNFV w. 40 lags
	(1)	(2)	(3)
Flow	0.001	0.001	0.001
	(0.004)	(0.004)	(0.004)
ΔTrading volume (t)	0.006^{***}	0.006^{***}	0.006^{***}
	(0.002)	(0.002)	(0.002)
ΔPast 12M return (t-1)	-0.468^{***}	-0.478^{***}	-0.480^{***}
	(0.132)	(0.135)	(0.136)
ΔBNFV w. 10 lags (t-1)	-0.624^{***}		
	(0.024)		
ΔBNFV w. 10 lags (t-2)	-0.363^{***}		
	(0.026)		
ΔBNFV w. 10 lags (t-3)	-151.568^{***}		
	(23.541)		
ΔBNFV w. 20 lags (t-1)		-0.624^{***}	
		(0.024)	
ΔBNFV w. 20 lags (t-2)		-0.363^{***}	
		(0.026)	
ΔBNFV w. 20 lags (t-3)		-0.152^{***}	
		(0.024)	
ΔBNFV w. 40 lags (t-1)			-0.624^{***}
			(0.024)
ΔBNFV w. 40 lags (t-2)			-0.363^{***}
			(0.026)
ΔBNFV w. 40 lags (t-3)			-0.152^{***}
			(0.024)
Constant	-0.0004	-0.0004	-0.0004
	(0.002)	(0.002)	(0.002)
Observations	1,746	1,746	1,746
R^2	0.287	0.287	0.287
Adjusted R^2	0.285	0.285	0.284
Residual Std. Error (df = 1739)	0.095	0.097	0.098
F Statistic (df = 6; 1739)	116.653^{***}	116.647^{***}	116.641^{***}
Note:			*p<0.1; **p<0.05; ***p<0.01