BI Norwegian Business School - campus Oslo

GRA 19703

Master Thesis

Thesis Master of Science

Does ETF Ownership Increase S&P 500 Stock Volatility?

Navn: Anette Normann, Emilie Gravdal Plaum

Start: 15.01.2020 09.00

Finish: 01.09.2020 12.00

Does ETF Ownership Increase S&P 500 Stock Volatility?

Master Thesis

by

Anette Normann and Emilie Gravdal Plaum

MSc in Business with Major in Finance

Oslo, September 1, 2020

ABSTRACT

This thesis studies the relation between exchange-traded fund ownership and volatility for 396 equity ETFs in the United States, using a data set comprising 11 years, from January 2008 to December 2018. We study stocks listed on the S&P 500 Index, using OLS regressions to investigate whether ETF ownership increase the volatility of underlying stocks. We find that a one-standard-deviation increase in ETF ownership would lead to a shift in the volatility of the median stock in the sample to a place between the 60^{th} and 73^{rd} percentiles. We conclude that in the period from 2008 to 2018 ETF ownership increases S&P 500 Index stocks volatility.

Acknowledgements

We would like to thank our thesis supervisor, Patrick Konermann, for the support and guidance through this process. Further, we thank our families and friends for their unconditional support and patience. The process has allowed us to increase our knowledge in areas we previously had not explored, thereby giving us a substantial understanding of the ETF market.

Contents

List of Abbreviations **IV**

List of Figures **V**

List of Tables **VI**

List of Symbols **VII**

1 Introduction **1**

2 Literature Review **3**

2.1 Stock volatility 4
2.2 ETFs impact on volatility of underlying stocks 5

3 Empirical Methodology **7**

3.1 Hypothesis 7
3.2 The econometric model 8
3.3 Measures of volatility and ETF ownership 11
3.3.1 Daily volatility 11
3.3.2 ETF ownership 11
3.4 Testable hypothesis 12

4 Data **12**

4.1 ETF data sample 12
4.2 Survivorship bias 13
4.3 Daily volatility 13
4.3.1 Measuring daily volatility 13
4.3.2 Descriptive statistics of daily volatility 14
4.4 ETF ownership 16
4.4.1 Measuring ETF ownership 16
4.4.2 Descriptive statistics of ETF ownership 19
4.5 Control variables 21
4.6 Preparation before empirical analysis 21
4.6.1 Multicollinearity 21
4.6.2 Fixed effects 22
4.6.3 Autocorrelation and heteroskedasticity 22
4.6.4 Lags of dependent variable 23

5 Empirical Results and Analysis 24

5.1 Empirical findings 24
5.2 Magnitude estimation and discussion of OLS estimates 27
5.3 Robustness 29
5.3.1 Realized Volatility as measure of daily volatility 29
5.3.2 Control for financial crisis 30
5.4 Limitations 32
5.4.1 Reliable data 32
5.4.2 Index, Active and Hedge Fund 32

6 Conclusion and Recommendation 33

6.1 Conclusion 33
6.2 Recommendations 34

Bibliography 35

Appendix 40

A.1 Institutional Details 40
A.1.1 ETFs vs. Mutual Funds 40
A.1.2 Creation/Redemption Function 41
A.2 Variable Description 42
A.3 List of ETFs in the Sample 44
A.4 Summary Statistics – Before Adjustment for Gaps 47
A.5 Correlation Matrix - Before Adjustment for Gaps 48
A.6 ETF Summary Statistics by Year 49
A.7 Number of ETFs 49
A.8 Detailed description of data collection for control variables 50
A.9 Summary Statistics for Control Variables 52
A.10 Correlation Matrix – Full Sample 53
A.11 Hausman (1978) specification test for fixed vs. random effects 54
A.12 Wald test for time-fixed effects 54
A.13 OLS Regression with Realized Volatility 55
A.14 Summary Statistics for Realized Volatility 56
A.15 Correlation Matrix with Realized Volatility 57
A.16 Summary Statistics for subsample “in crisis” 58
A.17 Summary Statistics for subsample “after crisis” 58
A.18 Correlation Matrix for subsample “In crisis” 59

A.19 Correlation Matrix for subsample "after crisis" 60
A.20 OLS Regression – NBER (n.d.) Sample Split by Recession 61

List of Abbreviations

AP	Authorized Participants
AUM	Assets Under Management
BASPRD	Bid-Ask Spread
BTM	Book-to-Market ratio
CRSP	The Center for Research in Security Prices
DVOL	Daily Volatility
ETF	Exchange Traded Funds
ETFOWN	ETF Ownership
GP	Gross Profitability as in Novy-Marx (2013)
ILLIQ	Amihud (2002) illiquidity measure of price impact
IP	Inverse Price
LMCAP	Logged Market Capitalization
NAV	Net Asset Value
NBER	The National Bureau of Economic Research
OLS	Ordinary Least Squares
P12MRET	Past 12-Month Returns
RDVOL	Realized Daily Volatility
S&P 500	S&P 500 Index
SEC	Securities and Exchange Commission
SD	Standard Deviation
SPDR	Standard & Poor's Depository Receipts
U.S.	United States of America
WRDS	Wharton Research Data Services

List of Figures

Figure 1: Growth of ETF Market in the U.S. 3

Figure 2: Variation in Daily Volatility 16

Figure 3: Volatility Clustering.. 24

List of Tables

Table 1: Daily Volatility (%) Summary Statistics 15

Table 2: ETF Holdings by Database 17

Table 3: ETF Ownership Summary Statistics 20

Table 4: ETF Ownership Equation Descriptive Statistics 20

Table 5: OLS Regression - Full Sample 27

Table 6: Magnitude Estimation 29

List of Symbols

α Parameter for intercept

β Parameter for slope

ε Idiosyncratic error term

u_i Unobservable stock effects

v_t Unobservable month effects

1 Introduction

Exchange-traded funds (ETFs) have grown rapidly in financial markets since their introduction in the early 1990s and they play an increasingly important role in the world's financial markets. The Investment Company Institute (2020) reported that the ETF market in the United States (U.S.) had $4.4 trillion in total net assets at the end of 2019, hence it remains the largest market in the world, accounting for 70 percent of the $6.3 trillion in ETF net assets worldwide. We observe from Figure 1 that there has been a 110-fold increase in U.S. ETFs asset under management (AUM), from $40 billion to $4.4 trillion, in the 21st century.

ETFs are passive investment vehicles that for the most part[1] aim to track the performance of a specific index, similar to index mutual funds, however they differ in fundamental ways (Lettau & Madhavan, 2018). ETFs trade throughout the day at market prices, whereas mutual funds can be purchased or redeemed only at the end of the trading day at its net asset value. In addition, ETFs differ from mutual funds as they do not trade with capital markets directly. Over the past decades we have seen a shift in investment strategies from active to passive investing (Stambaugh, 2014). Researchers argue that one of the reasons might be that investors have realized that the market is more efficient than previously thought, meaning that low-cost passive investments produce comparable or even superior performance to after-fees active funds (Ben-David, Franzoni & Moussawi, 2017).

The growth of ETFs has captivated both regulators and researchers. Regulators have raised concerns whether these innovations pose a threat to the financial market stability, especially after the Flash Crash on May 6 in 2010, when ETFs comprised approximately 60 percent of the trades that were subsequently cancelled. In the aftermath, the U.S. Securities and Exchange Commission (SEC) acknowledged that ETFs might contribute to market volatility and announced that they were investigating the issue. Further, a paper written by Ramaswamy for the Bank of International Settlements in 2011, raised the concern that ETFs can lead to a build-

[1] Further explained in Appendix A.1.1.

up of systemic risks in the financial system. Thus, investigating the issue on whether ETF ownership of stocks contributes to financial market instability is important for regulators.

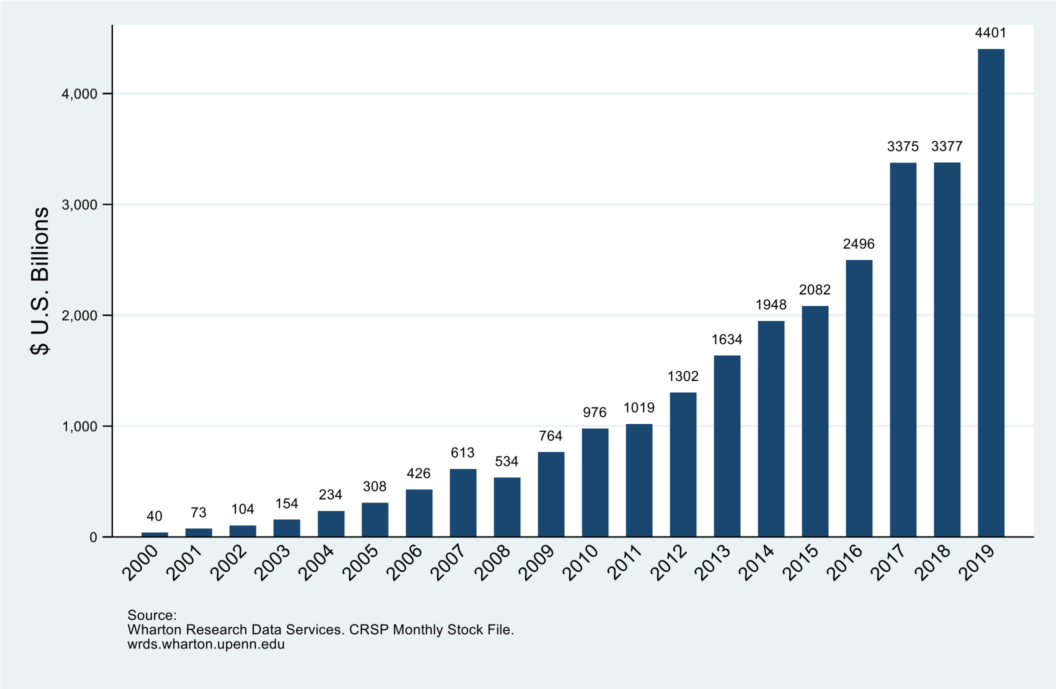
The total risk of financial markets is often measured by volatility (Brooks, 2019). As such, this thesis aims to contribute to the literature on ETFs impact on underlying securities, by investigating their influence on volatility. Our research question is as follows:

"Does ETF ownership increase the volatility of S&P 500 Index stocks?"

In a previous study conducted by Ben-David, Franzoni & Moussawi (2018a), they suggest that the liquidity of ETFs is likely to attract many investors because of their ease of trade. This demand can affect the prices of the underlying stocks through arbitrage, which may lead to higher volatility. For short-term investors, this increase in volatility could be attractive to the extent that it offers more trading opportunities. On the other hand, an increase in volatility would likely reduce the participation of long-term investors, who are often interested in the long-term prospects of firms. Thus, regulators with the goal of ensuring financial stability would worry about their reduced participation.

To address this research question, we investigate the U.S. ETF market, focusing on S&P 500 Index (S&P 500) stocks, in the period from 2008 to 2018. By fitting an OLS regression, we estimate the impact ETF ownership has on the volatility of underlying stocks. In the second chapter we provide a literature review of previous research relevant to the study. The third chapter defines the method we employ and the testable hypothesis, then we continue to present our data and its manipulations in chapter four. In the fifth chapter our results will be presented and discussed, together with an analysis of the robustness of our model. Last, we present our conclusion and recommendations for further studies of this subject.

Figure 1: Growth of ETF Market in the U.S.



The figure depicts the yearly (December 31st) AUM [2] in the U.S. in the time period from 2000 to 2019.

2 Literature Review

Most research has been conducted based on the U.S. ETF market, and focuses primarily on performance and efficiency as well as ETFs impact on other index markets. In many studies, researchers disagree about the effect ETFs have on the securities market (Ben-David et al., 2017). So far, these studies have highlighted the consequences ETFs have on *liquidity* (Hedge & McDermott, 2004; Hamm, 2014; Agarwal, Hanouna, Moussawi & Stahel, 2018;), *informational efficiency* (Israeli, Lee & Sridharan, 2017; Glosten, Nallareddy & Zoe, 2020) and *co-movements* (Israeli et al.,2017; Da & Shive, 2018). Additionally, some papers have emphasized the effect of ETFs on *the underlying stocks volatility* (Krause, Ehsani & Lien, 2014; Xu & Yin, 2017; Ben-David et al., 2018a). The following literature review discusses stock volatility and previous studies on ETFs impact on volatility of underlying stocks.

[2] AUM is the market capitalization of each ETF, computed as closing price times shares outstanding (Ben-David et al., 2018b).

2.1 Stock volatility

After the stock market crash in 1987, several researchers started to examine the relation between investment vehicles and stock market volatility (Harris, 1989; Edwards, 1988). Over the past decades, modelling and forecasting volatility has held the attention of academics (Poon & Granger, 2003). There exists a substantial amount of research on stock volatility, which reflects the importance of volatility in the stock market.

Volatility is considered as one of the most important concepts in finance and is often used as a crude measure of the total risk of financial markets (Brooks, 2019). It is also a key factor in pricing financial derivatives, such as options pricing.[3] Volatility in the market mainly reflects the deviation of the expected future value of an asset, and thus represents the uncertainty of an asset's future price. This uncertainty is commonly denoted by variance or standard deviation. More volatile securities are considered riskier, because the price of the security is expected to be less predictable. When stock prices display substantial volatility (i.e. the price of a stock change remarkably in either direction), over a short period of time, investors might be concerned about the future value of their investments (Edwards, 1988). This may cause investors to withdraw from their position in fear of losing their investments.

On the other hand, volatility can also attract investors, such as noise traders. Black (1985) states that investors with no access to inside information, irrationally act on noise as if it were information that would give them an edge. Consequently, this irrational trading creates an additional source of volatility that is priced in the marketplace (Brown, 1999). This volatility is often labeled *transitory volatility*, which is a source of volatility that regulators can substantially affect, depending on the policies adopted (Harris, 2003).

[3] Black-Scholes model is one of the most used mathematical models in pricing options, alongside Binomial pricing models, where volatility is a key input factor.

2.2 ETFs impact on volatility of underlying stocks

After the financial crisis in 2008, when volatility was historically high, some of the blame was placed on leveraged ETFs. Trainor (2010) investigates the link between leveraged ETFs and volatility of S&P 500 to find whether this holds true. By studying leveraged and inverse ETFs over a ten-year period, he finds no evidence that volatility of the S&P 500 has systematically increased due to the rebalancing issue[4] associated with leveraged ETFs. When using intraday volatility, the researcher saw the same spikes in volatility during periods not associated with rebalancing. Trainor (2010) states that despite the continued growth in levered ETFs, the abnormal market volatility has declined.

Malamud (2015) develops a dynamic equilibrium model for physical ETFs (i.e. not synthetic ETFs), where the ETF manager can create and redeem ETF shares through the authorized participants (AP)[5]. The interaction between the ETF manager and the AP serves as a shock propagation channel, where demand shocks are spreading into the underlying stocks. He shows that ETFs can affect the volatility of the underlying stocks through the arbitrage channel. Further, Malamud's (2015) model indicates that there is a positive relation between the shock propagation and the liquidity of the underlying stocks. He concludes that non-fundamental shocks propagate through arbitrage activity to the underlying stocks, which leads to higher stock volatility in the presence of ETFs.

Empirical evidence of spillovers from ETFs to the underlying stocks is documented by Krause et al. (2014). The researchers examine nine large sector ETFs and examine how the volatility information flows from these ETFs to their largest stocks. They estimate the volatility spillovers by applying the methodology of Diebold and Yilmaz (2012). According to their spillover estimations, they find that the largest stocks held by ETFs have higher volatility and higher volume. However, as Krause et al. (2014) state themselves, they cannot conclude from their findings

[4] According to Trainor (2010) the rebalancing issue of leveraged ETFs is that daily rebalancing is required to maintain a constant leverage ratio, which creates additional demand or selling pressure in the same direction as the market move.
[5] Details on the relationship between ETF managers and APs are found in Appendix A.1.2.

that ETFs increase the volatility of the underlying stocks. An identification strategy is required, to identify exogenous variation in ETF ownership.

Such an identification strategy is employed by Ben-David et al. (2018a). They study pure equity ETFs listed on U.S. exchanges in the period between 2000 and 2015, and investigate whether ETF ownership increase stock volatility. They propose that when a liquidity shock hits the ETF market, the price of the ETF will rise above the fundamental value. In this case, the arbitrageurs invest in the underlying stock and short ETF shares. In the long run, the ETF price and the underlying stock price will revert to their fundamental value. They provide evidence that the volatility increases when arbitrage most likely occur, which is when the ETF price diverge from the value of the underlying stocks.

First, Ben-David et al. (2018a) employ OLS regressions to show that a one standard-deviation change in ownership is associated with a 16.4% standard deviation increase in daily volatility for S&P 500 stocks. Second, the researchers use a two-stage least squares estimation, where they document in the first stage that stocks switching from the Russel 1000 to Russel 2000[6] increases ETF ownership by about 19.6% of a standard deviation. Using this identification strategy, the researchers conclude that this exogeneous increase in ETF ownership leads to a substantial increase in stock volatility. We complement their study by investigating a later time period.

Moreover, Xu and Yin (2017) investigate the relation between the volatility of a market index and the trading volumes of the ETFs that track the index, specifically the S&P 500. The results show an upward trend on the index's volatility, where the slopes are steeper after the introduction of ETFs. By employing generalized autoregressive conditional heteroskedasticity (GARCH) models and OLS regressions, they demonstrate that the trading volume is a critical determinant of the volatility of the S&P 500. Further, they find that a two-way Granger causality exists between the trading of ETFs and the volatility of the index. This two-way

[6] The Russell 1000 represents the first 1,000 top companies by market capitalization, while the Russell 2000 represents the following 2,000 largest stocks.

Granger causality between ETF trading and stock volatility exists for various market indices.

The above literature review has shown that there exists a limited number of studies investigating the relation between ETFs and underlying stocks' volatility. The studies conducted in the equity ETF area, has arrived at the same conclusion, that ETFs increase stock or index level volatility. Overall, the studies are new, i.e. conducted after 2010, and differ widely in methodology. We interpret this as a sign that there are shortcomings in this area of the literature. Our thesis aims to contribute to this literature by investigating the relation between equity ETF ownership and volatility of S&P 500 stocks.

3 Empirical Methodology

This chapter provides the methodological framework used to examine the impact ETF ownership has on stock volatility. First, we build our hypothesis based on the theory presented above. Second, we describe the methodology used to study the impact ETF ownership has on the volatility of the underlying stocks and define the employed measures of volatility and ETF ownership. Last, we present the econometric hypothesis to be tested.

3.1 Hypothesis

The aim of the thesis is to examine how ETFs impact the volatility of their underlying securities, focusing on ETFs holding S&P 500 stocks. The S&P 500 Index is a capitalization-weighted index constituting the 500 largest U.S. publicly traded companies. To conduct our research, we are motivated by the findings of Ben-David et al. (2018a), that ETF ownership increase the volatility of underlying stocks. We would like to investigate whether Ben-David et al.'s (2018a) model holds for our sample and test the hypothesis that stocks with higher ETF ownership exhibits increased volatility. Thus, we test whether ETF ownership of S&P 500 stocks contribute to increased volatility by investigating the following hypothesis:

$\mathbf{H_0}$: S&P 500 stocks with higher ETF ownership does not exhibit increased volatility.

$\mathbf{H_A}$: S&P 500 stocks with higher ETF ownership exhibit increased volatility.

3.2 The econometric model

To test the relation between ETF ownership and stock volatility, we conduct three different OLS regressions at a monthly frequency across S&P 500 stocks, motivated by Ben-David et al. (2018a). The model will consist of daily volatility as the dependent variable and ETF ownership, lagged control variables and fixed effects as the independent variables. Since we want to study the effect of ETF ownership in period *t* on the volatility in period *t+1*, we lag the control variables once. We estimate a two-way fixed effects model, using "within" transformation. An alternative model would be to use the least squares dummy variable model, but since we estimate more than 500 stocks across 11 years, degrees of freedom will suffer a great loss resulting in an inefficient model.

Since we have data comprising both cross-sectional elements and time series, our data is, by definition, panel data. The S&P 500 applies a floating index reconstitution. This implies that the inclusion of stocks in our sample can vary over time periods, causing our panel data to be rotating and unbalanced. This could potentially be a source of bias, however since the reconstitution is random (i.e. a firm cannot choose to leave or enter), this type of unbalanced panels is easily dealt with (Wooldridge, 2010) and STATA is able to make appropriate adjustments within the model.

In OLS regressions a common problem is the omitted variable bias, which is the bias that arises when the independent variable is correlated with an omitted variable and the omitted variable is a determinant of the dependent variable. This can cause on average too large or too small OLS estimates, depending on the direction of the correlation. One of the main motivations of employing panel data is to solve the omitted variable problem (Woolridge, 2010). To guard against potentially omitted variables in our model, we make the following three inclusions.

First, we include stock and month fixed effects as fixed effects models remove the omitted variable bias by measuring changes within the stocks across time. Stock fixed effects account for other cross-sectional differences between the stocks (Brooks, 2019). This means that time-invariant differences between the stocks, such as industry, are controlled for. Thus, the estimated coefficients are not biased due to the omission of such characteristics. Time fixed effects is used because the average value of the stock volatility changes over time.

Second, we include a set of control variables. The first control variable of interest is the logged market capitalization (LMCAP), which is the natural logarithm of the stock market capitalization. It is natural to include this as a control variable since the S&P 500 is a capitalization-weighted index, which means that it assigns a higher weight the higher the market capitalization. Further, to control for stock size and liquidity, we include the inverse of the stock price (IP), the Amihud (2002) illiquidity measure of price impact (ILLIQ) and the bid-ask spread (BASPRD). Additionally, we include the following three standard predictors of returns that can relate to volatility: the book-to-market ratio (BTM), gross profitability (Novy-Marx, 2013) (GP) and past 12-month returns (P12MRET).

Third, standard errors are double-clustered at the stock and month levels. This is because, when modelling panel data at the stock level one can expect correlations within stocks over time and across firms, however the patterns of variance and covariance are usually unknown.

The first regression employed to analyze the effect of past ETF ownership on the volatility of the stock is as follows:

$$\begin{aligned} DVOL_{i,t} = \alpha &+ \beta_1 ETFOWN_{i,t} + \beta_2 LMCAP_{i,t-1} + \beta_3 IP_{i,t-1} + \beta_4 ILLIQ_{i,t-1} \\ &+ \beta_5 BASPRD_{i,t-1} + \beta_6 BTM_{i,t-1} + \beta_7 GPROFIT_{i,t-1} \\ &+ \beta_8 P12MRET_{i,t-1} + u_i + v_t + \varepsilon_{i,t} \end{aligned}$$

Equation 1: Regression of daily volatility w/o lags

where $DVOL_{i,t}$ is the daily volatility of the stocks i in month t, $ETFOWN_{i,t}$ is the measure of ETF ownership, $LMCAP_{i,t-1}$, $IP_{i,t-1}$, $ILLIQ_{i,t-1}$, $BASPRD_{i,t-1}$, $BTM_{i,t-1}$, $GPROFIT_{i,t-1}$ and $P12MRET_{i,t-1}$ is the set of control variables, u_i and v_t are the unobservable stock and month effects respectively, and $\varepsilon_{i,t}$ are the idiosyncratic errors. In our regression u_i and v_t are treated as fixed effects, which means that they are allowed to be correlated arbitrarily with the observed independent variables. Daily volatility and the ETF ownership variable are standardized by subtracting the sample mean and dividing by the sample standard deviation, to ease interpretation. Details on the variables in our regression are provided in Appendix A.2.

Since we should be concerned that there might be persistence in the daily volatility, we should address this concern by including three lags of daily volatility (dependent variable). To employ this method, we first need to estimate a regression where we replicate the first regression using a subsample where three lags of the dependent variable are available. This is a way to check that the estimated slopes of the variables are not highly influenced by the change of observations. After controlling for this we can estimate the last regression, where we include up to three lags of the daily volatility to address the concern that the persistence in volatility might cause reverse causality. The regression equation is given by

$$\begin{aligned} DVOL_{i,t} = \alpha &+ \beta_1 ETFOWN_{i,t} + \beta_2 LMCAP_{i,t-1} + \beta_3 IP_{i,t-1} + \beta_4 ILLIQ_{i,t-1} \\ &+ \beta_5 BASPRD_{i,t-1} + \beta_6 BTM_{i,t-1} + \beta_7 GPROFIT_{i,t-1} \\ &+ \beta_8 P12MRET_{i,t-1} + \beta_9 DVOL_{i,t-1} + \beta_{10} DVOL_{i,t-2} \\ &+ \beta_{11} DVOL_{i,t-3} + u_i + v_t + \varepsilon_{i,t} \end{aligned}$$

Equation 2: Regression of daily volatility with lags

An issue with including lags of the dependent variable in our model is that the strict exogeneity assumption never holds in unobserved effects models with lagged dependent variables (Woolridge, 2010). Since we have restricted our analysis to S&P 500 stocks, we are not able to completely address the concern that ETF ownership may be endogenous, according to Ben-David et al. (2018a). The researchers conduct a quasi-natural experiment investigating the Russel indices. While the S&P 500 applies a floating index reconstitution, the Russel indices apply an annual reconstitution; thus, we cannot employ the same index-switching model for our sample. However, if our findings are consistent with Ben-David et al.'s (2018a), regarding the positive and statistically significant relation between ETF ownership on stock-level volatility, we choose to rely on their study to assume that there is an exogenous relation.

3.3 Measures of volatility and ETF ownership

3.3.1 Daily volatility

We follow Ben-David et al. (2018a) and employ standard deviation as a measure of daily volatility. Using the standard deviation as a measure of volatility has influenced the investment literature since the classic work of Markowitz (1959). When examining ETF ownership effects on stock volatility, this is the measure we employ. It is calculated as

$$DVOL_{i,t} = \sqrt{\frac{\sum_{d=1}^{n}(r_{i,d} - \bar{r}_{i,d})^2}{n-1}}$$

Equation 3: Daily volatility

where $DVOL_{i,t}$ is the daily volatility of each stock *i* at month *t*, $r_{i,d}$ is the intraday stock return, $\bar{r}_{i,d}$ is the average daily return, and n is the number of days in the month. There are however other measures of volatility, such as the measure introduced by Andersen, Bollerslev, Diebold and Labys (2001) and Barndorff-Nielsen and Shephard (2002), called realized volatility. We use realized volatility as a measure of daily volatility to control the robustness of our model.

3.3.2 ETF ownership

For accuracy we employ the measure of ETF ownership proposed by Ben-David et al. (2018a). They define ETF ownership as the fraction of a stock's capitalization that is held by ETFs. In other words, ETF ownership of stock *i* at time *t* is the sum of the dollar value of holdings by ETFs investing in a particular stock, divided by the stock's market capitalization at the end of the month. Therefore, the ETF ownership is computed as

$$ETF\ ownership_{i,t} = \frac{\sum_{j=1}^{J} w_{i,j,t} AUM_{j,t}}{MCAP_{i,t}}$$

Equation 4: ETF ownership

where J is the set of individual ETFs that hold stock I; $w_{i,j,t}$ is the weight of the stock in the portfolio of ETF j, which is extracted from the most recent quarterly report; and $AUM_{j,t}$ is the monthly market capitalization of ETF j, which equals the assets under management. The product $w_{i,j,t}AUM_{j,t}$ reflects the dollar ownership of ETF j in stock i in the current period. Finally, $MCAP_{i,t}$ is the stock's market capitalization at the end of the month calculated as shares outstanding times closing price.

3.4 Testable hypothesis

From the econometric models presented in 3.2, we infer that if coefficient β_1 is proven to be positive and statistically significant, it implies that ETF ownership increases the daily stock volatility. We can thus explicitly define the testable hypothesis as

$$\mathrm{H_0}: \beta_1 \leq 0 \qquad\qquad \mathrm{H_A}: \beta_1 > 0$$

4 Data

In the following chapter we give a detailed description of how the data used in this research is collected and how we construct the sample. In the first section, we explain how we choose the ETFs that will be the foundation of our sample. Second, we enlighten the potential survivorship bias in our sample. Third, we demonstrate how the data for the variables in our regression is retrieved, how they are measured and adjusted. Last, we explain the preparation before the empirical results.

4.1 ETF data sample

Using Bloomberg, we identify ETFs traded on U.S. exchanges. In our analysis, we focus on ETFs traded in the U.S. with a minimum AUM of $100 million on January 20th, 2020. Further, we restrict our sample to equity ETFs that engage in physical replication, which is ETFs that hold the underlying stocks physically. This means that we omit from our sample other exchange-traded products, such as exchange-traded notes and exchange-traded commodities. In addition, we exclude leveraged ETFs that uses financial derivatives and debt to generate the return of a certain index and we exclude currency hedged funds. Our final sample consists of 396 distinct equity ETFs, which are all still traded in the U.S. today (see full list of the ETFs in Appendix A.3). These ETFs are identified by their *ticker* from Bloomberg.

4.2 Survivorship bias

Since our sample consists solely of ETFs that still exists today, we might face the issue of survivorship bias. Survivorship bias is the tendency to view the existing stocks or funds in the market as a representative for a larger and more comprehensive sample. This occurs when non-surviving stocks and funds (incl. merged stocks and funds) are not included in the sample. Several researchers have argued the importance of survivorship bias and arrived at different conclusions. Brown, Goetzmann, Ibbotson and Ross (1992) argue that it is necessary to include both the existing and non-surviving funds to prevent an overestimation of a fund's performance. An overestimation of performance might occur as funds tend to close because of their poor performance or sufficiently low total market value (Elton, Gruber & Blake, 1996). Contrary, Wermers (1997) claim that survivorship bias is a relatively small problem, as he finds a minor difference in returns between the non-surviving funds and the surviving funds.

4.3 Daily volatility

4.3.1 Measuring daily volatility

Considering that we would like to establish the effect ETF ownership has on the volatility of underlying S&P 500 securities, the dependent variable is the *daily volatility* of those securities. To compute the daily stock volatility at the monthly frequency we measure the standard deviation of intraday returns over each month for each security. Before calculating the daily volatility, we first need to calculate the intraday return for each security at time *t*. The intraday return $r_{i,d}$ is calculated as

$$r_{i,d} = log\left(\frac{P_d}{P_{d-1}}\right)$$

Equation 5: Intraday returns

where P_d is the closing price at day *d*, and P_{d-1} is the open price at the same day *d*. Thus, the intraday return is calculated by taking the logarithm of the difference between the closing price and the open price at day *d*. The daily open price and daily closing price are downloaded from the Daily Stock File from Center for Research in Security Prices (CRSP) through Wharton Research Data Services (WRDS). Using intraday returns, we calculate the variance ($VAR_{i,t}$) by subtracting the mean from each intraday return and square the result. Then, we summarize the squared deviations and divide it by one less than the number of trading days in the corresponding month. The variance is thus calculated as

$$VAR_{i,t} = \frac{\sum_{d=1}^{n}(r_{i,d} - \bar{r}_{i,d})^2}{n-1}$$

Equation 6: Variance

where $r_{i,d}$ is the daily return, $\bar{r}_{i,d}$ is the average daily return, and n is the number of trading days in each month *t*. Finally, we calculate the volatility as the squared root of the variance as

$$DVOL_{i,t} = \sqrt{VAR_{i,t}}$$

Equation 7: Daily volatility

where $DVOL_{i,t}$ is the volatility for each stock *i* at month *t*.

4.3.2 Descriptive statistics of daily volatility

Table 1 reports the summary statistics for the monthly sample of the daily stock volatility (dependent variable) for each year from 2008 to 2018. Firstly, the statistics show that our sample comprises 59,320 observations, across the sample period. This confirms that we have a fairly large sample. We observe a difference in the average volatility across the years, where the average is 3.2% in 2008 and 1.5% in 2018. From 2008 to 2018, the average is 1.6%, with a standard deviation of 1.2%, while the median is 1.3%. This is a difference of approximately 0.3% between the mean and median, which might imply that we have some outliers in our sample driving the mean upward. We further notice that the years 2008 and

2009 display the highest volatility, with a maximum of 29.3% and 19.7%, respectively. Thus, these years remain the periods with the highest volatility in the sample, which coincide with the financial instability due to the financial crisis.

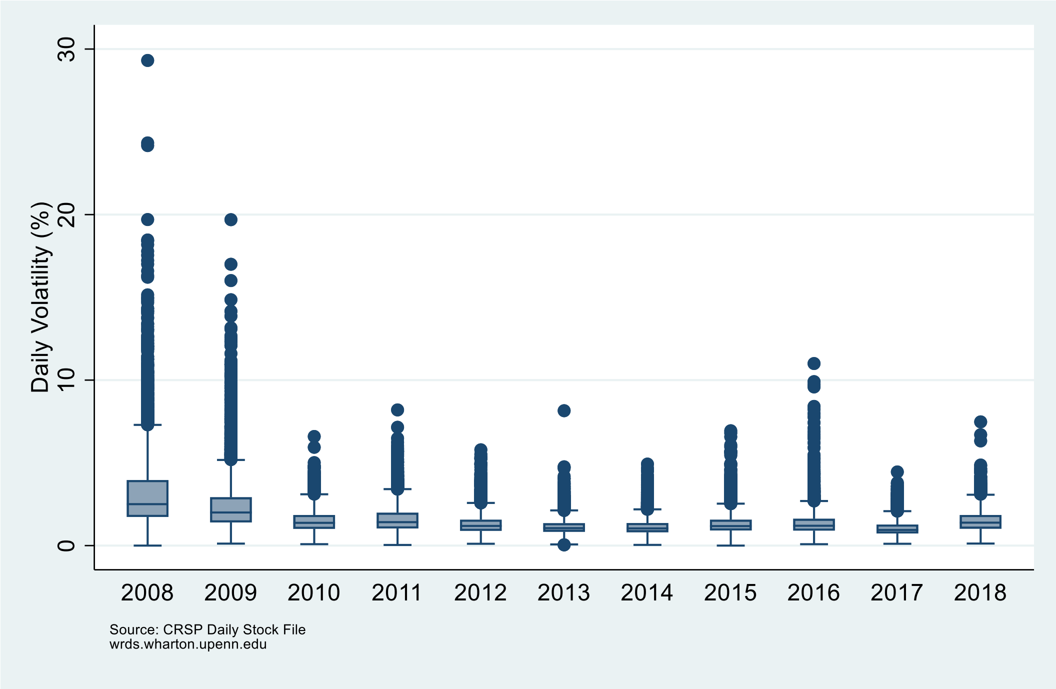
Figure 2 presents the yearly distribution and skewness of daily volatility of stocks in our sample. We observe that the daily volatility is left-skewed in the beginning of our sample, while it tends to be more symmetric in the recent years. As suspected, we notice a significant difference in the yearly ranges of daily volatility. There is a greater variability for volatility as well as larger outliers in 2008 and 2009 compared to the other years in the sample. The volatility in 2008 and 2009 ranges from zero to approximately 30% and 20%, respectively.

Table 1: Daily Volatility (%) Summary Statistics

The table reports summary statistics for the daily volatility in percentage terms. The summary statistics is reported in a monthly basis for S&P 500 stocks held by the ETFs in our sample. The summary statistics is reported at a monthly basis for S&P 500 stocks held by the ETFs in our sample. The sample cover the period between January 2008 to December 2018.

	N	Mean	SD	Median	Min	Max
2008	5,459	3.195	2.237	2.504	0.000	29.312
2009	5,475	2.432	1.637	2.000	0.120	19.692
2010	5,465	1.507	0.679	1.374	0.087	6.592
2011	5,439	1.612	0.825	1.415	0.038	8.197
2012	5,372	1.308	0.608	1.189	0.111	5.787
2013	5,370	1.153	0.483	1.053	0.045	8.150
2014	5,306	1.148	0.494	1.039	0.042	4.926
2015	5,334	1.322	0.634	1.178	0.000	6.938
2016	5,377	1.369	0.750	1.197	0.083	11.001
2017	5,327	1.058	0.476	0.943	0.109	4.450
2018	5,396	1.491	0.625	1.386	0.125	7.474
2008-2018	59,320	1.604	1.189	1.292	0.000	29.312

Figure 2: Variation in Daily Volatility



The figure shows the variation of daily volatility to the S&P 500 stocks in our sample in the years from 2008 to 2018. We note that the distributions are left-skewed and that the years 2008 and 2009 exhibit a high degree of volatility.

4.4 ETF ownership

4.4.1 Measuring ETF ownership

Our goal is to measure the effect ETF ownership has on the volatility of underlying securities; thus, our independent variable is ETF ownership. We calculate the ETF ownership variable following Equation 4. To calculate this variable, we need the weight that each ETF hold in a specific stock. This is commonly referred to as holding data and is known for being difficult to access for free. We access this data using the CRSP Mutual Fund Holdings database, through WRDS. Ben-David et al. (2018a) use Thomson Reuter Global Ownership database, however BI Norwegian Business School does not subscribe to it. CRSP Mutual Fund database seems to contain insufficient data[7] before 2008. Following this, we decide to restrict our sample to the period from January 2008 to December 2018. Moreover, CRSP

[7] Before 2008, CRSP Mutual Fund database does not contain the holdings of the funds in our sample.

Mutual Fund database does not contain the holdings for approximately 180 of our funds in the period from January 2008 to June 2010, these are extracted from Thomson Reuters Eikon (Table 2). Combining these sources, we find the holding data to be accurate in terms of providing a valid data sample for our purposes.

Table 2: ETF Holdings by Database

The table shows where each ETF is retrieved from. *TR Eikon* is Thomson Reuters Eikon and *CRSP* is CRSP Mutual Fund database. Year 2010 is a special case as CRSP Mutual Fund database reports all funds from mid-2010, thus Thomson Reuters Eikon is used to retrieve only the first half of the year.

Year	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
TR Eikon	180	181	88	0	0	0	0	0	0	0	0
CRSP	23	24	256	275	277	311	316	344	367	396	417
Total	203	205	256	275	277	311	316	344	367	396	417

Moreover, we need to ensure that we track the same equity ETFs in CRSP that we identified in Bloomberg. Using the *ticker* identifier from Bloomberg, we screen CRSP Mutual Fund Summary database in January 2020 to retrieve the *fundno* and *portno*'s for each fund. The *fundno* is a unique identifier distributed by CRSP for each fund in the database, while *portno* is the unique identifier for a security or a group of securities held in a fund's portfolio. The fund holdings in the CRSP Mutual Fund Holdings database are at the portfolio level, and fund portfolios can therefore include holdings of non-ETFs. For instance, several Vanguard ETFs are not standalone ETFs, because they are set up as share classes within the fund portfolio. To address this issue, we need to adjust the fund portfolio holdings to only consider the holdings of the ETF share class. First, we download the holding data using the *fundno* for each fund. Second, we find the ownership of each stock in the portfolio fund by using the percentage of the total net assets of all the holdings in the portfolio fund. Then, to compute the accurate ownership in each stock by the ETFs exclusively, we multiply the portfolio weights by the AUM of the ETFs. The monthly AUM for each ETFs is retrieved from CRSP Monthly Stock File database through WRDS.

We extract the holding data one year at the time for the set of ETFs as it returns a massive amount of data. For each security in the file CRSP assigns a permanent unique stock issue identifier, called *permno.* Our dataset is restricted to only investigate S&P 500 stocks. The yearly S&P 500 constituents are identified from Compustat – Capital IQ North America Daily database through WRDS using the code *i0003*, which exclusively defines S&P 500 constituents. Merging this sample with the CRSP Mutual Fund Holding data sample is challenging. Mainly because Compustat does not provide the *cusip* identifier for each stock at the corresponding historical time period, but the *cusip* at the time we download the data. The *cusip* identifier is changing over time, i.e. it is not permanent. While Compustat is always reporting the latest *cusip,* CRSP is reporting the historical *cusip.* Therefore, we have screened the holding data sample for matching *cusip, ticker,* and company names. We eliminate two of the stocks from our sample, who leaves and re-enters the S&P 500 during the sample period[8]. This is to reduce modelling difficulties regarding panel data sets with gaps and thereby provide more reliable estimates. Furthermore, we calculate the market capitalization for each of the S&P 500 stocks in our sample, as shares outstanding times closing price. We retrieve shares outstanding and the share price on the last trading day in a month from CRSP Monthly Stock File database through WRDS. From this data, we calculate the ETF ownership for each stock, as presented in Equation 4, by utilizing a self-written macro code in MS Excel.

In the beginning of our sample's time span ETFs were only reporting their holdings of stocks quarterly, while more recently ETFs reports their holdings monthly. Following Ben-David et al (2018a), who retrieves the most recent quarterly reports, we forward fill the quarterly holding data for the two missing months to obtain a monthly sample. Further, we screen the S&P 500 stocks by share code such that our stock sample only consist of ordinary common shares. The share codes are downloaded from CRSP Monthly Stock File database and we include stocks with

[8] Summary statistics and correlation matrix before adjustment for these gaps are provided in Appendix A.4 and A.5.

share codes 10 and 11[9]. Arbitrarily, no stock in our sample holds share code 10. Hence, we exclude stocks held by ETFs that are classified as closed end funds, incorporated outside the U.S., Americus Trust Components and Real Estate Investment Trusts (REITs).

4.4.2 Descriptive statistics of ETF ownership

Table 3 reports the summary statistics for ETF ownership. We observe that ETF ownership ranges from a minimum of 0.0% to a maximum of 21.7%. Thus, we have a higher maximum of ETF ownership than Ben-David et al. (2018a), who reports that the ETF ownership ranges from 0.0% to 11.2%. The average ETF ownership is 4.4%, which equals a difference of 1.8% compared to the reported average ETF ownership of Ben-David et al. (2018a). Further, we notice that we have a slightly higher median than Ben-David et al. (2018a), which coincide with the increased average of ETF ownership in our sample. The median ETF ownership across the years is 4.0%, while Ben-David et al (2018a) reports that the median is 2.3%.

Overall, our sample has a slightly higher maximum, average and median ETF ownership compared to Ben-David et al. (2018a). The differences might occur since we use a different sample period, and a different database to collect the holding data. However, since we look at a more recent period, where ETFs has continued to grow, it is reasonable that the ownership of ETFs in stocks has increased. Moreover, our sample consists only of funds that exists today (4.2), while Ben-David et al. (2018a) also include non-surviving funds. We assume that these differences occur because including non-surviving funds reduces the average and median ETF ownership, due to low AUM caused by bad performance. A yearly summary statistics of ETF ownership is provided in A.6.

[9] CRSP defines share code 10 as ordinary common shares which have not been further defined and share code 11 as ordinary common shares which need not be further defined.

Table 3: ETF Ownership Summary Statistics						
The table reports summary statistics for ETF ownership. The summary statistics is reported in a monthly basis for S&P 500 stocks held by the ETFs in our sample. The sample cover the period between January 2008 to December 2018.						
	N	Mean	SD	Median	Min	Max
ETF Ownership	59,320	0.044	0.021	0.040	0.000	0.217

Table 4 reports the descriptive statistics of the ETF ownership equation (Equation 4). The number of ETFs in the sample has increased substantially from 203 in 2008 compared to 372 in 2018, which aligns with the overall market increase of ETFs (A.7). Note that the number of ETFs in Table 4 is less than the number reported in Table 2, due to screening by S&P 500 stocks and share codes 10 and 11. Furthermore, the average AUM for ETFs has increased from $3.7 billion in 2008 to $11.2 billion in 2018. The average fraction of a stock's capitalization held by ETFs has increased by more than 200%, from 2.25% in 2008 to 7.25% in 2018.

Table 4: ETF Ownership Equation Descriptive Statistics

The table reports the descriptive statistics for the ETF ownership equation (Equation 4) of the S&P 500 stocks held by ETFs in our sample. For each year, the table reports the number of ETFs, the average of their AUM, the average weight of each stock in the ETF, the average market capitalization of the stocks, and the average percentage of each stock owned by ETFs.

Year	# ETFs	Average ETF AUM ($m)	Average Stock Weight in ETF (%)	Average Stock Market Cap ($m)	Average ETF Ownership (%)
2008	203	3,697.22	0.64	22,037.42	2.25
2009	205	3,679.66	0.66	17,662.37	2.86
2010	227	4,212.39	0.57	21,305.13	2.68
2011	245	4,074.39	0.59	24,180.27	3.28
2012	241	4,667.60	0.58	26,197.96	3.63
2013	270	6,052.21	0.59	31,025.23	4.19
2014	271	7,427.87	0.59	36,430.96	4.71
2015	302	7,868.21	0.57	38,488.86	5.20
2016	327	8,003.38	0.56	38,727.73	5.76
2017	352	10,192.59	0.57	45,341.56	6.66
2018	372	11,208.13	0.56	49,943.35	7.25

4.5 Control variables

We follow Ben-David et al. (2018a) and include seven control variables in our regression. Primarily, we want to control for other observable stock characteristics that might have an impact on the underlying stock volatility, besides ETF ownership. The characteristics we include in our regression are the following: LMCAP, IP, ILLIQ, BASPRD, BTM, GP and P12MRET. Appendix A.2 and A.8 gives a detailed description of the data collection of control variables used in this thesis, and Appendix A.9 reports the summary statistics for the control variables.

4.6 Preparation before empirical analysis

In this section, we provide an overview of different tests and considerations we employ to ensure that our empirical analysis is valid and reliable. We address issues related to multicollinearity, fixed effects, autocorrelation and heteroskedasticity, and volatility clustering.

4.6.1 Multicollinearity

When using the OLS estimation method an implicit assumption is that the explanatory variables are not correlated with each other (Brooks, 2019). Multicollinearity is a problem that occurs when the explanatory variables are highly correlated with one another, i.e. a correlation above 0.5 in absolute terms. The consequences when multicollinearity is present are that the regression becomes very sensitive to small changes in the specification and confidence intervals wide, leading to inappropriate conclusions (Brooks, 2019). One way to test for multicollinearity is to look at the correlation matrix between the individual variables and examine whether high correlations exists. In our dataset, we notice that the individual variables are not highly correlated (i.e. between -0.5 and 0.5). The highest correlation is between the Amihud (2002) ratio and the logged market capitalization and equals -0.482. Thus, we assume that multicollinearity is not an issue in our sample. The full correlation matrix is provided in Appendix A.10

4.6.2 Fixed effects

In panel data models it is important to test for whether to use the specification of fixed or random effects. Our methodology suggests that fixed effects is the proper specification, however we would like to check that this is in fact the case. We estimate a Hausman (1978) test on the full sample, without the inclusion of the lagged dependent variable. The null hypothesis of the Hausman (1978) test is that the difference in coefficients is not systematic, whereas the alternative is that it is. From estimating the Hausman (1978) test we find a p-value of 0.000 suggesting that we reject the null hypothesis and conclude that the differences in coefficients are systematic (A.11). The test results suggest that under our model specification a fixed effects model is appropriate. Moreover, we would also like to check if month-fixed effects should be included in our model. We estimate a regression using stock-fixed effects and include month-dummies[10]. Then we apply the Wald test, which is a joint test to see if the dummies for all months are equal to zero, if they are not, month-fixed effects should be applied in our regressions. The null hypothesis of the Wald test is that coefficients for all months are jointly equal to zero, whereas the alternative is that they are not. The test statistics returns a p-value of 0.000, which suggest that the month-dummies are in fact not jointly equal to zero (A.12), thus we include month-fixed effects in our model.

4.6.3 Autocorrelation and heteroskedasticity

Autocorrelation in linear panel-data is present when errors are not uncorrelated with each other, which can bias the standard errors of estimations (Brooks, 2019). In two-way fixed effects models, the error term exhibits a three-part structure: u_i: unobservable stock effects, v_t: unobservable time fixed effects, $\varepsilon_{i,t}$: idiosyncratic error. The advantage of employing a fixed effects model is that serial correlation in u_i and v_t can be ruled out. However, serial correlation in $\varepsilon_{i,t}$ cannot be ruled out by applying the fixed effect estimator. To control whether autocorrelation is present in our data, we estimate the Woolridge test implemented by Drukker (2003) for

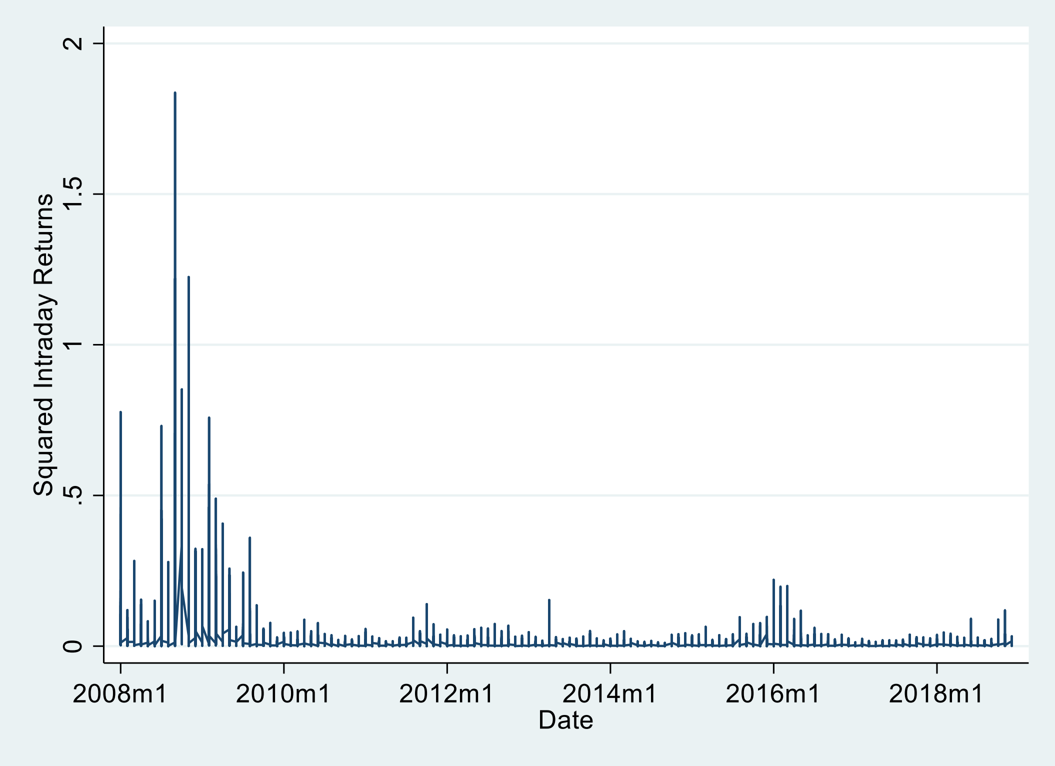
[10] The dummy variables (or indicator variables) takes the value 1 for a specific month, and 0 for all other months.

serial correlation in panel-data models. We test the null hypothesis that there is no first-order autocorrelation. The test statistic reveals a p-value of 0.000, hence we reject the null hypothesis and conclude that autocorrelation is present. To deal with this issue, a solution is to employ clustered standard errors in our estimations. These standard errors are derived from a robust variance matrix suggested by Arellano (1987) and follows from the general results of White (1984). It is valid in the presence of any heteroskedasticity and serial correlation, given that the number of time periods (T) is small relative to the number of stocks (N) (Woolridge, 2010). Since our sample is large N (665 stocks) and small T (132 months), cluster robust standard errors will be a suitable solution to the issues regarding autocorrelation and heteroskedasticity.

4.6.4 Lags of dependent variable

For financial data, a stylized fact is volatility clustering. According to Brooks (2019), volatility clustering is the tendency in asset prices of large changes (in either direction) to follow large changes and small changes (in either direction) to follow small changes. Thus, resulting in persistence in the amplitudes of asset price changes. It follows that the volatility today tends to be positively correlated with the volatility in immediately preceding periods (Brooks, 2019). In Equation 2 we included three lags of daily volatility, to account for this persistence. We would like to control that this is in fact the case for our sample. From Figure 3, we observe that the squared aggregated returns, across stocks and months in our sample, seems to appear in clusters. By including an arbitrary number of lags of the dependent variable equal to 12, we estimate a regression to control how many numbers of lags is appropriate in our model. We find that up to three lags of the dependent variable are significant at the 5% level, while the remaining lags are not significant at this level. Thus, we find it reasonable to include three lags of the dependent variable in our model.

Figure 3: Volatility Clustering



The figure depicts volatility clustering from the aggregated squared returns across months and stocks in our sample.

5 Empirical Results and Analysis

This section presents and discusses the empirical results from our regressions, with regards to our research question:

Does ETF ownership increase S&P 500 Index stocks volatility?

This section is divided into three parts. The first part evaluates the effects between ETF ownership and volatility, controlled for different stock characteristics. Further, we provide an analysis of the economic magnitudes of the OLS estimates. Last, we evaluate the robustness of our analysis to check whether our models hold with different conditions.

5.1 Empirical findings

In the following we discuss the regression results from the full sample of S&P 500 stocks in the period from 2008 to 2018. Table 5 reports the estimation results for the regressions given in Equation 1 and Equation 2. The sample consists of all the 396 funds, and daily volatility is used as the dependent variable. In regression 1, we estimate the daily volatility on ETF ownership, while controlling for common stock characteristics and fixed effects. In regression 2, we replicate the analysis of regression 1 by using a subsample where three lags of daily volatility are available. In regression 3, we include three lags of the daily volatility to address the concern that the persistence in volatility might cause reverse causality. The variable of interest is the coefficient of ETF ownership since it captures the relation between ETF ownership and volatility when controlling for stock characteristics.

To begin we examine regression 1, where we are studying a panel consisting of 665 stocks over 132 months. First, we observe that all variables are statistically significant at the 10% level. The R^2 of 0.720 suggests that the variables in our regression explain some of the variation in daily volatility, however a great part is still unexplained. The ETF ownership variable is statistically significant at the 1% level, which indicates that we have strong evidence against the null hypothesis that ETF ownership does not increase stock volatility. Empirically, a one-standard-deviation change in ETF ownership is associated with a 12.4% increase in the standard deviation of daily volatility.

Moreover, in regression 2 we are estimating the same number of stocks as in regression 1. The reduced number of observations comes from eliminating observations where three lags of the dependent variable are not available. By looking at the slope of the ETF ownership variable, we find that the reduction in observations is not significantly impacting the variable estimates from regression 1. Thus, we can be sure that it is not the reduced number of observations that are driving the results in the following regression including three lags of the dependent variable.

Continuing with regression 3, we estimate the regression with three lags of the dependent variable. Compared to the first regression the R^2 has increased, implying that this model explains more of the variation in daily volatility. In this regression the statistical significance of the control variables is substantially reduced. However, the ETF ownership variable is still statistically significant at the 1% level, but the t-statistic is somewhat lower than in regression 1. The empirical interpretation of the ETF ownership variable in this case is that a one-standard-deviation change in ETF ownership is associated with a 4.9% increase in the standard deviation of daily volatility. This is a reduction of 7.5%, which indicates that there is some persistence in daily volatility.

Overall, we observe that the ETF ownership coefficients are positive and highly statistically significant at the 1 % level in all three regressions. The positive and significant relation between ETF ownership and the daily stock volatility provides evidence in support of our hypothesis that ETF ownership increase S&P 500 stock volatility. As such, we have found the answer to our research question:

ETF ownership increases S&P 500 Index stock volatility.

Our findings are consistent with Krause et al. (2014), that there is a statistically significant relation between ETFs and underlying stocks. Similar to Krause et al. (2014), an identification strategy is required to confirm whether there is an exogenous relation. However, our findings are consistent with Ben-David et al.'s (2018a), thus we choose to rely on their study to assume that there is an exogeneous relation between ETF ownership and stock volatility. The increase in volatility suggests that stocks held by ETFs can be attractive to noise traders and short-term investors, as they offer more trading opportunities. Hence, our analysis so far suggests that regulators should be cautious about the growth of ETFs regarding financial market stability.

Table 5: OLS Regression - Full Sample

The table reports OLS estimates from regressions of daily volatility on ETF ownership and control variables with a monthly lag. The sample consists of S&P 500 stocks, and it covers the period from January 2008 to December 2018. The frequency of observations is monthly, and volatility is estimated using the daily returns within the month. The control variables include logged market capitalization, the lagged inverse price, the lagged Amihud (2002), the lagged bid-ask spread, the lagged book-to-market, lagged gross profitability (Novy-Marx, 2013), and lagged past 12-month returns. Regression 3 includes three lags of the dependent variable. The dependent variable and ETF ownership variable are standardized, and standard errors are double-clustered at the month and stock levels. The t-statistics are shown in parentheses, and significance at 10%, 5%, and 1% is indicated by ***, ** and *, respectively.

Dependent variable:		$DVOL_{i,t}$	
Sample:		Full Sample	
Regression (#)	(1)	(2)	(3)
$ETFOWN_{i,t}$	0.124***	0.123***	0.049***
	(4.62)	(4.49)	(3.50)
$LMCAP_{i,t-1}$	-0.100***	-0.100**	-0.042*
	(-2.61)	(-2.54)	(-1.83)
$IP_{i,t-1}$	4.578***	4.674***	1.469*
	(3.98)	(3.86)	(1.69)
$ILLIQ_{i,t-1}$	399.2***	394.4***	87.75
	(4.12)	(4.04)	(1.37)
$BASPRD_{i,t-1}$	0.473***	0.499***	0.196
	(2.68)	(2.72)	(1.55)
$BTM_{i,t-1}$	0.178***	0.177***	0.056
	(3.77)	(3.64)	(1.44)
$GPROFIT_{i,t-1}$	-0.665**	-0.636**	-0.122
	(-2.36)	(-2.23)	(-0.75)
$P12MRET_{i,t-1}$	-0.073*	-0.080*	-0.040
	(-1.92)	(-1.94)	(-1.56)
$DVOL_{i,t-1}$			0.340***
			(8.90)
$DVOL_{i,t-2}$			0.179***
			(3.59)
$DVOL_{i,t-3}$			0.115***
			(4.75)
Month fixed effects	Yes	Yes	Yes
Stock fixed effects	Yes	Yes	Yes
Standard Error	Clustered	Clustered	Clustered
Observations	58,655	57,325	57,325
R^2	0.720	0.721	0.792

5.2 Magnitude estimation and discussion of OLS estimates

In the following, we provide an analysis of the economic magnitude ETF ownership has on stock volatility, based on the results reported in Table 5. To assess the economic significance, we measure the shift in volatility relative to the median stock in the sample (Ben-David et al., 2018b). First, we allocate percentiles of the volatility distribution to the stock in each month. Further, we estimate the standard deviation of ETF ownership for each of the percentiles and focus on the 50th and 51st percentiles. Then we multiply the estimated standard deviation with the OLS coefficient. To obtain the new level of volatility, we add this term to the median volatility in the last month of the sample. Utilizing this calculation, we can examine the effect of a one-standard-deviation change of ETF ownership for the median stock in the sample. Table 6 reports the median volatility, which is the starting point before applying the variation in ETF ownership, the standard deviation of ETF ownership and the new level of volatility with the corresponding volatility quantile.

When examining the economic magnitudes, we begin with regression 1, where the lags of daily volatility are not included. First, we observe that the standard deviation of ETF ownership is equal to 2.1%. Second, this causes the median stock to increase by 2.3%, which corresponds to a shift in volatility to the 73rd percentile of the volatility distribution for S&P 500 stocks. When including the lags of daily volatility in regression 3, the median stock increases by about 2.2%. This corresponds to a new level of volatility to the 60th percentile. Thus, the economic magnitude appears to be slightly weaker when the lagged volatility is in included in the regression, however it is still significant.

We note that the shift in volatility induced by a one-standard-deviation change is greater when the lags of daily volatility are not included in the regression of the full sample. This corresponds with the OLS estimates reported in Table 5, where ETF ownership has a greater impact on the stock volatility when lags are not included. We conclude that a normal shock to ETF ownership causes the median stock volatility to shift to a place between the 60th and 73rd percentiles of the volatility distribution for S&P 500 stocks, where the lower bound coincides with the regression including three lags of daily volatility. This implies that the economic magnitudes of the OLS regressions are significant. Our results for the economic magnitudes are similar to the findings of Ben-David et al. (2018a), as they conclude that the median stock shifts to a place between the 58th and 64th percentiles for S&P 500 stocks, in the period from 2000 to 2015. These differences might be due to the higher values of ETF ownership in our sample. This strengthens our belief that regulators should be cautious of the growth of ETFs, as the economic magnitudes are of significance.

Table 6: Magnitude Estimation

The table reports the economic magnitude of ETFs on S&P 500 stock volatility. It reports the median volatility, the estimated standard deviation of ETF ownership for the median stock, the new level of volatility after applying the variation, and the new percentile in the volatility distribution which is achieved after applying the change in ETF ownership.

Sample:	Full sample	
Regression (#)	(1)	(3)
Lags of dep. variable in regression:	No lags	Three lags
Median volatility	0.021	0.021
Std. dev of ETF ownership	0.021	0.021
New level of volatility	0.023	0.022
New quantile of volatility	73	60

5.3 Robustness

To control the validity of our model, we employ two robustness tests. First, we test if another measure of volatility has an impact on the OLS estimates. Second, we split the sample in two periods, namely “in crisis” and “after crisis”, to control whether the relation between ETF ownership and stock volatility is persistent throughout the sample period.

5.3.1 Realized Volatility as measure of daily volatility

As briefly mentioned in 3.3.1, there are other ways to measure volatility than the definition we employed in our study. To control the robustness of our regression models, we estimate the regressions again with another common measure of volatility. Specifically, we employ the measure known as realized volatility, which is derived from the realized variance introduced by Andersen, Bollerslev, Diebold and Labys (2001) and Barndorff-Nielsen and Shephard (2002). It is a backward-looking metric, which measures how much the price has moved over a particular period in the past (Iqbal, 2018).

Realized volatility is given as

$$RDVOL_{i,t} = \sqrt{\sum_{d=1}^{n} r_{i,d}^2}$$

Equation 8: Realized Volatility

where $RDVOL_{i,t}$ is the realized volatility of each stock i at month t, and $r_{i,d}^2$ is the squared return at time d.

The regression estimates, given in Appendix A.13, report the same three regression models as in Table 5, with realized volatility as the dependent variable. The summary statistics and correlation matrix for the subsamples are provided in Appendix A.14 and A.15. We find that the slope of the coefficient for ETF ownership is close to exactly equal to the original model (Table 5), for all three regressions (A.13). It is positive and statistically significant at the 1% level. The estimation results show that our model is robust to other measures of volatility.

5.3.2 Control for financial crisis

According to NBER (n.d.) the U.S. financial market was in a recession from December 2007 to June 2009. This might have an impact on our results, as recessions are often accompanied by high volatility. As observed in Figure 2, the years 2008 and 2009 exhibit significantly higher levels of volatility as well as higher variability in volatility compared to the rest of the years in the sample. To test whether the relation between ETF ownership and stock volatility is persistent throughout the sample period, we divide the sample into two subsamples. The first sample includes stocks held by ETFs in the period from January 2008 to June 2009, while the second sample includes the period from July 2009 to December 2018. We define the two periods as "in crisis" and "after crisis". The summary statistics and correlation matrices for the subsamples are provided in Appendix A.16-A.19. We estimate the same three regression models as in the full sample (Equation 1 and Equation 2) and estimates are given in Appendix A.20. From the results, we infer that ETF ownership's impact on stock volatility is somewhat different in the two subsamples.

Beginning with the "in crisis" period from January 2008 to June 2009, we have a small sample compared with the full sample, with only 7,713 observations. It comprises 497 stocks and 18 time periods. The frequent rule of thumb for a cluster variable is that it needs at least 50 different "categories". This leads us to make an adjustment to the clustering, where we only apply clustering on the stock level to avoid misleading estimates. The relatively high R^2 implicates that the variables in our regression explain some of the variation in stock volatility. We observe that the impact ETF ownership has on stock volatility is substantially weaker in all three regressions relative to the full sample (Table 5). Due to the financial crisis one would assume that the financial markets exhibit higher illiquidity, hence exploiting the channel of ETF arbitrage would have a larger impact on stock prices[11]. However, because coefficient estimates for the ETF ownership variable are not significant, we are careful about drawing any conclusions from the model. We suspect that these results are largely impacted by the smaller sample size, with only 1 year and 6 months of data.

Investigating the "after crisis" sample, we are looking at a total of 628 stocks. Here the number of months is greater than 50, thus we employ double clustering on the standard errors. The regressions suggest that the independent variables explain around 60% of the variation in daily volatility, which is lower than in the full sample. Compared to the full sample the coefficient estimates, for regression 1 and 2, are slightly lower at a significance level of 1%. When three lags of daily volatility are not included, a one-standard-deviation increase in ETF ownership is associated with an increase of 11.1% in the standard deviation of daily volatility. These findings are in line with our results for the full sample, indicating that the relation between ETF ownership and stock volatility holds.

[11] Details on the channel of ETF arbitrage is provided in Appendix A.1.2.

5.4 Limitations

We acknowledge that our analysis has some limitations. We consider our main challenge and following implications to be regarding the data sampling. However, we believe that these limitations have not caused major disadvantages to our research, as such our findings and ideas might be helpful and inspirational for future research. In the following we discuss some of these limitations and their implications.

5.4.1 Reliable data

The main concern about our analysis is regarding the holding data and its accuracy. Retrieving reliable holding data for funds is difficult and time consuming. We have used data from CRSP Mutual Fund Database which is a different source than Ben-David et al. (2018a), who we compare our results to. In our data sample we found that CRSP does not contain data for all funds in the first 2.5 years. We relied on Thomson Reuters Eikon derived holdings as our data source for these missing funds. Additionally, the forward filling of quarterly data might contribute to some degree of inaccuracy. We have ensured that there are no double-entries in our forward filling process, thus we find that the holding data should be reliable enough for our purpose assuming that the holdings do not vary a lot in the short-run. We would also like to note that our sample of funds is subject to survivorship bias, which might have influenced our results. If non-surviving funds were included, the total ETF ownership would have increased, suggesting that our regression estimates would have increased as well.

5.4.2 Index, Active and Hedge Fund

Since ownership of other institutional investors can influence the stock volatility, it would be valuable to examine whether these ownerships capture some of the impact ETF ownership has stock volatility. Institutional investors similar to ETFs are open-ended funds, such as index mutual funds and other active mutual funds, since they tend to receive funds on a daily basis (Ben-David et al., 2018a). In addition, it could be valuable to control for hedge funds since they are also likely to trade at a high frequency like ETF arbitrageurs (Ben-David et al., 2018a). However, because of limited data availability, we are not able to obtain data on mutual fund's total asset value. This restricted us from controlling for other institutional ownership in the OLS regression.

6 Conclusion and Recommendation

6.1 Conclusion

ETFs are passive investment vehicles that have been increasing in popularity since their invention in the 1990's. While volatility in financial markets have been studied by economists and researchers for decades, the relation between ETFs and volatility is still an area to be further explored. Using a panel data set comprising 396 U.S. equity ETFs in the period from 2008 to 2018, this thesis has established a two-way fixed effects model that captures the effect ETF ownership has on volatility of S&P 500 stocks and investigated their impacts.

Our analysis finds that, overall ETF ownership does increase the volatility of underlying S&P 500 stocks. For the full sample between 2008 and 2018, we estimate that a one-standard deviation increase in ETF ownership leads to an increase in the standard deviation of daily volatility to somewhere between 4.9% and 12.4%, where the lower bound corresponds to the regression with three lags of daily volatility. The increase in daily volatility is of size and statistically significant. Further, our findings suggest that for the median stock in our sample, a one-standard-deviation increase in ETF ownership will shift the median stock to a place between the 60th and 73rd percentiles of the volatility distribution. Our findings are somewhat consistent with Ben-David et al. (2018a), who finds evidence that ETF ownership increases volatility of underlying stocks and a shift in the median stock between 58th and 64th percentiles.

Further, investigating the robustness of our analysis we find that our results are independent of changes in conditions, such as employing other measures of volatility. However, we find that in periods of financial instability our findings are not consistent with the full sample. For our sample, we can only conclude that during crisis there is no statistical significance of a change in volatility of underlying stocks due to ETF ownership.

Based on our analysis, we have shown that ETFs does contribute to the volatility of underlying stocks. As such, regulators should be concerned about their impact on the financial market stability.

6.2 Recommendations

As the ETF market is growing in an exponential-like pattern, it is important that further studies are conducted investigating their impact in financial markets. For future studies we would suggest that an even longer sample period is employed to estimate the long-run effects of ETF ownership on volatility, such that economic cycles are accounted for. Until this date, the long-run effects are yet to be determined.

Bibliography

Abner, D. J. (2016). *The ETF handbook: how to value and trade exchange traded funds.* John Wiley & Sons.

Amihud, Y. (2002). Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets,* 5(1), 31-56.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, 96(453), 42-55.

Agarwal, V., Hanouna, P., Moussawi, R., & Stahel, C. W. (2018). Do ETFs increase the commonality in liquidity of underlying stocks? *In 28th Annual Conference on Financial Economics and Accounting.*

Arellano, M. (1987). Computing robust standard errors for within-groups estimators. Oxford bulletin of Economics and Statistics, 49(4), 431-434.

Bae, K., Chan, K., & Ng, A. (2004). Investibility and return volatility. *Journal of Financial Economics,* 71(2), 239-263.

Barndorff-Nielsen, O. E., & Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(2), 253-280.

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

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

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

Black, F. (1986). Noise. *The Journal of Finance*, 41(3), 528-543.

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

Brown, G. W. (1999). Volatility, sentiment, and noise traders. *Financial Analysts Journal*, 55(2), 82-90.

Brown, S. J., Goetzmann, W., Ibbotson, R. G., & Ross, S. A. (1992). Survivorship bias in performance studies. *The Review of Financial Studies*, 5(4), 553-580.

Correia, S. (2017). reghdfe: Stata module for linear and instrumental-variable/gmm regression absorbing multiple levels of fixed effects. *Statistical Software Components s457874,* Boston College Department of Economics.

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

Deville, L. (2008). Exchange traded funds: History, trading, and research. *In Handbook of financial engineering* (pp. 67-98). Springer, Boston, MA.

Dannhauser, C. D. (2016). The impact of innovation: Evidence from corporate bond ETFs. *Journal of Financial Economics (JFE)*, Forthcoming.

Diebold, F. X., & Yilmaz, K. (2012). Better to give than to receive: Predictive directional measurement of volatility spillovers. *International Journal of Forecasting*, 28(1), 57-66.

Drukker, D. M. (2003). Testing for serial correlation in linear panel-data models. *The Stata Journal*, 3(2), 168-177.

Edwards, F. R. (1988). Does futures trading increase stock market volatility?. *Financial Analysts Journal*, *44*(1), 63-69.

Elton, E. J., Gruber, M. J., & Blake, C. R. (1996). Survivor bias and mutual fund performance. *The Review of Financial Studies*, 9(4), 1097-1120.

Ferri, R. A. (2011). *The ETF book: all you need to know about exchange-traded funds.* John Wiley & Sons.

Glosten, L., Nallareddy, S., & Zou, Y. (2020). ETF activity and informational efficiency of underlying securities. *Management Science.*

Hamm, S. (2014). The effect of ETFs on stock liquidity. *Social Science Research Network Working Paper Series.*

Harris, L. (1989). S&P 500 cash stock price volatilities. *The Journal of Finance*, 44(5), 1155-1175.

Harris, L. (2003). *Trading and exchanges: Market microstructure for practitioners.* OUP USA.

Hausman, J. A. (1978). Specification Tests in Econometrics. *Econometrica*, 46(6), 1251-1271.

Hehn, E. (2005). Exchange Traded Funds: Structure, Regulation and Application of a New Fund Class (1st ed. 2005. ed.). Berlin, Heidelberg: Springer Berlin Heidelberg : Imprint: Springer.

Hegde, S., & Mcdermott, J. (2004). The market liquidity of DIAMONDS, Q's, and their underlying stocks. *Journal of Banking and Finance*, 28(5), 1043-1067.

Hill, J. M., Nadig, D., & Hougan, M. (2015). A Comprehensive Guide to Exchange Traded Funds. CFA Institute Research Foundation.

Investment Company Institute (2019) Investment company fact book–a review of trends and activities in the U.S. investment company industry, vol. 59.

Israeli, D., Lee, C. M., & Sridharan, S.A. (2017). Is There a Dark Side to Exchange Traded Funds? An information perspective. *Review of Accounting Studies* 22(3), 1048-083.

Iqbal, A. S. (2018). Volatility: Practical Options Theory. John Wiley & Sons.

Krause, T., Ehsani, S., & Lien, D. (2014). Exchange-traded funds, liquidity and volatility. *Applied Financial Economics*, *24*(24), 1617-1630.

Lettau, M., & Madhavan, A. (2018). Exchange-traded funds 101 for economists. *The Journal of Economic Perspectives, 32*(1), 135-154. doi: http://dx.doi.org.ezproxy.library.bi.no/10.1257/jep.32.1.135

Malamud, S. (2015). A dynamic equilibrium model of ETFs. Swiss Finance Institute Research Paper No. 15-37.

Markovitz, H. M. (1959). *Portfolio selection: Efficient diversification of investments.* John Wiley.

Mazumder, I. (2014). Investing In Exchange Traded Funds. *Applied Finance Letters*, *3(2)*, 16-23.

NBER. (n.d). US Business Cycle Expansions and Contractions. Retrieved from https://www.nber.org/cycles.html

Novy-Marx, R. (2013). The other side of value: The gross profitability premium. *Journal of Financial Economics*, 108(1), 1-28.

Poon, S. H., & Granger, C. W. (2003). Forecasting volatility in financial markets: A review. *Journal of Economic Literature, 41(2)*, 478-539.

Stambaugh, R. F. (2014). Presidential address: Investment noise and trends. *The Journal of Finance*, 69(4), 1415-1453.

State Street Global Advisors. (n.d.). SPDR® S&P 500® ETF Trust. Retrieved from https://www.ssga.com/us/en/individual/etfs/funds/spdr-sp-500-etf-trust-spy

Stratmann, T., & Welborn, J. W. (2012). Exchange-traded funds, fails-to-deliver, and market volatility.

Trainor Jr, W. J. (2010). Do leveraged ETFs increase volatility. *Technology and Investment*, *1(3)*, 215.

U.S. Securities and Exchange Commission. (n.d). Exchange-Traded Funds (ETFs). Retrieved from https://www.investor.gov/introduction-investing/investing-basics/investment-products/mutual-funds-and-exchange-traded-2

Wermers, R. (1997). Momentum investment strategies of mutual funds, performance persistence, and survivorship bias. *University of Colorado. Working Paper.*

White, H. (1984). *Asymptotic Theory for Econometricians*. Orlando, FL: Academic Press

Wooldridge, J. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). Cambridge, Mass.: MIT Press.

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

Appendix

A.1 Institutional Details

ETFs were introduced on U.S. and Canadian exchanges in the early 1990s and have become a popular investment product in recent years (Deville, 2008). The first ETF listed on a U.S. exchange was the SPDR S&P 500 ETF Trust, commonly known as Standard & Poor's Depository Receipt (SPDR, ticker: SPY), and it was developed to track the S&P 500 stock market index (Lettau & Madhavan, 2018). SPY is one of the most actively traded ETFs today, and as of July 2nd, 2020 it accounted for \$273 billion in AUM (State Street Global Advisors, n.d). U.S. Securities and Exchange Commission (SEC, n.d) identifies two types of ETFs: Index-Based ETFs and actively managed ETFs. Index-based ETFs investment objective is to track a securities index, like the S&P 500, and generally invest primarily in the component securities of the index. Conversely, actively managed ETFs are not based on an index, instead they seek to achieve a stated investment objective by investing in a portfolio of stocks, bonds, and other assets. The main difference between the two is that in an actively managed ETF the components in the portfolio can be traded by an advisor daily, without regard to conformity with an index. At year-end 2019, 1,708 index-based ETFs and 320 actively managed ETFs were registered with SEC under the Investment Company Act of 1940.

A.1.1 ETFs vs. Mutual Funds

Similar to (index) mutual funds, index-based ETFs are designed to replicate the performance of an index as closely as possible. An index-based ETF can either invest in all stocks of an index or invest in a representative sample of the index (Hehn, 2005). ETF managers are, like mutual fund managers, required by the SEC to announce a NAV of their funds at the end of each trading day (Lettau & Madhavan, 2018). However, ETFs differ in some important respects from mutual funds, such as pricing. For instance, ETFs are traded on stock exchanges and can be traded continuously throughout the trading day, while mutual fund transactions only occur at the end of the day at the NAV (Deville, 2008). In addition, ETFs do not trade with capital markets directly, instead the fund managers issue or acquire shares in large blocks, known as creation or redemption units, to authorized participants (APs) (Lettau & Madhavan, 2018). APs are a small group of institutions that have signed an agreement with the ETF provider who can trade bundles of ETF shares, often 50,000 shares, with the ETF manager (Ben-David et al., 2018a).

A.1.2 Creation/Redemption Function

An ETF can be traded in two markets, the primary and the secondary market (Dannhauser, 2016). The primary market is used by ETFs to manage liquidity shocks in the secondary market and to engage in arbitrage activity if the market price deviates from the NAV. The primary market links ETFs to the underlying stocks. Only the APs and managers of the ETFs participate in the primary market for the creation/redemption of ETF shares (Israeli et al., 2017). The APs are responsible for buying a basket of shares for the ETF manager, and in exchange the APs receive a number of ETF creation units. In the secondary market, the APs can buy or sell ETF shares directly on the exchange (Dannhauser, 2016). This mechanism, by which the shares of the ETF are adjusted in response to supply and demand, is known as the creation/redemption mechanism (Lettau & Madhavan, 2018). “Creations” is here referring to increasing the supply of ETF shares, while “redemptions” refers to a decrease in the ETF shares outstanding.

Since an ETF can be traded in two markets, it has two prices: the market price and the NAV (Deville, 2008). The market price of an ETF often deviates from its NAV, and an arbitrage opportunity then arise between the ETF shares and the underlying basket of securities (Ben-David et al., 2017). This deviation may be explained by time inconsistencies and inefficiencies in the share creation/redemption process (Stratmann & Welborn, 2012). APs create and redeem shares to ensure that the market price of the ETF remain extremely close to the NAV (Mazumder, 2014).

To demonstrate the arbitrage process through the creation and redemption of ETF shares, we differentiate between two cases: ETF premium and ETF discount (Ferri, 2011). When the ETF is trading at a premium (a price above the NAV per share of the ETF), the APs can purchase the cheaper underlying stocks while simultaneously sell the expensive ETFs shares. These actions will reduce the ETF price and increase the price of the underlying stocks, and thereby bring the price of the ETF and the price of the underlying stocks closer together. In the case of an ETF discount (a price below the NAV per share of the ETF), it can be profitable for the APs to purchase ETF shares and sell the underlying stocks. The increased demand for the ETF will raise the ETF price and lower the price of the underlying stocks, which narrows the gap between the ETF and its underlying value. Thus, the equilibrium market price of an ETF will be restored by the dynamic market mechanisms with respect to its creation and redemption units' NAV.

A.2 Variable Description

This table shows a description of the variables used in this thesis.

Variable	**Description**	**Formula**	**Source**
Volatility	Standard deviation of intraday returns.	$DVOL_{i,t} = \sqrt{\frac{\sum_{d=1}^{n}(r_{i,d} - \bar{r}_{i,d})^2}{n-1}}$	CRSP
ETF ownership	ETF ownership represents how much each stock is owned by the ETFs, and it is measured by using the most recent quarterly reports for the companies.	$ETFOWN_{i,t} = \frac{\sum_{j=1}^{J} w_{i,j,t} AUM_{j,t}}{Mkt\ Cap_{i,t}}$	CRSP, Bloomberg, Thomson Reuters Eikon
Log (Market Capitalization)	The logged market capitalization of the stocks (in millions), using the closing price.	$LMCAP_{i,t} = log\left(LMCAP_{i,t}\right)$	CRSP
Inverse Price	The inverse price is the inverse of the *price alternate*, which the last available price at the end of the month, derived from daily prices.	$IP_{i,t} = \frac{1}{Price_{i,t}}$	CRSP
Amihud (2002)	The absolute daily return divided by the total daily dollar volume (in millions),	$ILLIQ_{i,t} = \frac{1}{N}\sum_{j=1}^{d_t}\frac{\lvert r_{i,j}\rvert}{\$V_{i,j}}$	CRSP

	following Amihud (2002).		
Book-to-Market	Book-to-Market ratio is measured by dividing total assets less total liabilities, intangible assets and preferred shares by the market capitalization.	$BTM_{i,t} = \frac{Total\ Book\ Value_{\ i,t}}{Market\ capitalization_{i,t}}$	Compustat
Gross Profitability	Gross profits (revenues less cost of goods sold) scaled by total assets, following Novy-Marx (2013).	$GPROFIT_{i,t} = \frac{Revenue_{i,t} - Cost\ of\ Goods\ Sold_{i,t}}{Total\ assets_{i,t}}$	Compustat
Past 12-Month Returns	The total return of the stock between the close of t_1 and the close of t_{12}.	$P12MRET_{i,t} = \log\left(\frac{price_{i,t}}{price_{i,t-12}}\right)$	CRSP
Bid-Ask Spread	The difference between the bid price for the stocks and its ask price, divided by the bid-ask spread average.	$BASPREAD_{i,t} = \frac{Ask\ Price_{i,t} - Bid\ Price_{i,t}}{Bid-Ask\ Spread\ Average_{i,t}}$	CRSP

A.3 List of ETFs in the Sample

This table presents the 396 ETFs included in the sample.

AI Powered Equity ETF	First Trust NASDAQ Technology Div. Index Fund
Alpha Architect US Quantitative Value ETF	First Trust NASDAQ-100 Equal Weighted Index F.
ALPS Sector Dividend Dogs ETF	First Trust NASDAQ-100 Ex-Technology Sector
American Century STOXX US Quality Growth ETF	First Trust NASDAQ-100 Technology Index Fund
American Century STOXX US Quality Value ETF	First Trust North American Energy Infrastr. Fund
Barron's 400 ETF	First Trust NYSE Arca Biotechnology Index Fund
Cambria Shareholder Yield ETF	First Trust Rising Dividend Achievers ETF
Communication Services Select SPDR Fund	First Trust S&P REIT Index Fund
Consumer Discretionary Select SPDR Fund	First Trust Small Cap Core AlphaDEX Fund
Consumer Staples Select Sector SPDR Fund	First Trust Small Cap Growth AlphaDEX Fund
Davis Select US Equity ETF	First Trust Technology AlphaDEX Fund
DBX ETF Trust - Xtrackers Russell 1000 US	First Trust US Equity Opportunities ETF
Deltashares S&P 500 Managed Risk ETF	First Trust Utilities AlphaDEX Fund
Direxion NASDAQ-100 Equal Weighted Index Shares	First Trust Value Line Dividend Index Fund
Energy Select Sector SPDR Fund	First Trust Water ETF
ERShares Entrepreneur 30 ETF	FlexShares Morningstar US Market Factor Tilt Index
ETF Series Solutions - Deep Value ETF	FlexShares Quality Dividend Defensive Index Fund
Fidelity Dividend ETF for Rising Rates	FlexShares Quality Dividend Index Fund
Fidelity High Dividend ETF	Franklin LibertyQ US Equity ETF
Fidelity Low Volatility Factor ETF	Global SuperDividend US ETF
Fidelity Momentum Factor ETF	Global X Adaptive US Factor ETF
Fidelity MSCI Communication Services Index ETF	Global X MLP & Energy Infrastructure ETF
Fidelity MSCI Consumer Discretionary Index ETF	Global X MLP ETF
Fidelity MSCI Consumer Staples Index ETF	Global X Nasdaq 100 Covered Call ETF
Fidelity MSCI Energy Index ETF	Global X S&P 500 Catholic Values ETF
Fidelity MSCI Financials Index ETF	Global X S&P 500 Covered Call ETF
Fidelity MSCI Health Care Index ETF	Global X US Infrastructure Development ETF
Fidelity MSCI Industrials Index ETF	Goldman Sachs ActiveBeta U.S. Large Cap Equity
Fidelity MSCI Information Technology Index ETF	Goldman Sachs Equal Weight US Large Cap Equity
Fidelity MSCI Materials Index ETF	Goldman Sachs JUST US Large Cap Equity ETF
Fidelity MSCI Real Estate Index ETF	Hartford Multifactor US Equity ETF
Fidelity MSCI Utilities Index ETF	Health Care Select Sector SPDR Fund
Fidelity NASDAQ Composite Index Track. S. ETF	Industrial Select Sector SPDR Fund
Fidelity Quality Factor ETF	InfraCap MLP ETF
Fidelity Value Factor ETF	Innovator IBD 50 ETF
Financial Select Sector SPDR Fund	Invesco Active US Real Estate Fund
First Trust Capital Strength ETF	Invesco Aerospace & Defense ETF
First Trust Cloud Computing ETF	Invesco BuyBack Achievers ETF
First Trust Consumer Discretionary AlphaDEX Fund	Invesco Defensive Equity ETF
First Trust Consumer Staples AlphaDEX Fund	Invesco Dividend Achievers ETF
First Trust Dorsey Wright Momentum & Low Vol.	Invesco DWA Healthcare Momentum ETF
First Trust Dow Jones Internet Index Fund	Invesco DWA Industrials Momentum ETF
First Trust Financial AlphaDEX Fund	Invesco DWA Momentum ETF
First Trust Health Care AlphaDEX Fund	Invesco DWA SmallCap Momentum ETF
First Trust Horizon Managed Volatility Dom. ETF	Invesco DWA Technology Momentum ETF
First Trust Industrials/Producer Durables AlphaDEX	Invesco DWA Utilities Momentum ETF
First Trust Large Cap Core AlphaDEX Fund	Invesco Dynamic Biotechnology & Genome ETF
First Trust Large Cap Growth AlphaDEX Fund	Invesco Dynamic Building & Construction ETF
First Trust Large Cap Value AlphaDEX Fund	Invesco Dynamic Large Cap Growth ETF
First Trust Materials AlphaDEX Fund	Invesco Dynamic Large Cap Value ETF
First Trust Mid Cap Core AlphaDEX Fund	Invesco Dynamic Market ETF
First Trust Mid Cap Growth AlphaDEX Fund	Invesco Dynamic Pharmaceuticals ETF
First Trust Morningstar Dividend Leaders Index	Invesco Dynamic Semiconductors ETF
First Trust MultiCap Growth AlphaDEX Fund	Invesco Dynamic Software ETF
First Trust NASDAQ ABA Community Bank Index	Invesco FTSE RAFI US 1000 ETF
First Trust Nasdaq Bank ETF	Invesco FTSE RAFI US 1500 Small-Mid ETF
First Trust NASDAQ Clean Edge Green Energy I.	Invesco High Yield Equity Dividend Achievers ETF
First Trust NASDAQ Cybersecurity ETF	Invesco KBW Bank ETF

Invesco KBW High Dividend Yield Financial ETF	iShares Morningstar Large-Cap Value ETF
Invesco KBW Property & Casualty ETF	iShares Morningstar Mid-Cap ETF
Invesco Nasdaq Internet ETF	iShares Morningstar Mid-Cap Growth ETF
Invesco QQQ Trust Series 1	iShares Morningstar Mid-Cap Value ETF
Invesco RAFI Strategic US ETF	iShares Morningstar Small-Cap Value ETF
Invesco Russell 1000 Dynamic Multifactor ETF	iShares Morningstar Small-Cap ETF
Invesco Russell 1000 Equal Weight ETF	iShares Morningstar Small-Cap Growth ETF
Invesco S&P 500 BuyWrite ETF	iShares Mortgage Real Estate ETF
Invesco S&P 500 Equal Weight Consumer S. ETF	iShares MSCI USA Equal Weighted ETF
Invesco S&P 500 Equal Weight Consumer	iShares MSCI USA ESG Select ETF
Invesco S&P 500 Equal Weight ETF	iShares Nasdaq Biotechnology ETF
Invesco S&P 500 Equal Weight Energy ETF	iShares North American Natural Resources ETF
Invesco S&P 500 Equal Weight Financials ETF	iShares PHLX Semiconductor ETF
Invesco S&P 500 Equal Weight Health Care ETF	iShares Residential Real Estate ETF
Invesco S&P 500 Equal Weight Industrials ETF	iShares Russell 1000 ETF
Invesco S&P 500 Equal Weight Materials ETF	iShares Russell 1000 Growth ETF
Invesco S&P 500 Equal Weight Technology ETF	iShares Russell 1000 Value ETF
Invesco S&P 500 Equal Weight Utilities ETF	iShares Russell 2000 ETF
Invesco S&P 500 ex-Rate Sensitive Low Vol. ETF	iShares Russell 2000 Growth ETF
Invesco S&P 500 High Beta ETF	iShares Russell 2000 Value ETF
Invesco S&P 500 High Dividend Low Vol. ETF	iShares Russell 3000 ETF
Invesco S&P 500 Low Volatility ETF	iShares Russell Mid-Cap ETF
Invesco S&P 500 Pure Growth ETF	iShares Russell Mid-Cap Growth ETF
Invesco S&P 500 Pure Value ETF	iShares Russell Mid-Cap Value ETF
Invesco S&P 500 Quality ETF	iShares Russell Top 200 ETF
Invesco S&P 500 Revenue ETF	iShares Russell Top 200 Growth ETF
Invesco S&P 500 Top 50 ETF	iShares Russell Top 200 Value ETF
Invesco S&P MidCap 400 Pure Growth ETF	iShares S&P 100 ETF
Invesco S&P MidCap 400 Pure Value ETF	iShares S&P 500 Growth ETF
Invesco S&P MidCap Low Volatility ETF	iShares S&P 500 Value ETF
Invesco S&P Smallcap 600 Revenue ETF	iShares S&P Mid-Cap 400 Growth ETF
Invesco S&P SmallCap Health Care ETF	iShares S&P Mid-Cap 400 Value ETF
Invesco S&P Ultra Dividend Revenue ETF	iShares S&P Small-Cap 600 Growth ETF
Invesco Water Resources ETF	iShares S&P Small-Cap 600 Value ETF
Invesco WilderHill Clean Energy ETF	iShares Select Dividend ETF
Invesco Zacks Mid-Cap ETF	iShares Transportation Average ETF
Invesco DWA Consumer Staples Momentum ETF	iShares Trust - iShares MSCI KLD 400 Social ETF
Invesco S&P Midcap 400 Revenue ETF	iShares Trust iShares ESG MSCI USA ETF
IQ Chaikin US Large Cap ETF	iShares U.S. Basic Materials ETF
iShares Cohen & Steers REIT ETF	iShares U.S. Broker-Dealers & Securities Exch.
iShares Core Dividend Growth ETF	iShares U.S. Consumer Services ETF
iShares Core High Dividend ETF	iShares U.S. Energy ETF
iShares Core S&P 500 ETF	iShares U.S. Financial Services ETF
iShares Core S&P Mid-Cap ETF	iShares U.S. Healthcare ETF
iShares Core S&P Small-Cap ETF	iShares U.S. Healthcare Providers ETF
iShares Core S&P Total US Stock Market ETF	iShares U.S. Home Construction ETF
iShares Core S&P U.S. Growth ETF	iShares U.S. Industrials ETF
iShares Core S&P U.S. Value ETF	iShares U.S. Medical Devices ETF
iShares Core US REIT ETF	iShares U.S. Oil & Gas Exploration & Prod. ETF
iShares Dow Jones U.S. ETF	iShares U.S. Real Estate ETF
iShares Edge MSCI Min Vol USA ETF	iShares US Aerospace & Defense ETF
iShares Edge MSCI Min Vol USA Small Cap ETF	iShares US Consumer Goods ETF
iShares Edge MSCI Multifactor USA ETF	iShares US Financials ETF
iShares Edge MSCI Multifactor USA Small-Cap	iShares US Pharmaceuticals ETF
iShares Edge MSCI USA Momentum Factor ETF	iShares US Regional Banks ETF
iShares Edge MSCI USA Quality Factor ETF	iShares US Technology ETF
iShares Edge MSCI USA Size Factor ETF	iShares US Telecommunications ETF
iShares Edge MSCI USA Value Factor ETF	iShares US Utilities ETF
iShares ESG MSCI USA Small-Cap	Janus Detroit Street Trust Janus Henderson
iShares Expanded Tech Sector ETF	John Hancock Multi-Factor Large Cap ETF
iShares Expanded Tech-Software Sector ETF	John Hancock Multi-Factor Mid Cap ETF
iShares Micro-Cap ETF	John Hancock Multifactor Small Cap ETF
iShares Morningstar Large-Cap Growth ETF	JPMorgan Diversified Return US Equity ETF
iShares Morningstar Large-Cap ETF	JPMorgan Diversified Return US Mid Cap Equity

JPMorgan Diversified Return US Small Cap Equity	SPDR S&P MidCap 400 ETF Trust
JPMorgan US Minimum Volatility ETF	SPDR S&P Oil & Gas Equipment & Services ETF
JPMorgan US Quality Factor ETF	SPDR S&P Oil & Gas Exploration & Production
Legg Mason Low Volatility High Dividend ETF	SPDR S&P Pharmaceuticals ETF
Materials Select Sector SPDR Fund	SPDR S&P Regional Banking ETF
Motley Fool 100 Index ETF	SPDR S&P Retail ETF
Nationwide Maximum Diversification US Core Equity	SPDR S&P Semiconductor ETF
Nationwide Risk-Based US Equity ETF	SPDR S&P Software & Services ETF
Nuveen ESG Large-Cap Growth ETF	SPDR S&P Transportation ETF
Nuveen ESG Large-Cap Value ETF	SPDR SSGA Gender Diversity Index ETF
Nuveen ESG Small-Cap ETF	SPDR SSGA US Large Cap Low Volatility Index
O'Shares US Small-Cap Quality Dividend ETF	Technology Select Sector SPDR Fund
O'Shares US Quality Dividend ETF	TrimTabs All Cap US Free-Cash-Flow ETF
Pacer CFRA-Stovall Equal Weight Seasonal Rot.	US Diversified Real Estate ETF
Pacer US Cash Cows 100 ETF	Utilities Select Sector SPDR Fund
PIMCO RAFI Dynamic Multi-Factor U.S. Equity	VanEck Vectors Biotech ETF
Principal US Mega-Cap Multi-Factor Index ETF	VanEck Vectors Morningstar Wide Moat ETF
ProShares Large Cap Core Plus	VanEck Vectors Oil Services ETF
ProShares S&P 500 Dividend Aristocrats ETF	VanEck Vectors Pharmaceutical ETF
Proshares S&P Midcap 400 Dividend Aristocrats	VanEck Vectors Retail ETF
Real Estate Select Sector SPDR Fund	VanEck Vectors Semiconductor ETF
RiverFront Dynamic US Dividend Advantage ETF	Vanguard Communication Services ETF
RiverFront Dynamic US Flex-Cap ETF	Vanguard Consumer Discretionary ETF
Schwab 1000 Index ETF	Vanguard Consumer Staples ETF
Schwab Fundamental U.S. Broad Market Index ETF	Vanguard Dividend Appreciation ETF
Schwab Fundamental U.S. Large Company Index	Vanguard Energy ETF
Schwab Fundamental U.S. Small Company Index	Vanguard ESG US Stock ETF
Schwab U.S. Large-Cap Growth ETF	Vanguard Extended Market ETF
Schwab U.S. Large-Cap Value ETF	Vanguard Financials ETF
Schwab U.S. Mid-Cap ETF	Vanguard Growth ETF
Schwab US Broad Market ETF	Vanguard Health Care ETF
Schwab US Dividend Equity ETF	Vanguard High Dividend Yield ETF
Schwab US Large-Cap ETF	Vanguard Industrials ETF
Schwab US Small-Cap ETF	Vanguard Information Technology ETF
SPDR Dow Jones Industrial Average ETF Trust	Vanguard Large-Cap ETF
SPDR Dow Jones REIT ETF	Vanguard Materials ETF
SPDR MSCI USA StrategicFactors ETF	Vanguard Mega Cap ETF
SPDR NYSE Technology ETF	Vanguard Mega Cap Growth ETF
SPDR Portfolio Large Cap ETF	Vanguard Mega Cap Value ETF
SPDR Portfolio Mid Cap ETF	Vanguard Mid-Cap ETF
SPDR Portfolio S&P 500 Growth ETF	Vanguard Mid-Cap Growth ETF
SPDR Portfolio S&P 500 High Dividend ETF	Vanguard Mid-Cap Value ETF
SPDR Portfolio S&P 500 Value ETF	Vanguard Real Estate ETF
SPDR Portfolio Small Cap ETF	Vanguard Russell 1000
SPDR Portfolio Total Stock Market ETF	Vanguard Russell 1000 Growth ETF
SPDR Russell 1000 Low Volatility Focus ETF	Vanguard Russell 1000 Value
SPDR Russell 1000 Momentum Focus ETF	Vanguard Russell 2000 ETF
SPDR Russell 1000 Yield Focus ETF	Vanguard Russell 2000 Growth
SPDR S&P 400 Mid Cap Value ETF	Vanguard Russell 2000 Value
SPDR S&P 400 Mid Cap Growth ETF	Vanguard Russell 3000
SPDR S&P 500 ETF Trust	Vanguard S&P 500 ETF
SPDR S&P 500 Fossil Fuel Reserves Free ETF	Vanguard S&P 500 Growth ETF
SPDR S&P 600 Small Cap Growth ETF	Vanguard S&P 500 Value ETF
SPDR S&P 600 Small Cap ETF	Vanguard S&P Mid-Cap 400 ETF
SPDR S&P 600 Small Cap Value ETF	Vanguard S&P Mid-Cap 400 Growth ETF
SPDR S&P Aerospace & Defense ETF	Vanguard S&P Mid-Cap 400 Value ETF
SPDR S&P Bank ETF	Vanguard Small-Cap Value ETF
SPDR S&P Biotech ETF	Vanguard Total Stock Market ETF
SPDR S&P Dividend ETF	Vanguard U.S. Minimum Volatility ETF
SPDR S&P Health Care Equipment ETF	Vanguard US Multifactor ETF
SPDR S&P Health Care Services ETF	Vanguard Utilities ETF
SPDR S&P Homebuilders ETF	Vanguard Value ETF
SPDR S&P Insurance ETF	VictoryShares US Large Cap High Div Volatility
SPDR S&P Metals & Mining ETF	VictoryShares US Multi-Factor Minimum Volatility

VictoryShares USAA MSCI USA Small Cap Value	WisdomTree US Dividend ex-Financials Fund
VictoryShares USAA MSCI USA Value Momentum	WisdomTree US High Dividend Fund
Victoryshares Dividend Accelerator ETF	WisdomTree US LargeCap Dividend Fund
VictoryShares US 500 Enhanced Volatility Wtd ETF	WisdomTree US LargeCap Fund
VictoryShares US 500 Volatility Wtd ETF	WisdomTree US MidCap Dividend Fund
VictoryShares US EQ Income Enhanced Volatility	WisdomTree US MidCap Fund
Vident Core US Equity ETF	WisdomTree US SmallCap Dividend Fund
WisdomTree US Multifactor Fund	WisdomTree US SmallCap Fund
WisdomTree U.S. Quality Dividend Growth Fund	WisdomTree US Total Dividend Fund
WisdomTree U.S. SmallCap Quality Div. Growth	Xtrackers Russell 1000 Comprehensive Factor ETF

A.4 Summary Statistics – Before Adjustment for Gaps

Full sample, before adjustment for gaps

The table reports summary statistics for the variables used in this thesis, before adjustment of gaps. The summary statistics is reported in a monthly basis for S&P 500 stocks held by the ETFs in our sample. The sample cover the period between January 2008 to December 2018.

	N	Mean	SD	Median	Min	Max
Daily volatility (%)	59,500	1.606	1.190	1.293	0.000	29.312
ETF ownership	59,500	0.044	0.021	0.040	0.000	0.217
log (MCAP ($m))	59,500	9.666	1.099	9.534	5.622	13.910
Inverse Price	59,500	0.031	0.033	0.022	0.001	0.826
Amihud (2002)	59,500	0.000	0.000	0.000	0.000	0.012
Bid-Ask spread (%)	59,500	0.052	0.128	0.026	-1.550	11.765
Book-to-Market	59,500	0.215	0.503	0.147	-4.085	29.128
Gross Profitability	59,500	0.074	0.058	0.065	-0.493	0.572
Past 12-month returns	59,500	0.108	0.369	0.091	-0.991	12.174

A.5 Correlation Matrix - Before Adjustment for Gaps

Correlation Matrix – Before Adjustment for Gaps

The table reports the correlation matrix for the variables used in this thesis for the S&P 500 stocks held by the ETFs in our sample. The sample include the variables before the adjustment for gaps, and it covers the period between January 2008 to December 2018.

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Daily volatility (%)	(1)	1.000								
ETF ownership (%)	(2)	-0.187	1.000							
log (MCAP ($m))	(3)	-0.320	-0.030	1.000						
Inverse Price	(4)	0.425	-0.168	-0.380	1.000					
Amihud (2002)	(5)	0.477	-0.091	-0.482	0.441	1.000				
Bid-Ask spread (%)	(6)	0.415	-0.156	-0.185	0.316	0.348	1.000			
Book-to-Market	(7)	0.253	-0.025	-0.101	0.219	0.130	0.142	1.000		
Gross Profitability	(8)	-0.076	-0.076	-0.011	-0.103	-0.019	-0.044	-0.281	1.000	
Past 12-month return	(9)	-0.283	0.034	0.147	-0.211	-0.174	-0.160	-0.128	0.095	1.000

A.6 ETF Summary Statistics by Year

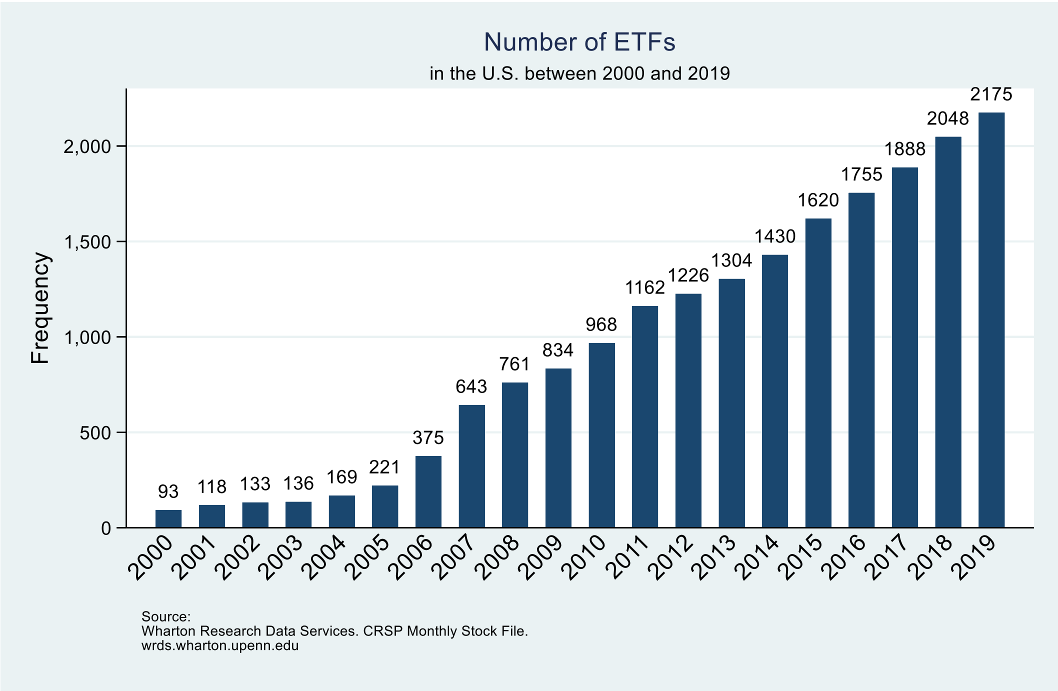
ETF Summary Statistics by Year

The table reports summary statistics for ETF ownership each year. The summary statistics is reported in a monthly basis for S&P 500 stocks held by the ETFs in our sample. The sample cover the period between January 2008 to December 2018.

	N	Mean	SD	Median	Min	Max
2008	5,459	0.023	0.010	0.021	0.000	0.106
2009	5,475	0.029	0.012	0.028	0.000	0.108
2010	5,465	0.027	0.009	0.025	0.000	0.107
2011	5,439	0.032	0.012	0.031	0.002	0.145
2012	5,372	0.036	0.014	0.034	0.003	0.177
2013	5,370	0.042	0.015	0.050	0.000	0.206
2014	5,306	0.047	0.015	0.045	0.006	0.163
2015	5,334	0.052	0.015	0.050	0.000	0.169
2016	5,377	0.058	0.016	0.055	0.000	0.188
2017	5,327	0.067	0.017	0.065	0.012	0.187
2018	5,396	0.073	0.018	0.071	0.030	0.217
2008-2018	59,320	0.044	0.021	0.040	0.000	0.217

A.7 Number of ETFs

This figure presents the number of ETFs in the U.S. from the period 2000 to 2019.



A.8 Detailed description of data collection for control variables

To obtain the daily and monthly values for the control variables, we use two different unique identifiers to identify the stocks held by the ETFs in our sample. When we collect data from the CRSP database, we use *permno* for each security to retrieve the information needed. When we collect data from the Compustat database, we use *gvkey* (The Global Company Key), which is a unique six-digit number assigned to each company in Compustat.

The first control variable of interest is the LMCAP, which is the natural logarithm of the stock market capitalization. It is natural to include this as a control variable since the S&P 500 is a capitalization-weighted index, which means that it assigns a higher weight the higher the market capitalization. Our sample of ETFs contain 17 funds that explicitly mention equal weighting, this can be concerning because Equation 4 relies on the weights in the numerator to grow at the same pace as market capitalization in the denominator (Ben-David et al., 2018a). If they do not, a spurious link could exist between ETF ownership and volatility, because of the correlation between stock size and volatility. Thus, including LMCAP controls for the issues that might be related to weighting schemes. Our sample consist of stocks that vary widely in size, from the lowest, $276 million (American Capital LTD), to the highest, $1.1 trillion (Apple Inc.). Due to this vast variation in firm size, we log the market capitalization to narrow this range and to ensure normality. The shares outstanding and closing price for each stock is collected from CRSP Monthly Stock File.

The second stock characteristic is the inverse price, which is used to control for the stock size influence on the volatility. There is a considerable difference between the share prices in our sample, from the lowest price at $1.21 (Genworth Financial Inc) to the highest price at $2,178 (Booking Holdings Inc). Inverse price is calculated using data from CRSP Monthly Stock File. The variable price alternate is used as the price value, which is an alternate monthly price derived from daily prices and contains the last non-missing price in the month. Further, we want to control for liquidity, which is measured by the Amihud (2002) illiquidity measure of price impact, and the bid-ask spread. We download the closing price, open price and share volume from CRSP Daily Stock File database to calculate the Amihud ratio, while we download the closing ask and closing bid prices from CRSP Monthly Stock File database to calculate the bid-ask spread.

Moreover, we include three standard predictor of returns that might also relate to volatility: book-to-market ratio, gross profitability and past 12-month returns (Ben-David et al., 2018a). The fifth characteristic of interest is the Book-to-Market ratio (BTM) defined as a company's book value divided by its market value. We use Compustat Fundamentals Quarterly database to download the data to calculate book-to-market. The collected data is reported quarterly, consequently we forward fill the two missing months in each quarter to construct monthly data. We find the forward filling appropriate since we assume that each companies' financial statements are stable between two reporting quarters. The sixth stock characteristic is gross profitability, presented by Novy-Marx (2013), and it is defined as the gross profit scaled by total assets. The data to calculate this variable is downloaded from Compustat Fundamentals Quarterly database, and it is reported quarterly. To avoid any predictions of future profits, we backward fill the two missing months in each quarter. The last stock characteristic we include is the past 12-month returns (P12MRET). Since it is well documented that there is a positive relation between stock return volatility and trading volume (Bae, Chan & Ng, 2004), we include the past 12-month returns to control for the effect trading volume has on stock volatility. The closing price and open price are used to calculate this variable is collected from CRSP Daily Stock File.

A.9 Summary Statistics for Control Variables

Control Variables Summary Statistics

The table reports summary statistics for the control variables used in this thesis. The summary statistics is reported in a monthly basis for S&P 500 stocks held by the ETFs in our sample. The sample cover the period between January 2008 to December 2018.

	N	Mean	SD	Median	Min	Max
log (MCAP ($m))	59,320	9.668	1.099	9.536	5.622	13.910
Inverse Price	59,320	0.031	0.032	0.022	0.001	0.826
Amihud (2002)	59,320	0.000	0.000	0.000	0.000	0.012
Bid-Ask spread (%)	59,320	0.052	0.128	0.026	-1.550	11.765
Book-to-Market	59,320	0.216	0.504	0.148	-4.085	29.128
Gross Profitability	59,320	0.074	0.058	0.065	-0.493	0.572
Past 12-month returns	59,320	0.107	0.367	0.091	-0.991	12.174

A.10 Correlation Matrix – Full Sample

Correlation Matrix – Full Sample

The table reports the correlation matrix for the variables used in this thesis for the S&P 500 stocks held by the ETFs in our sample.

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Daily volatility (%)	(1)	1.000								
ETF ownership (%)	(2)	-0.187	1.000							
log (MCAP ($m))	(3)	-0.318	-0.032	1.000						
Inverse Price	(4)	0.427	-0.169	-0.379	1.000					
Amihud (2002)	(5)	0.477	-0.091	-0.482	0.447	1.000				
Bid-Ask spread (%)	(6)	0.414	-0.155	-0.183	0.313	0.347	1.000			
Book-to-Market	(7)	0.254	-0.025	-0.101	0.228	0.130	0.143	1.000		
Gross Profitability	(8)	-0.078	-0.075	-0.009	-0.112	-0.019	-0.045	-0.281	1.000	
Past 12-month return	(9)	-0.286	0.032	0.147	-0.214	-0.173	-0.160	-0.128	0.094	1.000

A.11 Hausman (1978) specification test for fixed vs. random effects

```
. hausman fixed random, sigmamore

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |     fixed        random       Difference          S.E.
-------------+----------------------------------------------------------------
    zetfownp |    .0130958     .0041915        .0089043        .0008489
       lmcap |
         L1. |   -.3139471    -.2572814       -.0566657        .0048555
          ip |
         L1. |    3.836631     4.485702       -.6490709        .0768368
       illiq |
         L1. |    793.3924      777.217         16.1754        3.040689
   baspreadp |
         L1. |    1.871264     1.893148       -.0218845        .0017642
         btm |
         L1. |    .1081986     .1300558       -.0218573        .0025514
     gprofit |
         L1. |    .4629885     .2620413        .2009473        .0380668
     p12mret |
         L1. |   -.3394107     -.346685        .0072743        .0010866
------------------------------------------------------------------------------
                           b = consistent under Ho and Ha; obtained from xtreg
            B = inconsistent under Ha, efficient under Ho; obtained from xtreg

    Test:  Ho:  difference in coefficients not systematic

                  chi2(8) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =      247.69
                Prob>chi2 =      0.0000
```

A.12 Wald test for time-fixed effects

H_0: Coefficients for all months are jointly equal to zero

H_A: Coefficients for all months are not jointly equal to zero

```
       F(130, 57853) =   406.44
            Prob > F =    0.0000
```

A.13 OLS Regression with Realized Volatility

OLS Regression – Realized Volatility

The table reports the OLS estimates from regressions of realized volatility on ETF ownership and control variables with a monthly lag. The sample consists of S&P 500 stocks, and it covers the period January 2008 to December 2018. The frequency of observations is monthly, and volatility is estimated using the daily returns within the month. The control variables include logged market capitalization, the lagged inverse price, the lagged Amihud (2002), the lagged bid-ask spread, the lagged book-to-market, lagged gross profitability (Novy-Marx, 2013), and lagged past 12-month returns. Regression 3 includes three lags of the dependent variable. The dependent variable and ETF ownership variable are standardized, and standard errors are double-clustered at the month and stock levels. The t-statistics are shown in parentheses, and significance at 10%, 5%, and 1% is indicated by ***, ** and *, respectively.

Dependent variable:	$RDVOL_{i,t}$		
Sample:	Full Sample		
Regression (#)	(1)	(2)	(3)
$ETFOWN_{i,t}$	0.124***	0.124***	0.050***
	(4.62)	(4.49)	(3.47)
$LMCAP_{i,t-1}$	-0.103***	-0.104***	-0.045**
	(-2.69)	(-2.62)	(-2.00)
$IP_{i,t-1}$	4.560***	4.636***	1.311
	(3.93)	(3.79)	(1.52)
$ILLIQ_{i,t-1}$	403.6***	398.2***	83.90
	(4.26)	(4.18)	(1.50)
$BASPRD_{i,t-1}$	0.492**	0.521***	0.211
	(2.55)	(2.60)	(1.50)
$BTM_{i,t-1}$	0.174***	0.173***	0.045
	(3.61)	(3.48)	(1.14)
$GPROFIT_{i,t-1}$	-0.662**	-0.640**	-0.128
	(-2.38)	(-2.28)	(-0.81)
$P12MRET_{i,t-1}$	-0.076**	-0.083**	-0.042*
	(-2.02)	(-2.01)	(-1.69)
$DVOL_{i,t-1}$			0.349***
			(9.01)
$DVOL_{i,t-2}$			0.169***
			(3.32)
$DVOL_{i,t-3}$			0.122***
			(5.04)
Month fixed effects	Yes	Yes	Yes
Stock fixed effects	Yes	Yes	Yes
Standard Error	Clustered	Clustered	Clustered
Observations	58,655	57,325	57,325
R^2	0.722	0.722	0.795

A.14 Summary Statistics for Realized Volatility

Realized Volatility Statistics

The table reports summary statistics for realized volatility used in the robustness test. The summary statistics is reported in a monthly basis for S&P 500 stocks held by the ETFs in our sample. The sample cover the period between January 2008 to December 2018.

	N	Mean	SD	Median	Min	Max
Realized volatility (%)	59,320	7.339	5.432	5.915	0.169	135.527

A.15 Correlation Matrix with Realized Volatility

Correlation Matrix – with Realized Volatility

The table reports the correlation matrix for the variables used in this thesis for the S&P 500 stocks held by the ETFs, where realized volatility is the dependent variable. The sample covers the period between January 2008 to December 2018.

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Realized volatility (%)	(1)	1.000								
ETF ownership (%)	(2)	-0.187	1.000							
log (MCAP ($m))	(3)	-0.319	-0.316	1.000						
Inverse Price	(4)	0.428	-0.169	-0.379	1.000					
Amihud (2002)	(5)	0.477	-0.091	-0.482	0.447	1.000				
Bid-Ask spread (%)	(6)	0.418	-0.155	-0.183	0.313	0.347	1.000			
Book-to-Market	(7)	0.256	-0.025	-0.101	0.228	0.130	0.143	1.000		
Gross Profitability	(8)	-0.078	-0.075	-0.009	-0.112	-0.019	-0.045	-0.281	1.000	
Past 12-month return	(9)	-0.287	0.032	0.147	-0.214	-0.173	-0.160	-0.128	0.094	1.000

A.16 Summary Statistics for subsample "in crisis"

Summary Statistics – "in crisis"

The table reports summary statistics for the variables used in this thesis. The summary statistics is reported in a monthly basis for S&P 500 stocks held by the ETFs in our sample. The sample cover the period between January 2008 to June 2009.

	N	Mean	SD	Median	Min	Max
Daily volatility (%)	8,208	3.190	2.124	2.593	0.000	29.312
ETF ownership	8,208	0.025	0.012	0.023	0.000	0.108
log (MCAP ($m))	8,208	9.148	1.145	9.034	5.622	13.115
Inverse Price	8,208	0.050	0.053	0.035	0.000	0.826
Amihud (2002)	8,208	0.000	0.000	0.000	0.000	0.012
Bid-Ask spread (%)	8,208	0.002	0.003	0.001	-0.016	0.118
Book-to-Market	8,208	0.305	0.751	0.195	-4.085	19.622
Gross Profitability	8,208	0.075	0.062	0.070	-0.493	0.388
Past 12-month returns	8,208	-0.190	-0.414	-0.227	-0.991	9.695

A.17 Summary Statistics for subsample "after crisis"

Summary Statistics – "after crisis"

The table reports summary statistics for the variables used in this thesis. The summary statistics is reported in a monthly basis for S&P 500 stocks held by the ETFs in our sample. The sample cover the period between July 2009 to December 2018.

	N	Mean	SD	Median	Min	Max
Daily volatility (%)	51,112	1.349	0.669	1.196	0.000	13.089
ETF ownership	51,112	0.047	0.021	0.044	0.000	0.217
log (MCAP ($m))	51,112	9,751	1.068	9.595	6.579	13.910
Inverse Price	51,112	0.027	0.026	0.021	0.000	0.472
Amihud (2002)	51,112	0.000	0.000	0.000	0.000	0.004
Bid-Ask spread (%)	51,112	0.000	0.000	0.000	-0.002	0.012
Book-to-Market	51,112	0.201	0.450	0.141	-3.057	29.128
Gross Profitability	51,112	0.073	0.057	0.064	-0.299	0.572
Past 12-month return	51,112	0.155	0.335	0.126	-0.975	12.174

A.18 Correlation Matrix for subsample "In crisis"

Correlation Matrix – " in crisis"

The table reports the correlation matrix for the variables used in this thesis for the S&P 500 stocks held by the ETFs in the period between January 2008 to June 2009.

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Daily volatility (%)	(1)	1.000								
ETF ownership (%)	(2)	0.204	1.000							
log (MCAP ($m))	(3)	-0.300	-0.312	1.000						
Inverse Price	(4)	0.421	0.186	-0.395	1.000					
Amihud (2002)	(5)	0.425	0.146	-0.534	0.482	1.000				
Bid-Ask spread (%)	(6)	0.301	-0.035	-0.134	0.195	0.253	1.000			
Book-to-Market	(7)	0.386	0.178	-0.162	0.284	0.183	0.191	1.000		
Gross Profitability	(8)	-0.249	-0.147	0.037	-0.122	-0.045	-0.091	-0.246	1.000	
Past 12-month return	(9)	-0.288	-0.194	0.194	-0.265	-0.114	-0.073	-0.215	0.154	1.000

A.19 Correlation Matrix for subsample "after crisis"

Correlation Matrix – " after crisis"

The table reports the correlation matrix for the variables used in this thesis for the S&P 500 stocks held by the ETFs in the period July 2009 to December 2018.

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Daily volatility (%)	(1)	1.000								
ETF ownership (%)	(2)	-0.056	1.000							
log (MCAP ($m))	(3)	-0.297	-0.089	1.000						
Inverse Price	(4)	0.316	-0.156	-0.353	1.000					
Amihud (2002)	(5)	0.324	-0.021	-0.646	0.372	1.000				
Bid-Ask spread (%)	(6)	0.307	-0.147	-0.301	0.629	0.396	1.000			
Book-to-Market	(7)	0.170	-0.027	-0.071	0.182	0.061	0.107	1.000		
Gross Profitability	(8)	-0.038	-0.085	-0.015	-0.124	-0.019	-0.074	-0.299	1.000	
Past 12-month return	(9)	-0.073	-0.085	0.070	-0.104	-0.091	-0.066	-0.079	0.093	1.000

A.20 OLS Regression – NBER (n.d.) Sample Split by Recession

OLS Regression – NBER (n.d.) Sample Split by Recession

The table reports the OLS estimates from regressions of daily volatility on ETF ownership and control variables with a monthly time lag. The subsamples consist of S&P 500 stocks and covers the period from January 2008 to June 2009 ("in crisis"), and from July 2009 to December 2018 ("after crisis"). The frequency of observations is monthly, and volatility is estimated using the daily returns within the month. The control variables include logged market capitalization, the lagged inverse price, the lagged Amihud (2002), the lagged bid-ask spread, the lagged book-to-market, lagged gross profitability (Novy-Marx, 2013), and lagged past 12-month returns. Regression 3 includes three lags of the dependent variable. The dependent variable and ETF ownership variable are standardized using the standard deviation corresponding to the relevant sample. Note that for the "in crisis", the standard errors are clustered at only the stock level, while in "after crisis" they are clustered at both the stock and month level. The t-statistics are shown in parentheses, and significance at 10%, 5%, and 1% is indicated by ***, ** and *, respectively.

Dependent variable:	$DVOL_{i,t}$					
Sample:	"in crisis"			"after crisis"		
Regression (#)	(4)	(5)	(6)	(7)	(8)	(9)
$ETFOWN_{i,t}$	0.021	0.015	0.031	0.111***	0.107***	0.057***
	(0.66)	(0.42)	(0.88)	(4.91)	(4.14)	(4.11)
$LMCAP_{i,t-1}$	-0.805***	-0.810***	-0.709***	-0.141**	-0.125**	-0.032
	(-10.41)	(-9.78)	(-10.43)	(-2.55)	(-2.14)	(-0.93)
$IP_{i,t-1}$	1.467	1.296	1.072	9.109***	10.14***	5.708***
	(1.54)	(1.28)	(1.22)	(5.12)	(5.58)	(5.92)
$ILLIQ_{i,t-1}$	33.73	28.44	-15.23	798.1***	1014.4***	299.8**
	(1.06)	(0.90)	(-0.55)	(5.23)	(4.86)	(2.17)
$BASPRD_{i,t-1}$	0.097***	0.093**	0.066*	1.478***	1.275***	0.467**
	(2.63)	(2.46)	(1.82)	(4.84)	(3.62)	(2.57)
$BTM_{i,t-1}$	-0.032	-0.053	-0.078	0.312***	0.056	0.067
	(-0.40)	(-0.65)	(-1.08)	(2.62)	(0.66)	(1.48)
$GPROFIT_{i,t-1}$	0.621**	0.733**	0.689***	-1.157***	-1.144**	-0.409
	(2.18)	(2.41)	(2.58)	(-2.72)	(-2.38)	(-1.33)
$P12MRET_{i,t-1}$	0.010	0.013	0.018	0.072**	0.070	0.036
	(0.33)	(0.24)	(0.42)	(2.51)	(1.62)	(1.18)
$DVOL_{i,t-1}$			0.172***			0.245***
			(5.76)			(13.09)
$DVOL_{i,t-2}$			0.040*			0.146***
			(1.76)			(10.67)
$DVOL_{i,t-3}$			-0.071***			0.148***
			(-3.66)			(11.49)
Month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Stock fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Standard Err.	Clustered (Stock)	Clustered (Stock)	Clustered (Stock)	Clustered	Clustered	Clustered
Observations	7,713	6,723	6,723	50,941	49,235	49,235
R^2	0.767	0.773	0.780	0.618	0.616	0.678