Faculty of Business and Economics – HEC Lausanne
University of Lausanne

Final thesis for the completion of the Master of Science in Finance (MScF)

Exchange-traded funds' expansion and their unintended effects over underlying stocks

Volatility, liquidity and informational efficiency

Grégoire Pichard

June 12, 2019

Supervisor :

Prof. Artem Neklyudov

Contents

List of Tables

List of Figures

1. Foreword

It is a moving event that the actual beginning of the research regarding this paper nearly coincides with the death of one of the fathers, if not the demiurge of that same research topic : Mr. John C.,"Jack", Bogle, who was born in 1929 and passed out on January 17th, 2019, best known as the founder of *The Vanguard Group* in 1975. He was already considered a legend in the investment world decades before his death and was duly praised by finance leaders and scholars, even by those who were far from sharing his views. It is interesting to notice that he was almost as famous as one of his contemporaries, namely Warren Buffett (born in 1930), chairman of conglomerate and investment company *Berkshire Hathaway*, for achieving success on a fundamentally opposite view of his role as an investor. Mr. Buffett has built an empire with active bets on companies that he considered, at the time he purchased them, to be undervalued and with a high book-to-market ratio, hence showing his unique ability to exploit what can be modelled as a small set of alternative risk premia, in activities which he has a deep understanding of. Recent research about his multi-decade long winning streak has shown his success is more systematic than genius stock-picking and relies on drivers that the academic literature considers systematic with low correlation to the market premium. Still, the conclusion is often that he is an overwhelmingly successful disciple of strict principles written in the 1930s by Graham and Dodd at the Columbia Business School (Graham and Dodd 1934). Since then, the research has caught up and shed light on the value risk factor for instance, which can explain a significant share (while not all) of Buffett's financial performance over the last 50 years. Mr. Buffett is the living proof that the investor can generate long term risk-adjusted performance above the market return thanks to a mix of information available to investors and a selection based on fundamental valuation.

On the other side, British weekly newspaper *The Economist*, has deemed Jack Bogle's curriculum as the founder of a business that

> has radically changed money management by being boring and cheap.

The worry that Bogle's ultimate success with widespread adoption of index funds leads to less liquidity and less research, has been expressed by Sam Zell, the billionaire chairman and founder of *Equity Group Investments*, a private investment firm. In an interview on *Bloomberg Television* in which, among various topics he was asked to react, was remembering Jack Bogle. He says that

> Bogle's concept and what he advocated was terrific as long as it did not get to be

> too big a percentage, and I think we are at that point now where there is significant risk.

Obviously, as any market actor and company representative or owner that will answer Bloomberg journalists' question, Sam Zell may have interests that do not match those of Vanguard's, not necessarily as a direct competitor. That issue is also raised by the hosts and Sam Zell answers that his group of firms' business is not "yet" strongly impacted by the rise of passive investing, but that

> everybody is pretty concerned about the percentage of ownership of New York Stock Exchange companies, particularly the REITs[1] owned by passive players. [...] If your percentage [in index funds] gets too high, then you institutionalize mediocrity.

Later he admits the long-run rise of passive investing may even cause a positive effect on capital flows to private equity, at the expense of active managers in public companies (concern with regard to hedge funds), so the overall effect on Zell's businesses is probably only clear to himself, as a manager of a private company.

Despite having developped the first index fund at time the idea was only nascent[2], Jack Bogle expressed harsh criticism of index-tracking ETFs, altough they present some tax advantages over open-end index funds. The so-called issues raised by these widespead and quick growth of their volume related to the use made by some investors rather than the very nature of ETF. Opposite to the idealized buy-and-hold (and most importantly but less often stated, "rebalance") passive portfolio, ETFs exhibit turnover rates in line with intraday trading. With the individual

[1] Real estate investment trusts own, manage and even in some cases finance real estate assets, for example commercial surfaces (offices), residential buildings as well as malls and hotels. Sam Zell's companies are especially active in real estate and another part of the interview to *Bloomberg* was about comments (strong words as he seems to be famous for that form of communication) regarding the trends in commercial real estate.

[2] As Jack Bogle himself wrote in a 2012 piece of the *Journal of Indexes* about his impact in the industry, the idea was probably not only his, but he was the first to achieve such a development, although as it was commonly viewed as a mere failure at that time. To quote his introduction, "ideas are a dime; implementation is everything" Among earlier attempts starting in the late 1960s, not always based on the S&P 500 Index, he recalls about modelling by *Wells Fargo Bank* for the pension fund of *Samsonite*, then by a mutual fund firm from... Boston (of course, as his own Master's thesis from 1951 was inspired by a *Fortune* magazine piece titled "Big Money in Boston"), the *American National Bank* in Chicago (later absorbed by *J.P. Morgan Chase*), *American Express* before its management stopped the project and finally an insurance and pension company for teachers advised by Milton Friedman. Mr. Bogle concludes : "In all of these forays of indexing: ideas A+; implementation F."

investor in mind during his whole career, which he served by cutting management fees and the fund *load* of his products, Jack Bogle noticed that they were tools enabling speculation. An abundant literature he authored condemned speculation as a net financial loss for fund investors and ultimately a serious lack of stewardship by the very industry he had been part of for more than six decades. While it is possible to track similar indices with traditional index funds and ETFs, the actual use of those securities differ to a great extent : while the former, in line with Bogle's, are generally long-term buy-and-hold investments, the latter, continuously tradeable, may serve for hedging, short-selling and thus for arbitrage purpose.

In this paper, we are especially curious to find, if not why, at least how this improved trading characteristic of ETFs may have changed the underlying securities' markets. Admittedly, studying an issue discussed by few people outside Jack Bogle himself means standing on the shoulder of a giant.

This recount of ongoing debate that flourish in the financial press only serves as an element of context. Additionally, it would have been a historical faux-pas not to mention Jack Bogle's impact in the passive investing world. Yet the fact that the relative size of index funds holdings in general and ETF holdings of stocks in particular is discussed outside scientific publications tends to show that the question deserves to be asked and is the rationale for further enquiry. This paper humbly seeks to contribute to this major debate.

2. Motivation

Remains to be developed : this is the proper introduction of the thesis, insisting on the reasons why the topic is relevant and a current issue in research : the rise of passive investing and alternative risk premia products availability for a broader segement of the market; recent notable evidence based on a track record of about 15 to 20 years of measurable ETF existence on the market; very important of volume on exchanges attributable to ETF raises the question of its impact; very visible benefits for ETF investors in terms of cost, intraday liquidity, less information about the disadvantages for the market as a whole; ETFs already linked with abrupt non-fundamental market movements, quickly reversed (flash crashes) : they are connected with concerns about high-frequency trading; fears expressed by some commentators if the market becomes more passive and the role of ETF continues to grow;

RESEARCH ABOUT EARLIER FORMS OF INDEX INVESTING (E.G. FUTURES) HAS YIELDED RESULTS THAT SHOULD BE UPDATED AND COMPARED.

DETAIL THE CHOICE OF VARIABLES OF INTEREST, DOCUMENTED EVIDENCE AND HOW THIS PAPER EXTENDS EXISTING RESEARCH BY 1. INVESTIGATING THREE SUB-TOPICS INTO A SINGLE DATASET AND 2. EXPANDING THE COVERAGE TO INTERNATIONAL (I.E. NON-U.S.) EQUITY.

2.1. Structure of the paper

The research questions imply a tranmission mechanism between two layers of trading, thus an overview of both the structure and the context is needed. First, focusing on the specific nature of ETFs and on distinctive characteristics of the ETF markets, Section 3 provides a brief institutional summary. Section 4 navigates along the time dimension in order to put the recent expansion in context. The review of literature in Section 5 has a narrow scope since the focus of this part is exclusively on recent published research about the impact of ETFs on stocks. Next, the data used are described in Section 6, before the more Section 7 explicits how the key variables are computed, what the best-known theoretical explanations for various controls are and which model specifications are estimated. This part draws heavily on multiple strands of the asset pricing literature. Logically, the outcome is disclosed and discussed in Section 8. Finally, Section 9 summarizes the main findings, draws conclusions regarding the research question, attempts both at showing how the current methodology could be extended and, admittedly a perilous exercise, tries discussing what the perspectives of ETF markets and their effects might be in the future.

3. Characteristics of Exchange-Traded Funds

The creation and more recently the significant ownership share of Exchange-Traded Funds in major stock markets (among other asset classes) yields several empirical statements. For about a decade, further investigation has started measuring the impact of ETFs and confronting hypotheses while theory has been following with models that embed the mechanisms observed in practice :

- how ETFs and mutual funds, as investment vehicles, can coexist in equilibrium, due to a trade-off between cost efficiency and exposure to liquidity shocks; more generally, the rise of ETFs coincides with debates about the effects of indexing (the rise of passive investment) and about the increasing concentration of institutional investors causing larger and more correlatied fund flows and subsequent comovements among securities.

- how participants cause both the dry-up and increased price movements of multiple securities held through ETFs during high volatility times, a behavior that propagates turmoil in an unpredictable way;
- how ETFs may also increase pricing efficiency by providing arbitrageurs a proxy in order to sell short some securities that are precisely subject to short-sale limitations.

These topics will be discussed in greater depth in the next section. First, in order to understand why ETFs have experienced such an outstanding growth[3] and raised such concerns, it is necessary to describe how this structure appeals to investors and why they were a timely innovation that matched deep trends in financial markets' research and practice from the late 20$^{\text{th}}$ century until now.

3.1. Goal

As it has been briefly exposed in the Foreword (Section 1), ETF are the most successful implementation of an objective initially aimed at and reached by institutional investors. This objective was, then, to design a contiuously, publicly tradable vehicle tracking a diversified value-weighted equity index. The first ETF issued on a U.S. exchange (back in 1993) was indeed the SPDR S&P 500 ETF, the archetype of the first products ever designed. Its net assets as of September 30, 2018, according to the trust prospectus, amount to nearly USD 280 billion, only USD 1.3 billion being held in cash. Several thousand products have appeared in the ETF category[4] as well as in the broader Exchange-Traded Products family. The assets under management amounted to more than USD 4.5 trillion. Although there is wide range investment objectives and sophisticated strategies, the fund with the highest absolute net inflows over the year (USD 30.2 billion) was the iShares Core S&P 500 ETF.

TBD : figure from ETFGI comparing recent inflows/outflows in ETFs and mutual funds

[3] . . . whether measured in the number of products issued, the value of assets under management, the share in exchanged volume or the variety of asset classes, sectors and invesment styles covered. . .

[4] At least 5400 ETFs were trading globally as of late 2017, according to research firm ETFGI. Source : https://www.marketwatch.com/story/efts-shattered-their-growth-records-in-2017-2017-12-11

ETFs differ from open-end index funds, sometimes called *traditional index funds*, in that they offer permanent liquidity throughout trading hours, they publish the fund's Net Asset Value at regular intervals and the creation/redemption of shares is the task of institutions entitled to exchange a certain in kind securities for a newly created batch of ETF shares or to do the opposite trade with the fund sponsor to redeem ETF shares. The close tracking of the index is made possible by the privilege that authorized participants, a designation essentially applied to market makers, have to perform an arbitrage between the NAV and the actual value of underlying securities on the exchange. (Ben-David, F. Franzoni, and Moussawi 2017)

Standardized blocks of shares created and redeemed by ETF sponsors are called creation units. In exchange for a basket of securities matching the current index composition and an amount in cash amounting to the value of dividends granted on stocks, the counterparty will receive a multiple of fixed number of shares, e.g. 50000 shares. The same process exist for in-kind redemption : in this case, the ETF hands back the basket of securities and erases a multiple of a creation unit. Actual delivery delays of assets vary according to regulation, allowing settlment within two trading days after the transaction date. In some situations, when participants can prove their failure to deliver is involuntary[5], trades may even be settled within three additional days. In the U.S.-incorporated ETFs, there is a clearinghouse processing transactions between authorized participants on one side and ETF issuers or their distributors on the other side; its name is the National Securities Clearing Corporation and it already was the major post-trade actor before ETFs existed. Once creations and redemptions are processed after the market close, the NSCC shares a portfolio composition file with all active corporations, based on information provided by the issuer. This file then serves as the list of every security and its relevant quantity to provide in order to receive a creation unit.

The primary market for creation units is an essential feature and it is considered by research as the channel allowing for shock transmission between ETFs and securities as well as between stocks that are commonly held. Commonality in liquidity is seriously supposed to araise because of the creation/redemption mechanism as will be discussed later in this paper.

3.2. Legal and regulatory framework

ETFs may have at least two legal forms, based on information relevant in the United States : they may be registered as regulated open-end investment companies – exactly the same as mutual

[5]The exact wording is :"their failure to deliver is the result of bona fide market making". (Ben-David, F. Franzoni, and Moussawi 2017)

funds, intuitively – or as unit investment trusts; for instance, the much-traded SPDR S&P 500 ETF (mnemonic code SPY), nowadays sponsored and administered through State Street Global Advisors, is a unit investment trust. This legal structure prevents the funds adopting it from engaging in securities lending, which is one of the main sources of revenues of ETF sponsors. The latter legal form, which does not, in principle, exclude active investment objectives, offers the redemption of shares at Net Asset Value to the owner, which will sometimes be called the primary market as opposed to the secondary market which enables buyers and sellers to trade shares independent from the trust advisor. Open-end funds and unit investment trusts are the two main legal structures used by ETFs focusing on stocks, which is a subset of the whole (expanding) ETF universe.

In terms of management, unit investment trusts differ from open-end investment companies since they do not have a board of directors; since the portfolio is assumed fixed and is especially straightforward if such an ETF engages in systematic physical replication, a board of directors would in essence seem superfluous. Even more anecdotally, unit investment trusts are required by law to have a fixed end date, however far away it is: for instance, the ultimate termination date of the SPY ETF is scheduled 125 years after its initial deposit, i.e. on January 22, 2118.

Since the majority of ETF, both in entities and in value terms, is registered and managed in the United States of America, more emphasis has been put on the structures allowed and regulated under the Investment Company Act of 1940. More generally, the legal form chosen by ETFs differs in every regulatory system[6], which unfortunately makes an exhaustive comparison of all eligible legal structures untractable for this study.

3.3. Market participants

So far, two essential actors of the primary market have been mentioned :

the ETF sponsor, or issuer is the firm managing the fund, which name is attached to the product because it decides which benchmark the ETF tracks and how the replication strategy is implemented. If the index is computed by a third party, the sponsor will benefit from

[6]European countries even seem to have at least kept their national regulatory frame, at least according to the names advertised by those structures. Harmonization nevertheless has already started through the so-called *Undertakings for the Collective Investment of Transferable Securities*, better known as the UCITS acronym which only labels compliant funds and not debt products such as notes, certificates, etc. Reportedly, as of 2017, three quarters of EU ordinary ETF investorshold UCITS funds. (Source : https://www.justeft.com/uk/news/etf/legal-structure-of-etfs-ucits.html [Consulted April 24, 2019])

tracking a well-known index since the demand exhibits a preference for widely used broad-based or sector-based indices. Replicating an index provided through another specialized company (for instance S&P, MSCI) comes at a cost, the license fee, which typically amounts for several basis points per year. The issuer is in charge of the SEC filing in the United States, before compliant documentation is published.They also appoint the otherparticipants presented in this subsection.

authorized participants (APs) are essential because they initiate share creations and redemptions with the fund, based on a formal agreement with the fund. One has to recall that the price of an ETF is not automatically and instantly determined through the name of the index it tracks; nor is it derived from the last known value of its assets, the NAV. Rather, ETFs are securities traded on an exchange as their name says, which mean their price stems from supply and demand. Therefore the market price of ETF shares may deviate from the underlying index; unlike closed-end funds, ETFs are known for the scheme allowing to correct this deviation. Precisely, APs' non-mandatory role is to adjust the supply of ETF shares by providing either securities portfolios corresponding to the track index in order to create new blocks of ETF shares (excess demand of ETF shares) or handing back blocks, the creation units, to the issuer in order to get in-kind reimbursement in the form of underlying securities (excess supply of ETF shares). They are not paid by ETF sponsors, not constrained to be constantly monitoring the net asset value (NAV) either: actually, authorized participants have to pay the sponsor a creation fee (which may be fixed, independent from traded value). Their incentive is the same as market makers, and some institutions may play several roles on an ETF market : they will engage in a creation process if demand for is in excess. In other terms, they will execute a profitable trade by being able to sell the created units above the NAV, regularly disseminated by the sponsor. Of course, such an arbitrage only takes place if the trade's costs, including transaction fees on individual stocks and ETF creation fees, do not offset the profit earned on the difference between the ETF price and its intraday NAV. Thus, authorized participants generally are large banks, broker-dealers and market makers.

Beside the two main actors and the clearing corporation, the proper daily trading and liquidity is made possible through :

custodians do not only hold the assets of the ETF on behalf of the fund issuer : they reconcile positions with the index rules, in case the product replicates an index from a third party (e.g. the S&P 500 index computed by *Standard & Poor's Dow Jones Indices LLC*), and more generally with ETF rules, which are made publicly available in the prospectus. Custodians, often banks, execute periodic rebalancings and adjust the portfolio when a corporate action (e.g. stock split, delisting, merger or acquisition) so demands.

market makers create liquidity in the secondary market by steadily providing quotes based on their calculation of the ETF's fair value. It is reported that a limited number of market markers trade each ETF, unlike (liquid) stocks for which a best bid and offer is computed out of various quotes. Nevertheless, there are exists a correlation between the volume traded, the spread on the ETF market and the number of market makers.

4. Chronological background

If one wants to understand how fundamental the shift to passive investment and ETFs is, it seems necessary to step back at least to the 1970s. Masses of workers in Western countries have accessed the financial markets during the strong post-war economic boom. Mutual funds were not invented in the second half of 20$^{\text{th}}$ century, rather were they formally regulated as collective investment schemes in the 1930s and more generally as investment companies with the 1940 Act in the U.S. Traces of the earliest closed-end funds are said to have been found in the Netherlands, first in 1774 in the form of a merchant's investment trust and then after the Congress of Vienna at the king's initiative. The scheme has then spread in Europe including Switzerland among early financial places with a 1849-established company called *Société civile genevoise d'emploi de fonds*. Modern open-end funds though more relevantly have a famous and long-lived ancestor with the *Wellington Fund*, established 1928. This company from Pennsylvania was famously led by the late Jack Bogle, before he founded *The Vanguard Group* which nowadays manages the Wellington fund.

The trend of institutionalization of investment really took on during after World War II and mutual funds gained a significant attraction power in the 1980s. Mutual funds have claimed to provide a performance relative to benchmarks that are reference indices, for example the market-capitalization-weighted stock index. On the other hand, as a consequence, passive investing has become more easily available for small, individual investors thanks to a new type open-end mutual fund : the index fund, first (successfully) implemented by Jack Bogle. ETFs appeared more than a decade later[7] and remained confidential as long as the trust into mutual funds was

[7] Charupat and Miu (2013) reports as common knowledge that the first ETF listed was the Toronto Stock

untouched. The narrative according to which fund managers only work with their shareholder's sole interest in mind suffered increasing doubt due to public scandals : in 2003, several dozen companies were found by regulators to have violated multiple customer diligence rules, being accused of late trading and market timing. Mutual fund prices, unlike ETFs, are computed once a day at close; late trading by informed traders allowed them to trade at the set price during the stock market "after-hours", instead of having to wait for the next session's fund price. Such certain profit opportunity is considered an unfair advantage granted by the fund to some traders, as is the market timing charge. Mutual funds investors, again unlike ETF investors, are generally prevented to trade back and forth beyond a certain frequence for cost reasons, since the processing of orders induces costs borne by all fund investors, and for cash management reasons too, since the cash balance needed to face frequent and potentially massive selling orders is larger. Violating their own rules, some funds arbitrarily allowed some privileged investors engage in market timing. The aim of such fund managers was reportedly to boast larger assets, which in turn positively influence their own fees. Scandals and the increased focus on managers' actual value added in exchange for their advisory fees following the 2008 financial crisis did not abruptly stop mutual funds' growth but slowed it at the benefit of Exchange-Traded Funds, considered by many as the most liquid and transparent vehicles for passive and alternative investment.

The advent of Exchange-Traded Funds takes place in an ongoing debate that have led modern financial economics at least since Fama's paper about the Efficient Market Hypothesis (hereafter EMH) around 1965. The 2013 Nobel Prize granted to Eugene Fama *and* Robert Shiller, EMH's main critic[8] has obviously acknowledged the secular importance of the debate but the discussion has certainly not been closed since then.

In a recent working paper called *The Active World of Passive Investing*, Easley et al. (2018) claim that the shift from mutual funds and individual securities to ETFs is not simply a shift from active to passive investing, that distinction being "antiquated". Popular statistical metrics such as the active share – which measures the absolute deviations of a portfolio holdings compared with a value-weighted benchmark – and the tracking error – i.e. the annualized standard deviation

Exchange Index Participation Units, a product that tracked the market-capitalization-weighted TSE 35 stock index. The ETF was launched in March 1990 and was replaced in 2000 with the iShares S&P/TSX 60 Index Fund as the benchmark on the Canadian stock market was changed.

[8] As a matter of completeness, let us not forget the third recipient, Lars Peter Hansen, who developped econometric methods used to test the EMH, namely the General Method of Moments (GMM) estimators. A summary of their literature and the fields they contributed most to is Economic Sciences Prize Committee of the Royal Swedish Academy of Sciences (2012)

of the daily difference between the portfolio and market returns – show that the decennial increase in ETFs share is actually fueled by active investing, including smart beta products whose "activeness" could be debated, because they essentially try to systematically capture risk premia justified by factors studied in the academic literature (most importantly, value and momentum factors). Nevertheless, a worldwide phenomenon has been at work leading to the mutual fund industry specializing towards passive investment and the ETF products, first designed to provide an efficient tracking of value-weighted stock indices, mutating to virtually any flavour of index, whether focusing on specific sizes, sectors, regions or factors.

In this paper, the focus is less on the mildly active aspect of ETF investing than on those products' impact on underlying assets. When agents shift from discretionary investing to index-tracking, the measure of success shifts from outperforming a benchmark to minimizing the difference between one's portfolio and a benchmark, and ETFs have shown that they are a convenient instruments for this latter goal. It has been said that, with the variety of products available, the actual decision is not to pick the right security but the right index. Such change is suspected to induce some correlation among securities, a claim that this paper will address after a summary of related research.

5. Review of literature

Different segments of research are relevant to this paper's questions, most of them, perhaps surprisingly, not considering ETFs specifically but studying the effects of indexing, derivatives, and institutional ownership.

Regarding the covariance of individual stock returns, a significant and positive impact of the contemporaneous share held by ETFs has been documented in several studies. An extensive *Journal of Finance* article that inspires this paper, Ben-David, F. A. Franzoni, and Moussawi (2018), brings consistent evidence of this impact by analyzing the shock propagation through arbitrage between underlying securities and funds, using a sample of U.S. stocks and ETFs. First of all, they show that ETF on average exhibit higher liquidity (measured in the form of a lower Amihud (2002) *illiquidity* ratio), a lower bid-ask spread and a higher turnover than the value-weighted baskets of securities they hold. This comparison method prevents an adverse selection bias due to ETF supposedly investing in the most liquid stock and being otherwise similar to their portfolio. This comparison is valid under the hypothesis that ETFs together hold a value-weighted basket of stocks, according to the traditional (and significant) use of ETFs.

Higher liquidity and turnover at the same time support the assumption of a clientele effect : as Amihud and Mendelson (1986) predicted in general and tested for NYSE stocks, investors with a shorter expected holding horizon have a preference for liquid securities and the relative bid-ask spread is a measure correlated with illiquidity. The focus on short term drives the investor's choice towards the most liquid stocks while buy-and-hold investors accept keeping less liquid assets since they do not expect and/or need to sell them immediately.

Ben-David, F. A. Franzoni, and Moussawi (2018) call their main hypothesis regarding a positive impact of ETF ownership over underlying stocks the *liquidity trading hypothesis*: volatility reflects the fact that the shocks occurring on asset prices are non-fundamental, i.e. not related with cash flows nor with any news about the company. Therefore, the price is expected to revert to its initial value, ceteris paribus, because the shock is purely driven by liquidity. This sort of shock propagation differs from a two alternative hypotheses:

the price-discovery hypothesis : permanent, i.e. fundamental, price adjustment related to some news spreading among investors. More precisely, this alternative hypothesis predicts that fundamental changes happen in the ETFs *before* they hit the concerned stocks themselves. Having shown that ETF trading have an impact on the prices of underlying assets, the tests try to assess whether the shocks due to ETF trading have a permanent (fundamentally driven) or temporary (liquidity-driven) impact. The acceleration in stock prices' response to fundamental news has been documented in theory and empirical results in Andrei and Hasler (2015) : in their model, there is a positive quadratic relationship between the time-varying degree of attention of investors and the stock's return variance, as well as a similar relation between learning uncertainty and variance, though weakly significant. This may come from the model, which specifies uncertainty as depending negatively from attention : participants will learn more information by being more attentive and therefore reduce their uncertainty. In other terms, stable prices require the market to take news into account (attention) as well as converging to similar conclusions based on news (the opposite of uncertainty, i.e. trust in your forecasts). It is perhaps more intuitive to think of attention and uncertainty in terms of the proxy used empirically than based on the stochastic processes defined in the model : time-varying attention to news is proxied using statistics of Google searches on a financial and economic set of words, whereas uncertainty is derivded from the dispersion in analyst forecasts.

the liquidity buffer hypothesis : instead of acting as a propagation channel for demannde shocks, ETF are supposed to act as substitutes for the underlying, less liquid securities since investors find the properties of stocks replicated in another liquid product. Thus, the introduction of an ETF captures part of the stocks's volatility non-fundamental volatility, which in turn decreases. This hypothesis was not developped in an ETF setting as it was first appeared in a commodity production model (Danthine (1978)) relating the introduction of futures with stabilized spot prices thanks to the information futures convey about rational expectations and therefore on supply – although it is recognized that futures attracts speculators too. In short, this theory is used as if ETFs were the new futures and as if they could reflect expectations on the underlying assets. Empirical evidence with futures and stocks is not unanimous and results are mixed : the introduction of futures has been linked with an increase in the volatility of Nikkei index stocks. What is more, if one assumes that the futures market's importance is proxied through traded volume and open interest, Bessembinder and Seguin (1992) find that indeed the introduction of futures decreased volumes exchanged on spot equity markets, that futures expected volume, resp. open interest, and spot volatility are negatively correlated together thanks to the increased market depth brought by futures. On the other side, the unexpected component of futures trading volume correlates positively with spot volatility. In general, their results are consistent with the hypothesis of the index arbitrage activity improving market depth highlighted in Grossman (1988) and they show no sign of support for increased instability due to liquid and relatively cheap correlated assets – at that time, futures, and nowadays ETFs ?

To be completed

6. Data

Most of the available literature regarding the impact of ETF ownership on stock's volatility and liquidity uses a specialized dataset that was unfortunately not available for this paper, namely the Thomson Reuters Mutual Fund Ownership Database also known as S12. Completeness is theoretically guaranteed in this dataset since its direct source is the filings that all investment companies[9] have to report quarterly to the U.S. Securities and Exchange Commission (SEC).

[9] A note by Wharton Research Data Services at the University of Pennsylvania lists the various types of investment companies reporting their holdings : "banks, insurance companies, parents of mutual funds, pension funds, university endomwents [as well as][...] professional investment advisors". Together, because of the form they report current assets held, they are called *13f institutions* and file with the regulator at least quarterly.

Whereas access to the Institutional Holdings (aggregated at the management companies' level), also known as S34, were available through WRDS at the University of Pennsyvania, they lacked granularity since most of if not all the major ETFs (according to the amount of funds held) are issued by a limited number of investment management companies. Nevertheless, Thomson Reuters data regarding the fund ownership of stocks can be found alternatively on their *Eikon* platform and the systematic lookup is implemented through an open application programmation interface (API) for which libraries have been written in common programming languages. Nevertheless, the origin of fund holdings data is less clear as I cannot ascertain that they come from S12 nor that they have not been modified before integration in *Eikon*[10].

The availability of ETF ownership is the critical variables when it comes to put time boundaries to the analysis : there are ETF holdings before 1999, whereas the Thomson Reuters Mutual Fund Ownership database starts around 1980 and would therefore theoretically allow to track the first ETFs active in the US ever, back in the early 1990s. Nevertheless, the aggregate value invested in those funds was almost insignificant during most of the decade and therefore I choose to collect month-end observations from January 1999 to December 2018, that is, over 240 months. Further restrictions due to lags will limit the effective length of the panel.

6.1. Exchange-Traded Funds characteristics

One of the most extensive sources of basic ETF data may be the items under share code `73` in the dataset maintained by the Center for Research in Security Prices (CRSP) at the University of Chicago, with access again granted through WRDS. Since the first Exchange-Traded Fund's inception in 1990 in Canada, they refer 2893 funds in this category, with the important restriction that CRSP only tracks securities in the United States. Without any restriction on the country of incorporation, Thomson Reuters exhibits 4085 active ETFs and it is not possible to get the sample of discontinued ETFs; on the other hand, the active sample seems to be a substantial share of all ETFs currently live worldwide.

In general, index-tracking ETFs[11] may engage in three different implementations of index-

[10] As of the date of redaction of this section, transition is ongoing towards a rebranding of the mentioned products, after the sale of the data unit of Thomson Reuters to Refinitiv.

[11] The context of this analysis still can be summarized as a gross distinction between mutual funds (actively managed, open-end funds without intraday liquidity and priced at their Net Asset Value at market close), index funds which are their passive counterpart (e.g. the iconic Vanguard 500 Index Fund with its beta of 1.00 to the S&P 500 equity index) and the ETFs that are open-end, index-tracking funds with a continuous arbitrage mechanism enabled by issuing their indicative intraday NAV every 15 seconds throughout the day.

tracking:

Full (or *physical*) replication The securities in the index portfolio are all held by the ETF in the actual proportions used to compute the index. We could say that this product is fully collateralized.

Optimized replication The ETF generally uses a proprietary algorithm to tilt its holdings towards the securities with the most contribution to the index volatility. Thus smaller securities encompassed by the index are neglected, in whole or in part.

Swap-based replication The ETF does not hold the "physical" securities but enters into derivatives such as a total return swap agreements which replicate the index performance including cash flows such as dividends. ETF shares creations and redemptions are therefore made in cash. Swap-based ETFs indirectly allow investors to access swaps while they - especially individuals - would not be able to enter such contracts with a bank on their own, but those products also expose them to the counterparty risk in exchage for tax advantages (i.e. income treated as capital gain).

While swap-based replication for instance allows leveraged and inverse ETFs, thefact those products do not actually trade the underlying securities prevents a demand shock happening on themselves from subsequently influencing the volatilityand liquidity of stocks. Since they constitute a minority of the broad ETF market in terms of value, especially in the United States, they are not kept in the sample of interest. Hence, there are 1798 full-replication, 1497 (explicitly) optimized-replication ETFs and 29 more exhibiting *Other*. Their documentation shows it means either an optimization regarding a specific sector or a mix between a majority of physical shares and derivatives in addition. It is relatively more worrying to acknowledge the 277 funds for which the value of the `Index replication method` variable is missing. At least a part among them may be actually swap- or option-based and therefore hold no physical asset – in which case, they could have been excluded from the sample of 3324 remaining ETFs.

TBD : FIGURE (AND TABLE) ABOUT THE AMOUNT OF ETFS IN NUMBER AND POSSIBLY CAPITALIZATION, AGAINST THE REMAINING TYPOLOGY OF FUNDS

This polarized landscape has been changing during the last decade though : preliminary results (Easley et al. (2018)) show that the share of ETFs explicitly labelled as active portfolios has been increasing (1.8% of assets under management in US equity ETF as of 2017, Ben-David, F. Franzoni, and Moussawi (2017)) while somewhat mixed performance relative to their benchmarks have made mutual funds less reliant on their managers' skills

Table 1: Summary of the variables' coverage in the ETF Characteristics table

Data columns	(total 20 columns)
Lipper_RIC	3601 non-null object
CUSIP	1814 non-null object
ISIN_Code	3577 non-null object
Asset_Name	3601 non-null object
Asset_Full_Name	1990 non-null object
SEC_Inception_Date	1378 non-null datetime64[ns]
Fund_Management_Company_Long_Name	3601 non-null object
Closed_Date	0 non-null float64
Exchange_Traded_Fund(ETF)	3601 non-null int64
ETF_Ticker	2509 non-null object
Asset_Status	3601 non-null object
UCITS	736 non-null float64
Legal_Structure	2221 non-null object
Index_Tracking	3333 non-null float64
Index_Replication_Method	3324 non-null object
Investment_Objective	3600 non-null object
Style_Matrix	1223 non-null object
Broad-Based_Index	1376 non-null object
Peer_Index	1369 non-null object

Type:	*<class 'pandas.core.frame.DataFrame'>*
RangeIndex:	*3601 entries, 0 to 3600*
dtypes:	*datetime64[ns](1), float64(3), int64(2), object(14)*
Memory usage:	*562.7+ KB*

6.2. Common stocks' fund ownership

The sample of live U.S.-traded stocks provided through the Screener function in Eikon is the basis for this analysis; it does not only include companies that are headquartered in the United States but also foreign (from the U.S. point of view) ones with a listing on a U.S. stock exchange. Overall, the sample contains 4978 active companies (as of March 4, 2019) with no conditioning on size, lifetime nor stock price; out of them, 4426 are U.S. companies and the remainder is split across 48 other countries, with almost 30% of them being located in China.

Fund ownership is available through the variable `TR.FundAdjShrsHeld`, meaning *the number of shares of a given stock held by a given fund, adjusted for corporate actions (e.g. stock splits).* For a stock at a given date, a query through the Thomson Reuters Eikon API allows to retrieve the whole fund ownership at once (ignoring the reporting issues. . .), with the fund ID, category and an essential information : the date of holdings reporting, which is a metadata of the shares held value. Enquiries in the database have shown that ETF holdings are generally up to date, whereas a larger share of other funds exhibit a reporting delay, i.e. a reporting date earlier than the query date. Indeed, in order not to introduce erroneous data, no spurious extrapolation (from the latest reported date to the query date) is decided, except if both dates lie in the same momth of the same year. Once they are grouped by month of report, only holdings of ETFs are kept, they are summed at the stock level and finally divided by the number of shares outstanding at month end (adjusted for corporate actions) to obtain the relative ETF ownership. Some outliers above or close to 100% ownership have been identified when specific companies went under the Chapter 11 protection of the US Bankruptcy Code and the fund amounts of shares held were not immediately updated for the equity being destroyed. Whenever this case occurred, the *de facto* invalid observations have been dropped. The final sample for U.S. stocks includes 712445 non-null observations.

6.3. Common stocks' market and accounting data

The Eikon API also provides an extensive access to time series of market data, of which daily close price, traded volume in terms of shares and – in order to compute the Amihud (2002) illiquidity ratio, a control variable – the volume-weighted average price (VWAP). Cross-sectional queries, similar to those regarding fund ownership, allowed to retrieve the remaining control variables, respectively the input necessary to compute them. Availability differs for every variable in the panel, which limits the number of observations actually used; by its very nature of gathering

Table 2: U.S. and International broad samples of underlying securities: categories of instruments

Asset Category Description	International	US	All
Ordinary Share	15225	4634	19859
Fully Paid Ordinary Share	1030	5	1035
Unit	57	114	171
American Depository Receipt	0	127	127
Preference Share	105	0	105
Global Depository Receipt	13	0	13
Participation Share	13	0	13
Depository Receipt	7	4	11
Closed-End Fund	6	3	9
Dutch Certificate	7	0	7
Brazilian Unit	5	0	5
Preferred Share	2	2	4
Stapled Security	3	1	4
Bond	2	1	3
Swedish Depository Receipt	3	0	3
Brazilian Depository Receipt	2	0	2
Genussschein	2	0	2
Savings Share	2	0	2
Convertible Preference Share	0	1	1
Exchange-Traded Commodity	1	0	1
Exchange-Traded Note	1	0	1
Non-Cumulative Preference Share	1	0	1
No available information	9	20	29
All	16496	4912	21408

stocks that have entered the sample at different times over twenty years, the panel is heavily unbalanced, which does not prevent the statistical analysis to hold as long as the availability of data is not correlated with variables of interest.

TBD : Summary of availability of data across time, size deciles (?) TBD : Table and figure for percentage of market cap across percentiles(abovebelow median, in the S&P 500/Russell 3000 fashion of Ben-David, F. A. Franzoni, and Moussawi (2018))

7. Methodology

7.1. Independent variables

7.1.1. Volatility

In order to keep the amount of data treated at a tractable level[12] and due to data availability limits, stock volatility is measured over a calendar month (end-to-end, adjusted for the number of trading days) as the standard deviation of simple returns using daily close series.

$$\text{Volatility}_{i,t} = \sqrt{\frac{1}{N\_d_{i,t} - 1} \sum_{d=1}^{N\_d_{i,t}} (r_{i,d} - \bar{r}_{i,t})^2} \tag{1}$$

Computing and analyzing intraday volatility is a different avenue for research, shown in Ben-David, F. A. Franzoni, and Moussawi (ibid.) : they use the US Trade and Quote (TAQ) database and compute intraday volatility over second-by-second returns, which which is used in a panel OLS regression with the following regressors : the absolute mispricing as a proxy for arbitrage activity, ETF ownership and the same controls included in their monthly database and in this paper (cf. subsection 7.4, p.35).

7.1.2. Liquidity : Amihud (2002) ratio

Liquidity has to be implied from various proxy variables. The illiquidity ratio introduced in Amihud (ibid.) is one of them and it has been used in literature both as a variable of interest in itself (e.g. Israeli, Lee, and Sridharan (2017)) and as a control for the volatility impact of ETF ownership (Ben-David, F. A. Franzoni, and Moussawi 2018).

$$\begin{aligned} \text{Illiq}_{i,t} &= \frac{1}{N\_d_{i,t}} \sum_{d=1}^{N\_d_{i,t}} \frac{\mid r_{i,d} \mid}{\text{Volume\_D}_{i,d}} \\ &= \frac{1}{N\_d_{i,t}} \sum_{d=1}^{N\_d_{i,t}} \frac{\mid r_{i,d} \mid}{\text{Volume}_{i,d} \cdot \text{VWAP}_{i,d}} \end{aligned} \tag{2}$$

The first line is Amihud (2002)'s original definition, whereas the second shows that the daily dollar volume, $Volume\_D$, is computed as the product between the volume expressed in terms of stock shares traded, $Volume$, and the volume-weighted adjusted price, or $VWAP$, since it is not an available data in the source database.

[12] A quick enquiry returns that the datasets overall more than 20 Gigabytes large, with monthly fund ownership files, segregated into US and international subsamples, being the largest arrays.

The method followed in liquidity regressions comes from Israeli, Lee, and Sridharan (2017), which document both liquidity and information-related effects due to ETF ownership : correlation between, on one side, higher ETF ownership and, on the other side :

lower liquidity : higher bid-ask spread and higher price impact of trades

lower price efficiency : higher stock returns synchronicity, lower future earnings response and, in the long run, lower analyst coverage.

The liquidity regressions will be explained in greater detail in the appropriate subsection (subsection 7.5, p.36).

7.1.3. Price efficiency

TBD; MEASURES INCLUDE THE FORWARD EARNINGS RESPONSE COEFFICIENT OF ISRAELI, LEE, AND SRIDHARAN (IBID.) AND THE ABSOLUTE VARIANCE RATIO OF BEN-DAVID, F. A. FRANZONI, AND MOUSSAWI (2018)

7.2. Regressor of interest : ETF ownership

The purpose of this paper is to identify and quantify the effect of the share of exchange-traded companies' equity being held through ETFs over several characteristics, namely the stocks' volatility, liquidity and price efficiency, all of three will be the dependent variables in dynamic panel regressions. Based on raw monthly fund-stocks number of shares held, the set of funds belonging to the ETF category and the overall number of shares outstanding of the given stock, the percentage of shares outstanding held is determined :

$$\mathsf{ETF\_Ownership}_{i,t} = \frac{\sum_{f=1}^{N_f} \#\_\mathsf{AdjShares\_Held}_{f,i,t} \cdot B_f}{\#\_\mathsf{Shares\_Out}_{i,t}} \tag{3}$$

$\forall i = 1 : N_i$ (stocks), $t = 1 : T$ (periods) with $B_f = 1$ if fund f is an ETF, 0 else.

The reporting frequency plays an important role as far as the accuracy of this variable is concerned: a substantial part of the reporting of their holdings to ETF, which is a regulatory task, is delayed each month : for example, the query for the number of shares owned by an ETF on February 28, 2018 yields a data pointas of the January 31, 2018; the delay can be more than one month. Two competing choices have been tried before it was decided to apply the same rule to the whole sample, out of a need of internal consistency. Even if both branches of this alternative contain disadvantages, a unique choice has to be made regarding the whole sample so that no spurious relationship appears later between one group of stocks and the other. Either unavailable monthly values are considered a marginal share of the aggregate, stock-level ownership ratios, and they are simply ignored, or the latest reported fund-level values are considered the "best guess" one can make regarding how many shares of this stock are currently owned through this fund, and they are included in our global, stock-level ratio; this guess will be referred to as *extrapolation*. In other terms, is the bias more severe by including slightly delayed, and therefore possibly outdated portfolios, or by only selecting the information that one has at hand with certainty, though incomplete ? The challenge is to understand, even *a priori*, whether the statistical and even economic interpretation of results will be less hamperedby a main regressor, total ETF ownership, computed precisely wrong (no extrapolation) or approximately correct (short term stale holdings). Without further intuition on the massive data collection and its coverage, one possible way of figuring how serious this issue actually is, is to compare several estimation results obtained with both methodologies. Even if the issue is known, its effects on a large panel are not guaranteed to be sizeable.

The granular data availability question is also relevant for other institutional ownership controls that will be used to capture a possible influence of other fund types over variables of interest. The fund categories distinguished in the database are, exhaustively, mutual funds, pension funds and hedge funds, all subject to an at least quarterly disclosure of their holdings. Due to the lower update frequency observed with those controls, it has been decided to assume the fund-stock holdings constant between the last available date and the current one, which reduces the volatility of aggregate ownership shares. Another strategy would be to reduce the frequency of observations from monthly to quarterly (at each calendar quarter end) over the whole sample. This estimation strategy remains to investigate in order to state whether such three-month periods are granular enough to perform relevant analyses.

7.3. Control variables

7.3.1. Bid-ask spread

The most straightforward way to compute the difference between the bid price – the highest price at which buyers on the market are agreeing to pay for the security on the spot market – and the ask price – the lowest price sellers agree to receive for their shares – is to compute the difference between the variables `TR.AskPrice` and `TR.BidPrice` at day close . Whenever both values returned by the data provider are not null, the absolute (difference) measure can be computed and in cross-sectional regressions, the relative measure is used as a liquidity control :

$$\mathsf{Pct\_BidAskSpread}_{i,t} = \frac{\mathsf{Ask}_{i,t} - \mathsf{Bid}_{i,t}}{\frac{\mathsf{Ask}_{i,t}+\mathsf{Bid}_{i,t}}{2}} \tag{4}$$

The denominator in the relative bid-ask spread is the mid-price approximated as the arithmetic average between both quoted prices, without considering the rounding that can occur due to the tick size.

The relative bid-ask spread accounts for the cost of trading, which is itself assumed to correlate positively with the illiquidity, i.e. the weak number of agents willing to trade on a market. The market makers require a higher price, the bid-ask spread, in order to compensate for the risk of note being able to net out their position in an asset rapidly. Thus, a higher bid-ask spread constitutes a limit to arbitrage across markets or assets.

Another explanation has been studied : according to the Glosten and Milgrom (1985) market-microstructure model, the bid-ask spread reveals the presence of informed traders and it can even exist in a competitive market without trading costs. If there are both informed and uninformed (so-called *noise*) traders and and market makers (such as the operators acting as middlemen on the NYSE) cannot tell whether a submitted order in which they are the counterparty comes from either group of traders, they (market makers) will infer determine their bid and ask prices based on the conditional expectation about the asset value based on the direction (buy or sell) of the order they are facing. In this model, the bid-ask spread accounts for the adverse selection, because transactions convey information, and creates a divergence between observed returns on securities and the returns that could be made by an uninformed trader – a difference that becomes relatively smaller, the longer the investor holds its asset, the authors show. The bid-ask spread can therefore be included in cross-sectional regressions in order to account for the unequal liquidity provision as well as the unequal availability of information regarding firms.

7.3.2. Fama-French factors

In the original paper about the cross-section of stock returns (Fama and French 1992) as well as in their generalization to bond returns using a new methodology (Fama and French 1993), Eugene Fama and Kenneth French introduce an empirically-founded, five-factor (counting the market return) extension to the Capital Asset Pricing Model. Three factors come from the equity universe : market return, size and value, while two factors are specific to bonds : a term (i.e. maturity) premium and a default risk premium. The methodology in the later paper provides a common framework for stocks and bonds and conclude that the explanatory power of the CAPM beta nearly disappears when the size and value factors are taken into account. They change the common, "CAPM-based" view by showing that, across sorted all portfolios, the residual sensitivity to the market return is the same and reflects a risk premium attached to any stock, compensating the investor for not investing in a bond instead. Here we will retain the three-factor model that is already powerful for explaining the differences of returns across stocks.

Size Small-capitalization firms have been shown to yield a higher return adjusted by their market exposure and this phenomenon justifies the existence of a risk premium according to Fama and French (1992). More fundamentally, the fact that the market β of small capitalizations is not able to fully account for their higher returns can already be found a decade earlier in Banz (1981), which reports its existence over more than forty years among NYSE common stocks.

Since the simple designation as an *anomaly* is not a viable reason for this persisting phenomenon, several explanations for the existence of a small-minus-big risk premium have been proposed.

Risk-based explanations Earlier theories (Berk 1995) claim that the size factor, since it is measured through market value, only reflects the fact that riskier firms have to pay higher expected returns, or conversely have a lower market value. If this hypothesis is true, the market value captures an individual risk premium but no other non-price-based measure of firms' size, whether the book value of equity, of total assets, the sales revenue or even the number of employees, will have any explanatory power regarding expected returns. Another theory yields the same relationship through growth options : they are assumed to be a risky component that is concentrated in small firms and, therefore, smaller companies have to pay their investors higher returns in order to compensate for the included growth options. Gârleanu, Panageas, and Yu (2012) support their model with consistent simulation results and empirical evidence of the size effect but fail to exhibit simulateneously a size and a value factor: both effects drive each other out.

Behavioral explanations In this strand of literature, the focus on size is relatively minor compared to value and momentum (over/underreaction, see subsubsection 7.3.3, although K. D. Daniel, Hirshleifer, and Subrahmanyam (2001) develop a model that jointly allows for the book-to-market ratio and the market value (our size effect) to predict returns. Their model incorporates overly confident insiders that cause mispricing and other risk-averse traders that do not completely eliminate mispricing.

Liquidity-based explanations The third branch of competing hypotheses regarding the size effect is based on liquidity, whether its level or its risk – and it is shown that both seem negatively correlated. Small firms are both less liquid on average and more risky, which explains a positive risk premium relative the rest of the market. Acharya and Pedersen (2005) propose a model with four betas, i.e. three additional covariances between the respective market and security returns and liquidity risks, those being proxied through a normalized version of the Amihud (2002) ratio.

After it had been considered less important in magnitude over history, yielding weaker returns than other risk premia (e.g. value, momentum) when targetted specifically, possibly arbitraged away after the 1980s or at least hard to find outside the U.S.[13] Cliff Asness, a principal at AQR Capital Management who wrote his PhD thesis under Eugene Fama's supervision, has tried to ressurrect size in a *Journal of Financial Economics* paper, C. Asness et al. (2018), with an eloquent title : *Size matters, if you control for your junk.* The authors show that the significance of the factor greatly increases once another, so far ignored, factor is taken into account : it is called quality and this factor can only be a composite statistic built over several more explicit measures: profitability, which is used as a control and developed later (subsubsection 7.3.4, p.33), earnings and cash flows growth, safety (low beta, leverage, bankruptcy risk and earnings volatility), payout: characteristics of a stock assumed to be sought for by investors. C. S. Asness, Frazzini, and Pedersen (2019) present empirical tests for 25 countries using a normalized metric that combines the aforementioned aspects and show that high-quality firms, which on average have a low beta and low exposures to alternative factors (size, value, momentum), perform better than so-called junk companies during market downturns.

As mentioned earlier, in this paper it would be too intensive in data for the expected benefit of such a metric, and size along with its best counterpart are only controls that should account for possible, yet extremely speculative influence over volatility, not expected returns. Thus I consider sufficient to proxy the quality composite metric of C. Asness et al. (2018) using one of its component only, i.e. gross profitability.[14]

[13] This is not a comprehensive inventory of somewhat mixed evidence regarding the size risk premium.

[14] Typically, precision is traded for tractability in the estimation methods and data availability issues would only make the sample even smaller, less representative – and so the conclusions drawn from results, because, precisely, small firms often lack data points. I am nevertheless conscious that such a shortcut may not account for the perceived quality-versus-junk judgement : a profitable company may exhibit earnings forecasts that are unstable, even around positive trend. It may profit from a market with entry barriers or a even a protected monopoly and commit little investment. These are examples of situations where the quality rating is impaired

Book-to-Market ratio Similar to the size effect, the inability of the CAPM to account for positive excess returns in high versus low book-to-market value firms, also called respectively value and growth companies, appears at least as early as in Rosenberg, Reid, and Lanstein (1985).[15]

It may further be noted, as an application for a long-short equity portfolio, that C. S. Asness, Frazzini, and Pedersen (2019) demonstrate improved risk-adjusted returns (measured through the Sharpe ratio) for a strategy they call *Quality at a reasonable price.* The baseline case is a portfolio based on sorting a portfolio on the HML (value) factor and adding a quality sort is a possible way to implement this improvement. The authors explain that quality (meant as desirable characteristics) actually complements value (the expensiveness of a company relative to its balance sheet) because this double sort excludes shares with high book-to-market ratio that score low on quality – in other terms, stocks that only look undervalued but are not.

7.3.3. Momentum

Momentum is, chronologically, the fourth factor that has been found to explain the cross-section of expected returns, identified in Carhart (1997), a paper about persistence in mutual funds' risk-adjusted returns; this paper shows that the one-year momentum effect from Jegadeesh and Titman (1993) makes the manager's skill or superior information irrelevant to explain the fund's performance, except for the worst-performing funds.

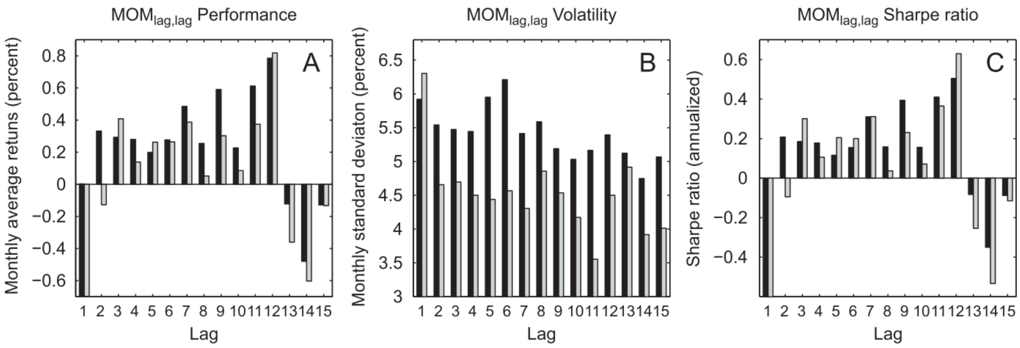
A portfolio based on buying the (relative, say top quintile) winners and selling the losers based on their return over the previous month only (Novy-Marx 2012), is a definitely losing strategy. The negative coefficients are significant with very little doubt (t-statistics between -10 and -20. If momentum has been found in the equity world, essentially the same in equal-weighted as well as value-weighted portfolios, this positive correlation is strictly bound between the past twelve to two months before the start of the securities holding period.

This concept of cross-sectional momentum corresponds to the traditional idea of momentum

and a good profitability ratio is the only good component – intellectual honesty commands to raise this flag.

[15] Even if both additional factors in this section nowadays bear his name, Eugene Fama writes in his academic biography, *My Life in Finance*, that his Fama and French (1992) "contains nothing new", precisely because the main results were already published individually in papers dating back to the early 1980s. Fama thinks that the collection of evidence into one paper (actually, the first of a series) spread the message, before Fama and French (1993) passed the test for replacing a model, the CAPM : introducing a new model with more explanatory power, at least in statistical terms (R^2). Indeed, the economic or behavioral justification attempts, though summarized by Fama in its address, are "unconvincing" in his opinion.

Figure 1: From Novy-Marx (2012), "Marginal strategy performance". Comparison of returns, standard deviation and Sharpe ratio of long top-decile/short bottom-decile of lagged performance strategies, varying the lag



Description from original paper:

> *Fig. 1. Marginal strategy performance. This figures shows the average monthly returns (Panel A), monthly standard deviations (Panel B) and annual Sharpe ratio (Panel C) to winners-minus-losers strategies. Winners and losers are defined as the top and bottom deciles of performance in a single month, respectively, starting lag months prior to portfolio formation. Dark bars show value-weighted results and light bars show equal-weighted results. Average monthly returns for the one month reversals are* -1.04% *(value-weighted) and* -2.82% *(equal-weighted). The sample covers April 1927 to December 2010.*

whereas Novy-Marx (2012) finds that positive correlation is essentially coming from the first five months of the previous year, i.e. the return between $t-12$ and $t-7$ (included). The expected return and Sharpe ratio of several trading strategies tested exhibit figures twice larger for 12-to-7 winners-minus-losers (WML) portfolios compared with 6-to-2 WML portfolios. Describing a term structure of momentum put in evidence thanks to CRSP data spanning from 1926 to 2010 Novy-Marx (ibid.) claims that

> Theoretically, the return predictability implied by the data, which looks *more like an echo than momentum*[16], poses a significant difficulty for stories that purport to explain momentum.

Results in Novy-Marx (ibid.) indeed defeat possible and competing explanations of momentum cited in his review of literature. Nevertheless, Goyal and Wahal (2015) have extensively tested and extended this surprising result: no evidence of this echo pattern has been brought outside the U.S. stock market and even in the U.S., the superior returns of 12-to-7-month sorted portfolios over their 6-to-2-month counterparts may be due to the fact that short-term reversal, usually considered over the previous month, sometimes extends further in the past over month $t-2$. I do not take any side without having tested the data from my own paper and therefore consider both alternatively in tests. The aim in this section is to explain why the momentum effect has to be included as a control in regressions involving returns or the volatility of returns, rather than provide a theoretical rationale for doing so. In general, let us summarize as follows : models "predicting short-term predictability" (i.e., attempting to explain momentum, as the author nuances) are in either of two categories:

Behavioral Momentum arises from the delayed and progressive incorporation of news into prices, meaning that the market underreacts to a news spreading out and catches up during the following months. Thanks to a model of investor sentiment based on the famous representativeness heuristics introducted by Tversky and Kahneman (1974),Barberis, Shleifer, and Vishny (1998) model simultaneously underreaction in the short term (up to 12 months after news) and overreaction in the longer term, between 3 and 5 years if the stream of news is consistently good or bad, highlighting conservatism among agents. K. Daniel, Hirshleifer, and Subrahmanyam (1998) attribute momentum to what they call *biased self-attribution*, which is a form of overconfidence in one's piece of privileged information and a common trait in behavioral finance (De Bondt and Thaler 1994). Themselves citing another arti-

[16] Emphasis added in the quote.

cle, K. Daniel, Hirshleifer, and Subrahmanyam (1998) summarize the self-attribution bias as follows: "Heads I win, tails it's chance." Confidence gets stronger, the more information confirms their private signal whereas so-called *disconfirming*, i.e. contrary, evidence only reduces confidence to a little extent. Continued confirmation therefore amplifies (in intensity) and extends (over time) the agents' initial overreaction, yielding positive price auto-correlation in the short run, i.e. momentum. Both of these behavioral models assume a single representative trader whose bias causes short-lag autocorrelation in returns; Hong and Stein (1999) choose a fundamentally different approach as they segregate investors into two categories, each being "boundedly rational": *newswatchers* are gathered into subgroups and gradually have access to private information and trade according to this piece of information but they do not condition their decision on past and current prices, only the information[17]. *Momentum traders* are the opposite category, i.e. they trade according to the price trend over a finite horizon but do not have access to any private information; in other terms, they observe, follow and extend an existing trend. Overall, in this model, the momentum effect therefore does not arise from either of the two groups of traders, rather does it appear from their interaction. While K. Daniel, Hirshleifer, and Subrahmanyam (1998) build their behavior model on psychological research results, Barberis, Shleifer, and Vishny (1998) and Hong and Stein (1999) support their models with testable hypotheses and empirical evidence.

Rational Predictions of the momentum that do not rely on investors' behavior are less frequent in literature but two contributions can be shortly summarized. First, the concern of Johnson (2002) is less about providing an equilibrium model with robust empirical evidence than show that momentum can arise even in a rational setting, along with mean-reversion approximately one-year after measurement. The fact growth rate risk rises with growth rate is at the core of the momentum effect and infrequent, persistent, e.g. technological shocks would cause most of this effect. Stocks prices depend on growth rates, because the former are modelled as a claim on a stochastic stream of dividends which grow at a random stationary rate. The growth rate risk is priced in the stock and, to quote this rather

[17] The authors seem conscious of the debatable realism of their modelling regarding newswatchers, which act without observing past and present prices and thus their own previous trades. It is suggested that they act like frontrunners, taking advantage of information eearly available and conscious that their trades will start a reaction (future asset price increases in case of good news). This attitude has a limited effect in the model, which means that short term underreaction still prevails.

quantitative paper :

> [M]omentum effects then follow because positive (negative) cumulative returns typically imply ex post that recent growth rate shocks have been positive (negative).

Another empirical paper with a real options model, Sagi and Seasholes (2007), aims specifically at identiying firm-specific variables that drive momentum and thus allow to build more profitable trading strategies conditioning on those drivers, thus outperforming the baseline Jegadeesh and Titman (1993) winners-minus-losers, 6-month holding portfolio. Positive return autocorrelation due to the company value being convex in an underlying risk factor is specified in this paper, borrowing to Johnson (2002) : the underlying risk factor was the growth rate risk. Empirical findings are the following : momentum varies positively with the volatility of revenue (sales) growth, negatively with costs of goods sold and negatively with the book-to-market ratio. For the latter, it means that a more profitable momentum strategy can be implemented on firms sorted according to their book-to-market ratio; this result does not invalidate the predictive power of the book-to-market ratio over expected returns, i.e. the HML factor from Fama and French (1992).

7.3.4. Gross profitability

According to Novy-Marx (2013), gross profitability, i.e.

$$\mathsf{Gross\_Profitability}_{i,t} = \frac{\mathsf{Revenues}_{i,t} - \mathsf{COGS}_{i,t}{}^{18}}{\mathsf{Total\_Assets}_{i,t}} \tag{5}$$

has a prediction power equal in magnitude and complementary to the book-to-market ratio over the cross-section of expected returns. Novy-Marx (ibid.) has deemed this factor the *other side of value* because it does not subsume it while it is linked to it. Both factors can be exploited together and improve each other's risk-adjusted performance in a portfolio. The value factor measures the market price of a company's assets and finances the purchase of inexpensive assets through the sale of expensive ones while the profitability ratio measures how productive assets within the firm are and finances the purchase of productive ones through the sale of unproductive (or at least, less productive) ones.

The influence of profitability had already been studied before Novy-Marx (ibid.) showed the existence of a predictive power over the cross-section of expected returns and thus the opportunity

[18] Cost of goods sold

of a trading strategy : indeed Fama and French (2006) treat the book-to-market, profitability and investment effects combined, although the authors remain agnostic on the mechanism, either rational or behavioral, underpinning their threefold statement. The statement originates from the dividend discount model, which states that the market value of the company is equal to the sum of discounted divided expected to be paid to shareholders in the future:

$$M_t = \sum_{\tau=0}^{\infty} \frac{\mathbb{E}_t(Y_{t+\tau} - dB_{t+\tau})}{(1+r)^{\tau}} \tag{6}$$

where M_t is the market value cum-dividend at time t; Y_t are the earnings; $dB_t = B_t - B_{t-1}$ is the increase of the book value of equity between $t-1$ and t, and r used in the discount factor is the required rate of return. Under the so-called clean surplus accounting, earnings are either retained within the company, thus increasing the book equity, or it is distributed to shareholders; subsequently, the difference within the expected value brackets is equal to dividends. Through this identity, Fama and French (ibid.) imply and later show that, *ceteris paribus*, firms with higher earnings must yield higher expected returns (the r variable) as long as their market valuations are the same, allowing the profitability premium to exist. Regarding the value premium, for fixed earnings and book value series over the whole (here, assumed infinite) time horizon, the lower the market value, the higher the expected return must be, which is gives birth to the value premium.

Overall, two measures of profitability seem to correlate with expected returns : earnings from the income statement were tested and Fama and French (ibid.) did not find an incremental predictive power beyond the size and book-to-market factors. Novy-Marx (2013) rationale for gross profitability can be summarized as follows : some expenses are treated as costs, such as R&D, human capital development, but in fact, they will likely yield the company higher profits in the future. Therefore, one has to look at a figure higher in the income statement in order to approach the pure operating performance of the company and filter out some irrelevant costs. The same happens for the computation of free cash flows : for example, capital expenditures are not the sign that the company is less profitable, rather the opposite can be expected in the future. The author's tests putting various profitability ratios lead to conclude that gross profit over assets (and not the book equity, thus the measure is independent from the debt level) is a less-biased proxy with the most predictive power.

7.4. Impact of ETF ownership on stocks' volatility

The theoretical basis of all main and control variables included in the models have been discussed and the inclusion of controls is justified through a summary of the abundant literature of empirical asset pricing. In this subsection, the first identification strategy attempts to answer the question : does the share of a stock globally owned by exchange-traded funds has a contemporaneous impact over the stocks' volatility, all other characteristics being equal ? The null hypothesis being conservative ("no effect"), its potential rejection would mean that the liquidity trading hypothesis is actually reflected in the panel data with a controlled error risk level that will be disclosed in the estimation results.

Due to the structure of dataset, which is an unbalanced panel of U.S., respectively international stocks, some unique characteristics inherent to each stock will be controlled through entity fixed effects. Since the trend over the measurement period has been an exponential increase in the weight of ETFs, both in terms of stock ownership and volume traded, some fixed effects over the time axis seem legitimate. Otherwise, one could imagine that, provided that the average volatility of each stock does not significantly varies, the higher ETF ownership would results in a spuriously negative coefficient (!). Another reason for month fixed effects deals with short term market events, such as several market crashes that occurred during the two decades of the sample.

$$\mathsf{Volatility}_{i,t} = \beta_0 + \beta_1 \mathsf{ETF\_ownership}_{i,t} + B_C^{\intercal} \mathsf{Controls}_{i,t} + \alpha_i + \gamma_t + \epsilon_{i,t} \tag{7}$$

Rather than a panel analysis, it would be more accurate to call this regression a dynamic panel analysis : several lags of volatility will be included among controls in some estimations in order to account for a potential case of autocorrelation in volatility. Further investigation will also control for a potential bias in the effect of the ETF ownership share over stock volatility : what if ETFs were only a symptom of the institutional ownership, which would in that case exhibit comovement with the ETF ownership. The collection of data allows us to account for certain other broad categories of institutional investors, legally forced to report their holdings: mutual funds, pension funds and hedge funds. Controlling for the aggregate ownership share of each category may help to distinguish between several channels between institutional trading and volatility.

7.4.1. Risk of endogeneity bias : the need for an instrument

The direct estimation of the effect of ETF ownership on volatility may bear a subtle and essential caveat: there may exist an unknown cause, an ommitted factor that would be correlated both with ETF ownership and volatility. A robust identification would alleviate this possible endogeneity bias by using an instrumental variable based on a restriction. For the instrument to be valid, it has to affect the volatility only through its correlation with ETF ownership. Since a significant share of ETFs track well-known indices, the literature (e.g. Ben-David, F. A. Franzoni, and Moussawi (2018)) uses the inclusion/exclusion event of a given stock into the index to explain the variation of ETF ownership. The second step consists in an panel regression similar to the previous ones in which the observed ETF ownership is replaced with the fitted value from the first stage regression.

7.5. Impact of ETF ownership on market and stock liquidity

This segment of the analysis focuses on evidence about of ETF ownership on proxies for liquidity : Israeli, Lee, and Sridharan (2017) have shown that an increase in ETF ownership is correlated with higher trading costs, measured using the bid-ask spread, and lower liquidity, measured through a price-impact variable is equal to the numerator of Amihud (2002) ratio.

The following regression is run regarding the effect of ETF ownership on the bid-ask spread:

$$\mathsf{PctBidAskSpread}_{i,t} = \beta_0 + \beta_1 \mathsf{ETF\_ownership}_{i,t} + B_C^{\intercal} \mathsf{Controls}_{i,t} + \alpha_i + \gamma_t + \epsilon_{i,t} \tag{8}$$

Israeli, Lee, and Sridharan (2017) measure this relationship in first differences using an annual panel. The stationarity concern will be investigated further and inspiration is evidently drawn from their research scheme. They add several controls and industry-level fixed effects.

The second proxy for liquidity is Amihud (2002) illiquidity ratio, which numerator and denominator may be correlated with ETF ownership : the numerator is the absolute daily return and the denominator, feared to be especially sensitive to ETF ownership's influence, is the daily dollar volume traded. A spurious improvement (lower illiquidity) could be linked with ETF ownership if the latter increases the denominator, possibly offsetting the simultaneous increase in absolute return. The proposed way to mitigate this inherent issue is to split the ratio and consider the denominator as a regressor.

$$\mathsf{Illiq^{Num}}_{i,t} = \beta_0 + \beta_1 \mathsf{ETF\_ownership}_{i,t} + \beta_2 \mathsf{Illiq^{Denom}} + B_C^{\intercal} \mathsf{Controls}_{i,t} + \alpha_i + \gamma_t + \epsilon_{i,t} \tag{9}$$

7.6. Concerns about informational efficiency

Pricing efficiency in Ben-David, F. A. Franzoni, and Moussawi (2018) is tested as follows : they test whether the increased ETF ownership yield a higher degree of mean-reversion in pprices. In other terms, a stronger negative auto-correlation in prices (over 5 days) is the sign that they become noisier due an increased non-fundamental volatility. Such evidence is considered a empirical confirmation of the liquidity-trading hypothesis.

The variance ratio is defined as :

$$VR_{i,t} = \frac{Var(r5_{i,t})}{5 \cdot Var(r1_{i,t})} \tag{10}$$

8. Panel regression results and discussion

In this whole section, descriptive statistics and results will always be given in the same order : first the U.S. stocks' sample, which is a tentative replication of Ben-David, F. A. Franzoni, and Moussawi (ibid.) and then the international stocks' sample, which is the newer contribution of this paper. The models tested are the same across both samples.

8.1. Summary statistics

This section aims first at providing an overview about the distribution of all variables in the upcoming regressions. In addition, although a thorough comparison with the emerging literature about ETF ownership effects on securities is beyond the scope of this paper, summary statistics help ensuring that the sample bears some resemblance with those used in other studies. Resemblance is obviously weaker, less precise and rigorous than similarity and one should for instance perform statistical tests such as the F-test or Bartlett's test for equal variances across populations, under normality assumptions, if two samples are available. Since only one sample is available in our case, such comparisons are actually impossible. Lastly, summary statistics are also useful in order to filter outliers at both ends of distributions.

Table 3: U.S. Sample (monthly) : Summary Statistics

(a) Summary statistics

	N (obs.)	Mean	St. dev.	Min.	25%	Median	75%	Max.
Volatility	300181	0.027	0.026	0.000	0.015	0.021	0.032	1.618
ETF Ownership	300181	0.028	0.035	0.000	0.001	0.015	0.042	0.990
Book-to-market	300181	1.001	156.778	-196.776	0.243	0.439	0.740	79321.012
Market cap. ($ Mln.)	300181	716.934	2846.498	0.013	30.072	107.263	374.055	109943.606
1/Price	300181	0.089	0.120	0.000	0.026	0.048	0.098	1.000
Rel. Bid-Ask spread	300181	0.005	0.014	-0.015	0.000	0.001	0.004	1.765
Amihud ratio	300181	0.000	0.000	0.000	0.000	0.000	0.000	0.101
Past 12-to-1-month return	300181	0.154	0.652	-1.000	-0.143	0.075	0.319	43.375
Past 12-to-7-month return	300181	-0.049	3.609	-1868.259	-0.119	0.037	0.161	0.965
Gross profitability	300181	0.332	0.349	-0.850	0.150	0.272	0.438	36.028

(b) Pearson correlation coefficients

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Volatility	(1)	1.000									
ETF Ownership	(2)	-0.229	1.000								
Book-to-market	(3)	-0.000	-0.003	1.000							
Market cap. ($ Mln.)	(4)	-0.001	0.028	-0.001	1.000						
1/Price	(5)	0.010	-0.217	0.004	-0.126	1.000					
Rel. Bid-Ask spread	(6)	0.437	-0.186	0.000	-0.072	0.312	1.000				
Amihud ratio	(7)	0.016	-0.001	-0.000	-0.001	0.005	0.032	1.000			
Past 12-to-1-month return	(8)	-0.044	-0.001	-0.005	0.002	-0.089	-0.065	-0.002	1.000		
Past 12-to-7-month return	(9)	-0.006	0.003	-0.005	0.005	-0.026	-0.013	-0.000	0.010	1.000	
Gross profitability	(10)	0.019	-0.008	0.000	0.001	0.003	0.002	-0.001	0.052	0.002	1.000

Table 4: International Sample : Summary Statistics

(a) Summary statistics

	N (obs.)	Mean	St. dev.	Min.	25%	Median	75%	Max.
Volatility	1338479	0.420	0.125	0.000	0.397	0.418	0.441	109.046
ETF Ownership	1338479	0.003	0.012	0.000	0.000	0.000	0.001	0.950
Book-to-market	1338479	0.875	33.362	-30363.758	0.360	0.693	1.237	1102.476
Market cap. ($ Mln.)	1338479	14351.004	214744.406	0.001	113.098	609.315	3385.835	39864005.168
1/Price	1338479	0.082	0.551	0.000	0.001	0.004	0.060	333.333
Rel. Bid-Ask spread	1338479	0.013	0.039	-2.000	0.002	0.005	0.013	2.000
Amihud ratio	1338479	0.000	0.004	-0.000	0.000	0.000	0.000	2.450
Past 12-to-1-month return	1338479	0.162	1.381	-1.000	-0.159	0.048	0.317	1249.000
Past 12-to-7-month return	1338479	-0.036	6.559	-5282.798	-0.128	0.019	0.151	0.997
Gross profitability	1338479	0.282	0.389	-79.835	0.121	0.215	0.365	76.409

(b) Pearson correlation coefficients

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Volatility	(1)	1.000									
ETF Ownership	(2)	-0.004	1.000								
Book-to-market	(3)	-0.001	-0.002	1.000							
Market cap. ($ Mln.)	(4)	-0.000	0.079	-0.000	1.000						
1/Price	(5)	0.000	-0.029	-0.002	-0.011	1.000					
Rel. Bid-Ask spread	(6)	0.018	-0.073	-0.002	-0.018	0.175	1.000				
Amihud ratio	(7)	0.000	-0.001	-0.000	-0.000	0.021	0.044	1.000			
Past 12-to-1-month return	(8)	-0.000	-0.003	0.000	0.001	-0.004	0.003	-0.000	1.000		
Past 12-to-7-month return	(9)	0.000	0.000	0.000	0.000	0.000	-0.011	-0.000	0.001	1.000	
Gross profitability	(10)	-0.000	-0.010	0.021	-0.004	-0.003	0.012	-0.001	0.003	-0.040	1.000

8.2. ETF ownership and underlying stocks' volatility

For the sake of completeness, more detail is reported in the appendix, subsection C.1, p.49, about the panel estimation results testing the relationship between the share of ETF ownership in securities and the held securities' daily returns volatility.

8.3. ETF ownership and underlying stocks' liquidity

Even though no significant effect has been found so far in this subsection, more detail is reported in the appendix, subsection C.2, p.54, about the panel estimation results testing the relationship between the share of ETF ownership in securities and the statistics linked with stock-level liquidity (bid-ask spread, Amihud (2002) ratio decomposition).

8.4. ETF ownership and concerns about pricing efficiency

	Abs. VR	**Abs. VR w/inst. o'ship**	**VR**	**VR w/inst. o'ship**
Dep. Variable	absVR	absVR	VR	VR
Estimator	PanelOLS	PanelOLS	PanelOLS	PanelOLS
No. Observations	126851	126847	126851	126847
Cov. Est.	Driscoll-Kraay	Driscoll-Kraay	Driscoll-Kraay	Driscoll-Kraay
R-squared	0.0079	0.0081	0.0040	0.0041
R-Squared (Within)	0.0106	0.0106	0.0040	0.0042
R-Squared (Between)	0.2814	0.2820	0.1376	0.1385
R-Squared (Overall)	0.0318	0.0320	0.0163	0.0165
F-statistic	123.79	91.734	61.843	46.323
P-value (F-stat)	0.0000	0.0000	0.0000	0.0000
Intercept	0.8523	0.8532	0.8735	0.8579
	(12.280)	(12.208)	(7.6512)	(7.4328)
PctSharesHeldETF__1lag	0.0096	0.0218	-0.1205	-0.1005
	(0.1484)	(0.3373)	(-1.4081)	(-1.1660)
np.log(CompanyMarketCap__1lag)	-0.0248	-0.0248	0.0044	0.0052
	(-7.5202)	(-7.4640)	(0.8038)	(0.9458)
InvClose__1lag	0.1276	0.1277	-0.2002	-0.1986
	(5.5223)	(5.5778)	(-6.0619)	(-5.9225)
AmihudRatio__1lag	9.9267	9.9488	-15.135	-15.117
	(3.1431)	(3.1455)	(-3.0288)	(-3.0279)
PctBidAskSpread__1lag	0.9314	0.9286	-1.2190	-1.2179
	(5.5878)	(5.6027)	(-3.5732)	(-3.5633)
BookToMarketRatio__1lag	-3.498e-06	4.433e-05	-1.257e-05	6.777e-05
	(-0.5195)	(3.9695)	(-1.4303)	(3.2826)
RetPast12to7M__1lag	0.0003	-9.153e-05	-0.0006	-0.0013
	(0.5981)	(-0.2511)	(-0.9013)	(-2.2311)
GrossProfitability__1lag	0.0011	0.0012	0.0064	0.0062
	(0.2252)	(0.2454)	(1.2767)	(1.2635)
PctSharesHeldOtherMutual__1lag		-0.0053		-0.0086
		(-3.9260)		(-3.6908)
PctSharesHeldPension__1lag		11.292		-3.7029
		(4.1468)		(-0.5362)
PctSharesHeldHedge__1lag		-0.1294		-0.1145
		(-0.8927)		(-0.6268)
Effects	Entity	Entity	Entity	Entity

T-stats reported in parentheses

Despite the, at best fragile, evidence exhibited in this subsection, more detail is reported in the appendix, subsection C.3, p.59, about the panel estimation results testing the relationship between the share of ETF ownership in securities and the statistics of mean reversion.

9. Conclusion and further research questions

References

Acharya, Viral V. and Lasse Heje Pedersen (2005). "Asset pricing with liquidity risk". In: *Journal of Financial Economics* 77.2, pp. 375–410. ISSN: 0304405X. DOI: 10.1016/j.jfineco.2004.06.007. URL: https://www.sciencedirect.com/science/article/pii/S0304405X05000334%20https://linkinghub.elsevier.com/retrieve/pii/S0304405X05000334.

Agarwal, Vikas et al. (2018). "Do ETFs Increase the Commonality in Liquidity of Underlying Stocks?" In: *SSRN Electronic Journal*. ISSN: 1556-5068. DOI: 10.2139/ssrn.3070550. URL: https://www.ssrn.com/abstract=3070550.

Amihud, Yakov (2002). "Illiquidity and stock returns: Cross-section and time-series effects". In: *Journal of Financial Markets* 5.1, pp. 31–56. ISSN: 13864181. DOI: 10.1016/S1386-4181(01)00024-6. URL: http://linkinghub.elsevier.com/retrieve/pii/S1386418101000246.

Amihud, Yakov and Haim Mendelson (1986). "Asset pricing and the bid-ask spread". In: *Journal of Financial Economics* 17.2, pp. 223–249. ISSN: 0304405X. DOI: 10.1016/0304-405X(86)90065-6. URL: https://doi.org/10.1016/0304-405X(86)90065-6%20https://linkinghub.elsevier.com/retrieve/pii/0304405X86900656.

Andrei, Daniel and Michael Hasler (2015). "Investor attention and stock market volatility". In: *Review of Financial Studies* 28.1, pp. 33–72. ISSN: 14657368. DOI: 10.1093/rfs/hhu059. URL: https://academic.oup.com/rfs/article-lookup/doi/10.1093/rfs/hhu059.

Asness, Clifford S., Andrea Frazzini, and Lasse Heje Pedersen (2019). "Quality minus junk". In: *Review of Accounting Studies* 24.1, pp. 34–112. ISSN: 1380-6653. DOI: 10.1007/s11142-018-9470-2. URL: http://link.springer.com/10.1007/s11142-018-9470-2.

Asness, Clifford et al. (2018). "Size matters, if you control your junk". In: *Journal of Financial Economics* 129.3, pp. 479–509. ISSN: 0304405X. DOI: 10.1016/j.jfineco.2018.05.006. URL: https://linkinghub.elsevier.com/retrieve/pii/S0304405X18301326.

Banz, Rolf W. (1981). "The relationship between return and market value of common stocks". In: *Journal of Financial Economics* 9.1, pp. 3–18. ISSN: 0304405X. DOI: 10.1016/0304-405X(81)90018-0. URL: https://www.sciencedirect.com/science/article/pii/0304405X81900180%20https://linkinghub.elsevier.com/retrieve/pii/0304405X81900180.

Barberis, Nicholas, Andrei Shleifer, and Robert Vishny (1998). "A model of investor sentiment". In: *Journal of Financial Economics* 49.3, pp. 307–343. ISSN: 0304-405X. DOI: 10.1016/S0304-405X(98)00027-0. URL: https://www.sciencedirect.com/science/article/pii/S0304405X98000270.

Ben-David, Itzhak, Francesco A. Franzoni, and Rabih Moussawi (2018). "Do ETFs Increase Volatility?" In: *The Journal of Finance* 73.6, pp. 2471–2535. ISSN: 00221082. DOI: r. URL: http://doi.wiley.com/10.1111/jofi.12727.

Ben-David, Itzhak, Francesco Franzoni, and Rabih Moussawi (2017). "Exchange-Traded Funds". In: *Annual Review of Financial Economics* 9.1, pp. 169–189. ISSN: 1941-1367. DOI: 10.1146/annurev-financial-110716-032538. URL: https://doi.org/10.1146/annurev-financial-%20http://www.annualreviews.org/doi/10.1146/annurev-financial-110716-032538.

Berk, Jonathan B. (1995). "A Critique of Size-Related Anomalies". In: *Review of Financial Studies* 8.2, pp. 275–286. ISSN: 0893-9454. DOI: 10.1093/rfs/8.2.275. URL: https://academic.oup.com/rfs/article-lookup/doi/10.1093/rfs/8.2.275.

Bessembinder, Hendrik and Paul J. Seguin (1992). "Futures-Trading Activity and Stock Price Volatility". In: *The Journal of Finance* 47.5, p. 2015. ISSN: 00221082. DOI: 10.2307/2329008. URL: https://www.jstor.org/stable/2329008?origin=crossref.

Bhattacharya, Ayan and Maureen O'Hara (2017). "Can ETFs Increase Market Fragility? Effect of Information Linkages in ETF Markets". In: *SSRN Electronic Journal* April, pp. 1–55. ISSN: 1556-5068. DOI: 10.2139/ssrn.2740699. URL: https://www.ssrn.com/abstract=2740699.

Bogle, John (2012). "The Economic Role Of The Investment Company. Excerpts from the author's 1951 college thesis". In: *Journal of Indexes* April, pp. 10–19.

Browne, Reginald M. (2017). " Exchange-Traded Funds (ETFs) ". In: *The Capital Markets*. DOI: 10.1002/9781119220589.ch27.

Carhart, Mark M. (1997). "On Persistence in Mutual Fund Performance". In: *The Journal of Finance* 52.1, p. 57. ISSN: 00221082. DOI: 10.2307/2329556. URL: http://www.jstor.org/stable/2329556?origin=crossref.

Charupat, Narat and Peter Miu (2013). "Recent developments in exchange-traded fund literature". In: *Managerial Finance* 39.5. Ed. by Peter Miu, pp. 427–443. ISSN: 0307-4358. DOI: 10.1108/03074351311313816. URL: https://www.emeraldinsight.com/doi/10.1108/03074351311313816.

Chinco, Alexander and Vyacheslav Fos (2016). "The Sound of Many Funds Rebalancing". In: *SSRN Electronic Journal*, pp. 1–56. ISSN: 1556-5068. DOI: 10.2139/ssrn.2764941. URL: https://www.ssrn.com/abstract=2764941.

Corwin, Shane A. and Paul Schultz (2012). "A Simple Way to Estimate Bid-Ask Spreads from Daily High and Low Prices". In: *The Journal of Finance* 67.2, pp. 719–760. ISSN: 00221082. DOI: 10.1111/j.1540-6261.2012.01729.x. URL: https://www.jstor.org/stable/

41419709%20https://doi.org/10.1111/j.1540-6261.2012.01729.x%20http://doi.wiley.com/10.1111/j.1540-6261.2012.01729.x.

Da, Zhi and Sophie Shive (2018). "Exchange traded funds and asset return correlations". In: *European Financial Management* 24.1, pp. 136–168. ISSN: 1468036X. DOI: 10.1111/eufm.12137.

Daniel, Kent D., David Hirshleifer, and Avanidhar Subrahmanyam (2001). "Overconfidence, arbitrage, and equilibrium asset pricing". In: *Journal of Finance* 56.3, pp. 921–965. ISSN: 00221082. DOI: 10.1111/0022-1082.00350. URL: http://doi.wiley.com/10.1111/0022-1082.00350.

Daniel, Kent, David Hirshleifer, and Avanidhar Subrahmanyam (1998). "Investor Psychology and Security Market Under- and Overreactions". In: *The Journal of Finance* 53.6, pp. 1839–1885. ISSN: 00221082. DOI: 10.1111/0022-1082.00077. URL: http://doi.wiley.com/10.1111/0022-1082.00077.

Danthine, Jean-Pierre (1978). "Information, futures prices, and stabilizing speculation". In: *Journal of Economic Theory* 17.1, pp. 79–98. ISSN: 00220531. DOI: 10.1016/0022-0531(78)90124-2. URL: https://linkinghub.elsevier.com/retrieve/pii/0022053178901242.

De Bondt, Werner F. and Richard Thaler (1994). "Financial Decision-Making in Markets and Firms: A Behavioral Perspective". Cambridge, MA. URL: http://www.nber.org/papers/w4777.pdf.

Easley, David et al. (2018). "The Active World of Passive Investing". In: *SSRN Electronic Journal*, p. 35. ISSN: 1556-5068. DOI: 10.2139/ssrn.3220842. URL: https://ssrn.com/abstract=3220842%20https://www.ssrn.com/abstract=3220842.

Economic Sciences Prize Committee of the Royal Swedish Academy of Sciences (2012). *Scientific Background on the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2013. Understanding Asset Prices.* Tech. rep. Stockholm: The Royal Swedish Academy of Sciences. URL: https://www.nobelprize.org/prizes/economic-sciences/2013/advanced-information/.

Evans, Richard B. et al. (2017). "ETF Short Interest and Failures-to-Deliver: Naked Short-Selling or Operational Shorting?" In: *SSRN Electronic Journal*. ISSN: 1556-5068. DOI: 10.2139/ssrn.2961954. URL: https://www.ssrn.com/abstract=2961954.

Fama, Eugene F. and Kenneth R. French (1992). "The Cross-Section of Expected Stock Returns". In: *The Journal of Finance* 47.2, pp. 427–465. ISSN: 00221082. DOI: 10.1111/j.1540-6261.1992.tb04398.x. URL: http://doi.wiley.com/10.1111/j.1540-6261.1992.tb04398.x.

Fama, Eugene F. and Kenneth R. French (1993). "Common risk factors in the returns on stocks and bonds". In: *Journal of Financial Economics* 33.1, pp. 3–56. ISSN: 0304405X. DOI: 10.1016/0304-405X(93)90023-5.

– (2006). "Profitability, investment and average returns". In: *Journal of Financial Economics* 82.3, pp. 491–518. ISSN: 0304405X. DOI: 10.1016/j.jfineco.2005.09.009. URL: https://linkinghub.elsevier.com/retrieve/pii/S0304405X06001164.

Galloway, Tina M. and James M. Miller (1997). "Index Futures Trading and Stock Return Volatility: Evidence from the Introduction of Mid Cap 400 Index Futures". In: *The Financial Review* 32.4, pp. 845–866. ISSN: 0732-8516. DOI: 10.1111/j.1540-6288.1997.tb00913.x. URL: http://doi.wiley.com/10.1111/j.1540-6288.1997.tb00913.x.

Gârleanu, Nicolae, Stavros Panageas, and Jianfeng Yu (2012). "Technological Growth and Asset Pricing". In: *Journal of Finance* 67.4, pp. 1265–1292. ISSN: 00221082. DOI: 10.1111/j.1540-6261.2012.01747.x. URL: http://doi.wiley.com/10.1111/j.1540-6261.2012.01747.x.

Glosten, Lawrence R. and Paul R. Milgrom (1985). "Bid, ask and transaction prices in a specialist market with heterogeneously informed traders". In: *Journal of Financial Economics* 14.1, pp. 71–100. ISSN: 0304405X. DOI: 10.1016/0304-405X(85)90044-3. URL: https://linkinghub.elsevier.com/retrieve/pii/0304405X85900443.

Goyal, Amit and Sunil Wahal (2015). "Is Momentum an Echo?" In: *Journal of Financial and Quantitative Analysis* 50.6, pp. 1237–1267. ISSN: 0022-1090. DOI: 10.1017/S0022109015000575. URL: http://www.journals.cambridge.org/abstract%7B%5C_%7DS0022109015000575%20https://www.cambridge.org/core/product/identifier/S0022109015000575/type/journal%7B%5C_%7Darticle.

Graham, Benjamin and David Dodd (1934). *Security analysis.* 1st ed. New York: McGraw-Hill. ISBN: 0070244960. DOI: 10.1036/0071592539. URL: http://cds.cern.ch/record/105688.

Grossman, Sanford J. (1988). "An Analysis of the Implications for Stock and Futures Price Volatility of Program Trading and Dynamic Hedging Strategies". In: *The Journal of Business* 61.3, pp. 275–298. URL: http://www.jstor.org/stable/2353040.

Grossman, Sanford J. and Joseph E. Stiglitz (1980). "On the Impossibility of Informationally Efficient Markets". In: *American Economic Review* 70.3, pp. 393–408. URL: http://www.jstor.org/stable/1805228.

Hong, Harrison and Jeremy C. Stein (1999). "A Unified Theory of Underreaction, Momentum Trading, and Overreaction in Asset Markets". In: *The Journal of Finance* 54.6, pp. 2143–2184.

ISSN: 00221082. DOI: 10.1111/0022-1082.00184. URL: http://doi.wiley.com/10.1111/0022-1082.00184.

Israeli, Doron, Charles M.C. Lee, and Suhas A. Sridharan (2017). "Is there a dark side to exchange traded funds? An information perspective". In: *Review of Accounting Studies* 22.3, pp. 1048–1083. ISSN: 13806653. DOI: 10.1007/s11142-017-9400-8.

Jegadeesh, Narasimhan and Sheridan Titman (1993). "Returns to buying winners and selling losers:implications for stock market efficiency". In: *The Journal of finance* 48.1, pp. 65–91.

Johnson, Timothy C. (2002). "Rational Momentum Effects". In: *The Journal of Finance* 57.2, pp. 585–608. ISSN: 00221082. DOI: 10.1111/1540-6261.00435. URL: http://doi.wiley.com/10.1111/1540-6261.00435.

Jordan, J.S (1982). "A dynamic model of expectations equilibrium". In: *Journal of Economic Theory* 28.2, pp. 235–254. ISSN: 00220531. DOI: 10.1016/0022-0531(82)90060-6. URL: https://linkinghub.elsevier.com/retrieve/pii/0022053182900606.

Koch, Andrew, Stefan Ruenzi, and Laura Starks (2016). "Commonality in liquidity: A demand-side explanation". In: *Review of Financial Studies* 29.8, pp. 1943–1974. ISSN: 14657368. DOI: 10.1093/rfs/hhw026.

Koehler, Derek J. and Nigel Harvey (2004). *Blackwell Handbook of Judgment and Decision Making*. Ed. by Derek J. Koehler and Nigel Harvey. Malden, MA, USA: Blackwell Publishing Ltd. Chap. 26, pp. 1–664. ISBN: 9780470752937. DOI: 10.1002/9780470752937. URL: http://www.blackwellpublishing.com%20http://doi.wiley.com/10.1002/9780470752937.

Krause, Timothy, Sina Ehsani, and Donald Lien (2014). "Exchange-traded funds, liquidity and volatility". In: *Applied Financial Economics* 24.24, pp. 1617–1630. ISSN: 0960-3107. DOI: 10.1080/09603107.2014.941530. URL: http://www.tandfonline.com/doi/abs/10.1080/09603107.2014.941530.

Kyle, Albert S. (1985). "Continuous Auctions and Insider Trading". In: *Econometrica* 53.6, p. 1315. ISSN: 00129682. DOI: 10.2307/1913210. URL: https://www.jstor.org/stable/1913210%20https://www.jstor.org/stable/1913210?origin=crossref.

Leippold, Markus, Lujing Su, and Alexandre Ziegler (2015). "Do Index Futures and ETFs Affect Stock Return Correlations?" In: *SSRN Electronic Journal* 41.44. ISSN: 1556-5068. DOI: 10.2139/ssrn.2620955. URL: http://www.ssrn.com/abstract=2620955.

Novy-Marx, Robert (2012). "Is momentum really momentum?" In: *Journal of Financial Economics* 103.3, pp. 429–453. ISSN: 0304405X. DOI: 10.1016/j.jfineco.2011.05.003. URL: http://dx.doi.org/10.1016/j.jfineco.2011.05.003.

Novy-Marx, Robert (2013). “The other side of value: The gross profitability premium”. In: *Journal of Financial Economics* 108.1, pp. 1–28. ISSN: 0304405X. DOI: 10.1016/j.jfineco.2013.01.003. URL: https://www.sciencedirect.com/science/article/pii/S0304405X13000044%20https://linkinghub.elsevier.com/retrieve/pii/S0304405X13000044.

Petajisto, Antti (2017). “Inefficiencies in the pricing of exchange-traded funds”. In: *Financial Analysts Journal* 73.1, pp. 24–54. ISSN: 0015198X. DOI: 10.2469/faj.v73.n1.7.

Rosenberg, Barr, Kenneth Reid, and Ronald Lanstein (1985). “Persuasive evidence of market inefficiency”. In: *The Journal of Portfolio Management* 11.3, pp. 9–16. ISSN: 0095-4918. DOI: 10.3905/jpm.1985.409007. URL: http://jpm.iijournals.com/lookup/doi/10.3905/jpm.1985.409007.

Rubinstein, Mark (1989). “Market Basket Alternatives”. In: *Research Program in Finance Working Papers* 45.5, pp. 20–29. URL: https://www.jstor.org/stable/4479255%20http://ideas.repec.org/p/ucb/calbrf/rpf-187.html.

Sagi, Jacob S. and Mark S. Seasholes (2007). “Firm-specific attributes and the cross-section of momentum”. In: *Journal of Financial Economics* 84.2, pp. 389–434. ISSN: 0304405X. DOI: 10.1016/j.jfineco.2006.02.002. URL: https://linkinghub.elsevier.com/retrieve/pii/S0304405X06002066.

Subrahmanyam, Avanidhar, Tarun Chordia, and Richard Roll (2000). “Commonality in liquidity”. In: *Journal of Financial Economics* 56, pp. 3–28. ISSN: 0304405X. DOI: 10.1016/S0304-405X(99)00057-4.

Tversky, A and D Kahneman (1974). “Judgment under Uncertainty: Heuristics and Biases”. In: *Science* 185.4157, pp. 1124–1131. ISSN: 0036-8075. DOI: 10.1126/science.185.4157.1124. URL: http://www.ncbi.nlm.nih.gov/pubmed/17835457%20http://www.sciencemag.org/cgi/doi/10.1126/science.185.4157.1124.

A. Institutional background

As a matter of conciseness, the description of the legal framework in which ETFs operate and the institutional design (market participants) in Section 3 (p.8) has been kept at its shortest in the main body of this paper. This appendix aims at providing a more exhaustive overview of these topics.

B. Econometric definitions

Due to the use of a (dynamic) panel regression method throughout the empirical analysis, it is useful to discuss specifically certain statistics given in summary outputs, to be found in Section 8 (p.37), as well as in the extended outputs gathered (as a matter of brevity) in an appendix section (Appendix C, p.49.).

C. Detailed estimation results

One of the many advantages provided through the statistical functions library, called `linearmodels`[19], accessed through the Python 3.7.3 programming language is the ability to extract summarized as well as extended regression estimates and related statistics. Since it would have been untractable to provide extended sample descriptions (the upper panel in following tables) and more insight into regression coefficients estimates (the lower panels), those extended results have been gathered in the current supporting section.

C.1. ETF ownership and underlying stocks' volatility

This subsection constitutes additional material about variants of the model attempting to explain stocks volatility, summarized and discussed in subsection 8.2 (p.40).

C.1.1. U.S. stocks sample

[19]The `linearmodels`, whose current stable version at the time of writing is 4.5, is developped and maintained by Prof. Kevin Sheppard, who provides the source code free for use at https://github.com/bashtage/linearmodels. I rely on the documentation included in the docstring as well as on https://bashtage.github.io/linearmodels/doc/

Dep. Variable:	Volatility	**R-squared:**	0.1582
Estimator:	PanelOLS	**R-squared (Between):**	0.7685
No. Observations:	302238	**R-squared (Within):**	0.1350
Date:	Wed, Jun 12 2019	**R-squared (Overall):**	0.2710
Time:	16:08:02	**Log-likelihood**	6.586e+05
Cov. Estimator:	Driscoll-Kraay		
		F-statistic:	4685.2
Entities:	2987	**P-value**	0.0000
Avg Obs:	101.18	**Distribution:**	F(12,299066)
Min Obs:	1.0000		
Max Obs:	174.00	**F-statistic (robust):**	196.78
		P-value	0.0000
Time periods:	174	**Distribution:**	F(12,299066)
Avg Obs:	1737.0		
Min Obs:	984.00		
Max Obs:	2794.0		

	Parameter	**Std. Err.**	**T-stat**	**P-value**	**Lower CI**	**Upper CI**
Intercept	0.0507	0.0066	7.7298	0.0000	0.0378	0.0635
PctSharesHeldETF	0.0398	0.0065	6.1376	0.0000	0.0271	0.0525
np.log(CompanyMarketCap__1lag)	-0.0019	0.0003	-6.3341	0.0000	-0.0024	-0.0013
InvClose__1lag	0.0247	0.0033	7.6046	0.0000	0.0184	0.0311
AmihudRatio__1lag	3.4497	1.2936	2.6667	0.0077	0.9143	5.9851
PctBidAskSpread__1lag	0.1795	0.0372	4.8311	0.0000	0.1067	0.2523
BookToMarketRatio__1lag	7.42e-07	1.781e-07	4.1651	0.0000	3.928e-07	1.091e-06
RetPast12to7M__1lag	6.481e-06	6.382e-06	1.0155	0.3099	-6.028e-06	1.899e-05
GrossProfitability__1lag	-0.0006	0.0002	-2.7450	0.0061	-0.0010	-0.0002
Volatility__1lag	0.1399	0.0204	6.8513	0.0000	0.0999	0.1800
Volatility__2lag	0.1613	0.0097	16.642	0.0000	0.1423	0.1803
Volatility__3lag	0.1210	0.0090	13.424	0.0000	0.1034	0.1387
Volatility__4lag	0.0796	0.0079	10.014	0.0000	0.0640	0.0952

F-test for Poolability: 49.051
P-value: 0.0000
Distribution: F(3159,299066)

Included effects: Entity, Time

Dep. Variable:	Volatility	**R-squared:**	0.1598
Estimator:	PanelOLS	**R-squared (Between):**	0.7643
No. Observations:	296629	**R-squared (Within):**	0.1339
Date:	Wed, Jun 12 2019	**R-squared (Overall):**	0.2707
Time:	16:08:12	**Log-likelihood**	6.475e+05
Cov. Estimator:	Driscoll-Kraay		
		F-statistic:	4295.2
Entities:	2933	**P-value**	0.0000
Avg Obs:	101.14	**Distribution:**	F(13,293510)
Min Obs:	1.0000		
Max Obs:	174.00	**F-statistic (robust):**	221.90
		P-value	0.0000
Time periods:	174	**Distribution:**	F(13,293510)
Avg Obs:	1704.8		
Min Obs:	965.00		
Max Obs:	2745.0		

	Parameter	**Std. Err.**	**T-stat**	**P-value**	**Lower CI**	**Upper CI**
Intercept	0.0604	0.0073	8.3118	0.0000	0.0462	0.0747
PctSharesHeldETF	-0.3879	0.0537	-7.2262	0.0000	-0.4931	-0.2827
np.log(CompanyMarketCap__1lag)	-0.0023	0.0003	-7.1990	0.0000	-0.0030	-0.0017
PctSharesHeldETF:np.log(CompanyMarketCap__1lag)	0.0207	0.0027	7.7590	0.0000	0.0155	0.0259
InvClose__1lag	0.0239	0.0032	7.4485	0.0000	0.0176	0.0302
AmihudRatio__1lag	3.4673	1.2888	2.6904	0.0071	0.9413	5.9932
PctBidAskSpread__1lag	0.1746	0.0370	4.7153	0.0000	0.1020	0.2471
BookToMarketRatio__1lag	7.429e-07	1.76e-07	4.2213	0.0000	3.98e-07	1.088e-06
RetPast12to1M__1lag	-0.0005	0.0002	-2.0383	0.0415	-0.0009	-1.739e-05
GrossProfitability__1lag	-0.0004	0.0002	-2.6128	0.0090	-0.0007	-0.0001
Volatility__1lag	0.1378	0.0205	6.7333	0.0000	0.0977	0.1779
Volatility__2lag	0.1599	0.0101	15.889	0.0000	0.1402	0.1796
Volatility__3lag	0.1219	0.0091	13.362	0.0000	0.1040	0.1398
Volatility__4lag	0.0801	0.0083	9.6749	0.0000	0.0639	0.0963

F-test for Poolability: 49.642
P-value: 0.0000
Distribution: F(3105,293510)

Included effects: Entity, Time

Dep. Variable:	Volatility	**R-squared:**	0.1584
Estimator:	PanelOLS	**R-squared (Between):**	0.7685
No. Observations:	302234	**R-squared (Within):**	0.1349
Date:	Wed, Jun 12 2019	**R-squared (Overall):**	0.2709
Time:	16:08:46	**Log-likelihood**	6.586e+05
Cov. Estimator:	Driscoll-Kraay		
		F-statistic:	3753.2
Entities:	2987	**P-value**	0.0000
Avg Obs:	101.18	**Distribution:**	F(15,299059)
Min Obs:	1.0000		
Max Obs:	174.00	**F-statistic (robust):**	216.76
		P-value	0.0000
Time periods:	174	**Distribution:**	F(15,299059)
Avg Obs:	1737.0		
Min Obs:	984.00		
Max Obs:	2794.0		

	Parameter	**Std. Err.**	**T-stat**	**P-value**	**Lower CI**	**Upper CI**
Intercept	0.0512	0.0066	7.8036	0.0000	0.0383	0.0641
PctSharesHeldETF	0.0410	0.0064	6.4401	0.0000	0.0285	0.0535
np.log(CompanyMarketCap__1lag)	-0.0019	0.0003	-6.4261	0.0000	-0.0025	-0.0013
InvClose__1lag	0.0247	0.0032	7.6653	0.0000	0.0184	0.0311
AmihudRatio__1lag	3.4512	1.2946	2.6658	0.0077	0.9138	5.9886
PctBidAskSpread__1lag	0.1792	0.0372	4.8222	0.0000	0.1063	0.2520
BookToMarketRatio__1lag	2.851e-06	6.618e-07	4.3085	0.0000	1.554e-06	4.148e-06
RetPast12to7M__1lag	7.771e-06	7.004e-06	1.1096	0.2672	-5.956e-06	2.15e-05
GrossProfitability__1lag	-0.0005	0.0002	-2.7599	0.0058	-0.0009	-0.0002
PctSharesHeldOtherMutual	-0.0005	0.0002	-2.9612	0.0031	-0.0008	-0.0002
PctSharesHeldPension	1.1527	0.3061	3.7651	0.0002	0.5526	1.7527
PctSharesHeldHedge	-0.0306	0.0086	-3.5588	0.0004	-0.0474	-0.0137
Volatility__1lag	0.1400	0.0204	6.8684	0.0000	0.1001	0.1800
Volatility__2lag	0.1612	0.0097	16.600	0.0000	0.1421	0.1802
Volatility__3lag	0.1210	0.0090	13.439	0.0000	0.1033	0.1386
Volatility__4lag	0.0795	0.0080	10.001	0.0000	0.0639	0.0951

F-test for Poolability: 49.072
P-value: 0.0000
Distribution: F(3159,299059)

Included effects: Entity, Time

Dep. Variable:	Volatility	**R-squared:**	0.1670
Estimator:	PanelOLS	**R-squared (Between):**	0.7571
No. Observations:	296011	**R-squared (Within):**	0.1354
Date:	Wed, Jun 12 2019	**R-squared (Overall):**	0.2661
Time:	16:08:59	**Log-likelihood**	-3.025e+05
Cov. Estimator:	Driscoll-Kraay		
		F-statistic:	3915.6
Entities:	2933	**P-value**	0.0000
Avg Obs:	100.92	**Distribution:**	F(15,292890)
Min Obs:	1.0000		
Max Obs:	174.00	**F-statistic (robust):**	178.01
		P-value	0.0000
Time periods:	174	**Distribution:**	F(15,292890)
Avg Obs:	1701.2		
Min Obs:	964.00		
Max Obs:	2735.0		

	Parameter	**Std. Err.**	**T-stat**	**P-value**	**Lower CI**	**Upper CI**
PctSharesHeldETF__std	0.0392	0.0056	7.0103	0.0000	0.0283	0.0502
CompanyMarketCap__1lag	0.0117	0.0031	3.8175	0.0001	0.0057	0.0177
InvClose__1lag	0.0902	0.0061	14.883	0.0000	0.0783	0.1021
AmihudRatio__1lag	0.0297	0.0085	3.4946	0.0005	0.0130	0.0464
PctBidAskSpread__1lag	0.0720	0.0120	6.0207	0.0000	0.0486	0.0954
BookToMarketRatio__1lag	0.0119	0.0038	3.1578	0.0016	0.0045	0.0193
RetPast12to7M__1lag	0.0004	0.0006	0.6446	0.5192	-0.0008	0.0016
GrossProfitability__1lag	-0.0119	0.0031	-3.7891	0.0002	-0.0181	-0.0058
PctSharesHeldOtherMutual__std	-0.0030	0.0032	-0.9212	0.3569	-0.0093	0.0034
PctSharesHeldPension__std	0.0071	0.0019	3.7474	0.0002	0.0034	0.0109
PctSharesHeldHedge__std	-0.0003	0.0010	-0.3096	0.7569	-0.0023	0.0017
Volatility__1lag	0.0902	0.0114	7.9095	0.0000	0.0679	0.1126
Volatility__2lag	0.1694	0.0096	17.597	0.0000	0.1505	0.1882
Volatility__3lag	0.1376	0.0105	13.159	0.0000	0.1171	0.1581
Volatility__4lag	0.0910	0.0097	9.3632	0.0000	0.0720	0.1101

F-test for Poolability: 53.110
P-value: 0.0000
Distribution: F(3105,292890)

Included effects: Entity, Time

C.1.2. International sample

C.2. ETF ownership and underlying stocks' liquidity

This subsection constitutes additional material about variants of the models using several liquidity proxies as dependent variables, summarized and discussed in subsection 8.3 (p.40).

C.2.1. U.S. stocks sample

Dep. Variable:	PctBidAskSpread	**R-squared:**	0.0512
Estimator:	PanelOLS	**R-squared (Between):**	0.5612
No. Observations:	335350	**R-squared (Within):**	0.0661
Date:	Wed, Jun 12 2019	**R-squared (Overall):**	0.1882
Time:	16:09:14	**Log-likelihood**	8.498e+05
Cov. Estimator:	Driscoll-Kraay		
		F-statistic:	4465.5
Entities:	3995	**P-value**	0.0000
Avg Obs:	83.942	**Distribution:**	F(4,331214)
Min Obs:	1.0000		
Max Obs:	138.00	**F-statistic (robust):**	88.621
		P-value	0.0000
Time periods:	138	**Distribution:**	F(4,331214)
Avg Obs:	2430.1		
Min Obs:	1474.0		
Max Obs:	3939.0		

	Parameter	**Std. Err.**	**T-stat**	**P-value**	**Lower CI**	**Upper CI**
Intercept	0.1054	0.0067	15.787	0.0000	0.0924	0.1185
PctSharesHeldETF	0.0061	0.0077	0.7923	0.4282	-0.0090	0.0212
np.log(CompanyMarketCap__1lag)	-0.0049	0.0003	-14.639	0.0000	-0.0056	-0.0042
BookToMarketRatio__1lag	-4.277e-08	4.747e-08	-0.9008	0.3677	-1.358e-07	5.028e-08
Volatility__1lag	0.0500	0.0049	10.104	0.0000	0.0403	0.0597

F-test for Poolability: 9.5234
P-value: 0.0000
Distribution: F(4131,331214)

Included effects: Entity, Time

Dep. Variable:	PctBidAskSpread	**R-squared:**	0.0512
Estimator:	PanelOLS	**R-squared (Between):**	0.5618
No. Observations:	335346	**R-squared (Within):**	0.0663
Date:	Wed, Jun 12 2019	**R-squared (Overall):**	0.1885
Time:	16:09:23	**Log-likelihood**	8.498e+05
Cov. Estimator:	Driscoll-Kraay		
		F-statistic:	2555.6
Entities:	3995	**P-value**	0.0000
Avg Obs:	83.941	**Distribution:**	F(7,331207)
Min Obs:	1.0000		
Max Obs:	138.00	**F-statistic (robust):**	61.762
		P-value	0.0000
Time periods:	138	**Distribution:**	F(7,331207)
Avg Obs:	2430.0		
Min Obs:	1474.0		
Max Obs:	3939.0		

	Parameter	**Std. Err.**	**T-stat**	**P-value**	**Lower CI**	**Upper CI**
Intercept	0.1057	0.0067	15.743	0.0000	0.0925	0.1188
PctSharesHeldETF	0.0060	0.0077	0.7832	0.4335	-0.0091	0.0211
np.log(CompanyMarketCap__1lag)	-0.0049	0.0003	-14.596	0.0000	-0.0056	-0.0043
BookToMarketRatio__1lag	-4.038e-08	4.562e-08	-0.8852	0.3760	-1.298e-07	4.903e-08
Volatility__1lag	0.0499	0.0049	10.107	0.0000	0.0402	0.0596
PctSharesHeldOtherMutual	-2.114e-06	4.672e-07	-4.5245	0.0000	-3.03e-06	-1.198e-06
PctSharesHeldPension	0.3943	0.3072	1.2835	0.1993	-0.2078	0.9963
PctSharesHeldHedge	0.0077	0.0044	1.7387	0.0821	-0.0010	0.0164

F-test for Poolability: 9.4898
P-value: 0.0000
Distribution: F(4131,331207)

Included effects: Entity, Time

Dep. Variable:	AmihudNumerator	**R-squared:**	0.0330
Estimator:	PanelOLS	**R-squared (Between):**	0.1873
No. Observations:	436857	**R-squared (Within):**	0.0436
Date:	Wed, Jun 12 2019	**R-squared (Overall):**	0.1276
Time:	16:09:38	**Log-likelihood**	6.731e+05
Cov. Estimator:	Driscoll-Kraay		
		F-statistic:	3688.0
Entities:	4079	**P-value**	0.0000
Avg Obs:	107.10	**Distribution:**	F(4,432600)
Min Obs:	1.0000		
Max Obs:	175.00	**F-statistic (robust):**	216.03
		P-value	0.0000
Time periods:	175	**Distribution:**	F(4,432600)
Avg Obs:	2496.3		
Min Obs:	1539.0		
Max Obs:	3986.0		

	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
Intercept	0.3438	0.0114	30.192	0.0000	0.3215	0.3661
PctSharesHeldETF	0.1870	0.0241	7.7614	0.0000	0.1398	0.2342
np.log(CompanyMarketCap__1lag)	-0.0153	0.0006	-26.847	0.0000	-0.0165	-0.0142
BookToMarketRatio__1lag	2.894e-07	2.556e-07	1.1320	0.2576	-2.116e-07	7.904e-07
AmihudDenominator	1.321e-11	3.147e-12	4.1972	0.0000	7.039e-12	1.937e-11

F-test for Poolability: 67.429
P-value: 0.0000
Distribution: F(4252,432600)

Included effects: Entity, Time

Dep. Variable:	AmihudNumerator	**R-squared:**	0.0330
Estimator:	PanelOLS	**R-squared (Between):**	0.1873
No. Observations:	436852	**R-squared (Within):**	0.0436
Date:	Wed, Jun 12 2019	**R-squared (Overall):**	0.1276
Time:	16:09:51	**Log-likelihood**	6.73e+05
Cov. Estimator:	Driscoll-Kraay		
		F-statistic:	2950.4
Entities:	4079	**P-value**	0.0000
Avg Obs:	107.10	**Distribution:**	F(5,432594)
Min Obs:	1.0000		
Max Obs:	175.00	**F-statistic (robust):**	195.22
		P-value	0.0000
Time periods:	175	**Distribution:**	F(5,432594)
Avg Obs:	2496.3		
Min Obs:	1539.0		
Max Obs:	3986.0		

	Parameter	**Std. Err.**	**T-stat**	**P-value**	**Lower CI**	**Upper CI**
Intercept	0.3438	0.0114	30.164	0.0000	0.3215	0.3661
PctSharesHeldETF	0.1871	0.0241	7.7638	0.0000	0.1398	0.2343
np.log(CompanyMarketCap__1lag)	-0.0153	0.0006	-26.822	0.0000	-0.0165	-0.0142
BookToMarketRatio__1lag	3.098e-07	2.71e-07	1.1432	0.2530	-2.213e-07	8.409e-07
AmihudDenominator	1.32e-11	3.146e-12	4.1968	0.0000	7.038e-12	1.937e-11
PctSharesHeldOtherMutual	-1.219e-05	2.171e-06	-5.6177	0.0000	-1.645e-05	-7.939e-06

F-test for Poolability: 67.427
P-value: 0.0000
Distribution: F(4252,432594)

Included effects: Entity, Time

C.2.2. International sample

C.3. ETF ownership and concerns about pricing efficiency

This subsection constitutes additional material about variants of the model studying two forms of variance ratios, summarized and discussed in subsection 8.4 (p.40).

C.3.1. U.S. stocks sample

Dep. Variable:	VR	**R-squared:**	0.0040
Estimator:	PanelOLS	**R-squared (Between):**	0.1376
No. Observations:	126851	**R-squared (Within):**	0.0040
Date:	Wed, Jun 12 2019	**R-squared (Overall):**	0.0163
Time:	16:09:53	**Log-likelihood**	-7.813e+04
Cov. Estimator:	Driscoll-Kraay		
		F-statistic:	61.843
Entities:	2966	**P-value**	0.0000
Avg Obs:	42.768	**Distribution:**	F(8,123807)
Min Obs:	1.0000		
Max Obs:	71.000	**F-statistic (robust):**	12.114
		P-value	0.0000
Time periods:	71	**Distribution:**	F(8,123807)
Avg Obs:	1786.6		
Min Obs:	1087.0		
Max Obs:	2836.0		

	Parameter	**Std. Err.**	**T-stat**	**P-value**	**Lower CI**	**Upper CI**
Intercept	0.8735	0.1142	7.6512	0.0000	0.6498	1.0973
PctSharesHeldETF__1lag	-0.1205	0.0856	-1.4081	0.1591	-0.2882	0.0472
np.log(CompanyMarketCap__1lag)	0.0044	0.0054	0.8038	0.4215	-0.0063	0.0150
InvClose__1lag	-0.2002	0.0330	-6.0619	0.0000	-0.2650	-0.1355
AmihudRatio__1lag	-15.135	4.9970	-3.0288	0.0025	-24.929	-5.3407
PctBidAskSpread__1lag	-1.2190	0.3412	-3.5732	0.0004	-1.8877	-0.5503
BookToMarketRatio__1lag	-1.257e-05	8.787e-06	-1.4303	0.1526	-2.979e-05	4.655e-06
RetPast12to7M__1lag	-0.0006	0.0006	-0.9013	0.3675	-0.0018	0.0007
GrossProfitability__1lag	0.0064	0.0050	1.2767	0.2017	-0.0034	0.0162

F-test for Poolability: 7.6661
P-value: 0.0000
Distribution: F(3035,123807)

Included effects: Entity, Time

Dep. Variable:	VR	**R-squared:**	0.0041
Estimator:	PanelOLS	**R-squared (Between):**	0.1385
No. Observations:	126847	**R-squared (Within):**	0.0042
Date:	Wed, Jun 12 2019	**R-squared (Overall):**	0.0165
Time:	16:09:55	**Log-likelihood**	-7.812e+04
Cov. Estimator:	Driscoll-Kraay		
		F-statistic:	46.323
Entities:	2966	**P-value**	0.0000
Avg Obs:	42.767	**Distribution:**	F(11,123800)
Min Obs:	1.0000		
Max Obs:	71.000	**F-statistic (robust):**	13.710
		P-value	0.0000
Time periods:	71	**Distribution:**	F(11,123800)
Avg Obs:	1786.6		
Min Obs:	1087.0		
Max Obs:	2836.0		

	Parameter	**Std. Err.**	**T-stat**	**P-value**	**Lower CI**	**Upper CI**
Intercept	0.8579	0.1154	7.4328	0.0000	0.6317	1.0841
PctSharesHeldETF__1lag	-0.1005	0.0862	-1.1660	0.2436	-0.2694	0.0684
np.log(CompanyMarketCap__1lag)	0.0052	0.0055	0.9458	0.3443	-0.0056	0.0160
InvClose__1lag	-0.1986	0.0335	-5.9225	0.0000	-0.2643	-0.1328
AmihudRatio__1lag	-15.117	4.9926	-3.0279	0.0025	-24.903	-5.3317
PctBidAskSpread__1lag	-1.2179	0.3418	-3.5633	0.0004	-1.8878	-0.5480
BookToMarketRatio__1lag	6.777e-05	2.065e-05	3.2826	0.0010	2.731e-05	0.0001
RetPast12to7M__1lag	-0.0013	0.0006	-2.2311	0.0257	-0.0024	-0.0002
GrossProfitability__1lag	0.0062	0.0049	1.2635	0.2064	-0.0034	0.0159
PctSharesHeldOtherMutual__1lag	-0.0086	0.0023	-3.6908	0.0002	-0.0132	-0.0041
PctSharesHeldPension__1lag	-3.7029	6.9063	-0.5362	0.5918	-17.239	9.8333
PctSharesHeldHedge__1lag	-0.1145	0.1826	-0.6268	0.5308	-0.4724	0.2435

F-test for Poolability: 7.6650
P-value: 0.0000
Distribution: F(3035,123800)

Included effects: Entity, Time

Dep. Variable:	absVR	**R-squared:**	0.0079
Estimator:	PanelOLS	**R-squared (Between):**	0.2814
No. Observations:	126851	**R-squared (Within):**	0.0106
Date:	Wed, Jun 12 2019	**R-squared (Overall):**	0.0318
Time:	16:09:57	**Log-likelihood**	-3.697e+04
Cov. Estimator:	Driscoll-Kraay		
		F-statistic:	123.79
Entities:	2966	**P-value**	0.0000
Avg Obs:	42.768	**Distribution:**	F(8,123807)
Min Obs:	1.0000		
Max Obs:	71.000	**F-statistic (robust):**	37.102
		P-value	0.0000
Time periods:	71	**Distribution:**	F(8,123807)
Avg Obs:	1786.6		
Min Obs:	1087.0		
Max Obs:	2836.0		

	Parameter	**Std. Err.**	**T-stat**	**P-value**	**Lower CI**	**Upper CI**
Intercept	0.8523	0.0694	12.280	0.0000	0.7163	0.9884
PctSharesHeldETF_1lag	0.0096	0.0646	0.1484	0.8820	-0.1170	0.1362
np.log(CompanyMarketCap_1lag)	-0.0248	0.0033	-7.5202	0.0000	-0.0312	-0.0183
InvClose_1lag	0.1276	0.0231	5.5223	0.0000	0.0823	0.1729
AmihudRatio_1lag	9.9267	3.1583	3.1431	0.0017	3.7365	16.117
PctBidAskSpread_1lag	0.9314	0.1667	5.5878	0.0000	0.6047	1.2580
BookToMarketRatio_1lag	-3.498e-06	6.733e-06	-0.5195	0.6034	-1.669e-05	9.699e-06
RetPast12to7M_1lag	0.0003	0.0005	0.5981	0.5497	-0.0006	0.0012
GrossProfitability_1lag	0.0011	0.0047	0.2252	0.8218	-0.0082	0.0104

F-test for Poolability: 4.7953
P-value: 0.0000
Distribution: F(3035,123807)

Included effects: Entity, Time

Dep. Variable:	absVR	**R-squared:**	0.0081
Estimator:	PanelOLS	**R-squared (Between):**	0.2820
No. Observations:	126847	**R-squared (Within):**	0.0106
Date:	Wed, Jun 12 2019	**R-squared (Overall):**	0.0320
Time:	16:09:59	**Log-likelihood**	-3.696e+04
Cov. Estimator:	Driscoll-Kraay		
		F-statistic:	91.734
Entities:	2966	**P-value**	0.0000
Avg Obs:	42.767	**Distribution:**	F(11,123800)
Min Obs:	1.0000		
Max Obs:	71.000	**F-statistic (robust):**	41.510
		P-value	0.0000
Time periods:	71	**Distribution:**	F(11,123800)
Avg Obs:	1786.6		
Min Obs:	1087.0		
Max Obs:	2836.0		

	Parameter	**Std. Err.**	**T-stat**	**P-value**	**Lower CI**	**Upper CI**
Intercept	0.8532	0.0699	12.208	0.0000	0.7162	0.9902
PctSharesHeldETF__1lag	0.0218	0.0647	0.3373	0.7359	-0.1050	0.1487
np.log(CompanyMarketCap__1lag)	-0.0248	0.0033	-7.4640	0.0000	-0.0313	-0.0183
InvClose__1lag	0.1277	0.0229	5.5778	0.0000	0.0829	0.1726
AmihudRatio__1lag	9.9488	3.1629	3.1455	0.0017	3.7496	16.148
PctBidAskSpread__1lag	0.9286	0.1657	5.6027	0.0000	0.6037	1.2534
BookToMarketRatio__1lag	4.433e-05	1.117e-05	3.9695	0.0001	2.244e-05	6.622e-05
RetPast12to7M__1lag	-9.153e-05	0.0004	-0.2511	0.8017	-0.0008	0.0006
GrossProfitability__1lag	0.0012	0.0047	0.2454	0.8062	-0.0081	0.0104
PctSharesHeldOtherMutual__1lag	-0.0053	0.0013	-3.9260	0.0001	-0.0079	-0.0026
PctSharesHeldPension__1lag	11.292	2.7230	4.1468	0.0000	5.9547	16.629
PctSharesHeldHedge__1lag	-0.1294	0.1450	-0.8927	0.3720	-0.4136	0.1548

F-test for Poolability: 4.7965
P-value: 0.0000
Distribution: F(3035,123800)

Included effects: Entity, Time

C.3.2. International sample