Essays on Exchange-Traded Funds

by

Hari Rozental

Thesis

Submitted to the University of Warwick

in partial fulfilment of the requirements

for admission to the degree of

Doctor of Philosophy

Warwick Business School

University of Warwick

September 2019

Contents

List of Tables

List of Figures

Acknowledgments

I would like to express my gratitude to my wife and my parents for their enormous support and for believing in me all these years. I am grateful to my supervisors Arie E. Gozluklu and Ilias Filippou for their unbelievable patience, invaluable pieces of advice and for constant pressure to be better. Finally, I would like to thank Jay Cao, Roman Kozhan, Michael Moore, Pedro Matos, Philippe Mueller, Alvaro Pedrasa, Alexi Savov, Alex Stremme, Thomas Williams, Chendi Zhang, as well as seminar participants at China International Conference in Finance 2019, 27^{th} Finance Forum (Madrid), The Hebrew University of Jerusalem, Fulcrum Asset Management, Jupiter Asset Management-Bristol University PhD conference, Warwick Business School and at the Lancaster-Warwick workshop for very useful comments.

Declarations

I declare that this thesis has not been submitted for a degree at another institution. I further declair that Chapter 1 and 3 are entirely my own work and Chapter 2 is co-authored with Arie E. Gozluklu and Ilias Filippou.

Hari Rozental

September, 2019

Abstract

This thesis investigates the consequences of exchange-traded fund industry growth. In particular, I study the ETF arbitrage mechanism, the impact of ETF trading on international diversification and on price efficiency of distressed stocks.

In the first chapter, I show that, although low on average, ETF premiums/discounts can be as high as 16% when considering international country-level ETFs. I propose a risk-based limits to arbitrage explanation of such deviations. I show that while currency and equity illiquidity risks are important in explaining ETF premiums there is still a large portion of premium that remains unexplained. I argue that ETF premiums represent a reward arbitrageurs demand for being exposed to financial frictions risk and show that the absolute value of ETF deviations is a good proxy for multiple dimensions of financial frictions such as funding illiquidity, credit risk and information uncertainty. I show that it can be used as an aggregate financial friction proxy at the country-level and that it is priced in the cross-section of stock returns internationally.

In the second chapter, I show that investment decisions of ETF market participants when trading country ETFs are driven by shocks to U.S. fundamentals, rather than local risks. Investors react only to negative news about local economies. When U.S. economic uncertainty increases, investors switch to Cash ETFs. I demonstrate that ETF arbitrage mechanism is one of the key channels through which U.S. shocks propagate to local economies leading to increased return correlation with the U.S. market, limiting the benefits from international diversification. I find that countries with stronger ETF price discovery and lower limits to arbitrage have a higher comovement with the U.S. market.

In the third chapter, I examine the effect of exchange traded funds on the underlying stocks conditional on the credit quality of securities in the basket. I show that U.S. industry ETFs help to alleviate the short-selling constraint present for distressed securities at the individual stock level by providing the alternative trading route to gain the negative exposure via cheap short-selling of ETFs. As a result, ETF basket membership has a positive effect on distressed stocks price efficiency. In addition, I show that distressed stocks that are members of ETF basket do not show signs of distress anomaly unlike the non-member securities.

Introduction

The Exchange-Traded Fund (ETF) industry has experienced a significant growth in the recent years.[1] According to Investment Company Institute, as of July 2019, the total assets under management of global ETFs reached 3,990.9 billion dollars including 807.1 billion in international ETFs and 390 billion in the U.S. industry ETFs[2], which are the focus of this thesis. As a result, this type of investment product became systematically important and is now under a close scrutiny of major global regulators. Yet, the ETFs are still understudied in the academic research. The aim of this thesis is to provide a deeper understanding of the consequences of such double-digit growth on the equity markets. This is important as recent regulations such as MiFID II, introduction of "non-fully transparent ETFs" (i.e. funds that do not need to disclose holdings daily) and lowering of fees to zero by major ETF sponsors will lead to even large inflows of funds from retail and active investors in the future.

The previous academic research is mainly focused on the negative effects of ETFs. For example, recent studies show that ETFs adversely affect the information efficiency, liquidity and volatility of the underlying stocks (e.g. Ben-David, Franzoni, and Moussawi, 2018; Israeli, Lee, and Sridharan, 2017). In contrast, this thesis consists of three empirical papers where I examine multidimensional impact of ETFs on the financial markets. In the first paper, using country-level funds, I look at the ETF arbitrage mechanism, show that it is not necessarily risk-free (as often

[1] Blackrock estimates the growth of 19% per annum from 2009 through 2017 https://www.blackrock.com/americas-offshore/insights/etf-growth

[2] https://www.ici.org/research/stats/etf/etfs_07_19

assumed) and relate the discrepancies in the ETF pricing to aggregate financial frictions. In the second paper, I continue with country-level ETFs and show the negative consequences of ETF arbitrage on the correlation of developed and emerging markets with the U.S., that limit the potential for diversification when investing internationally. Finally, in the third paper, I look at the positive consequences of ETF trading and show that industry ETF short-selling improves the efficiency of the distressed stocks by providing an alternative route to gain a negative exposure when it is restricted directly.

In Chapter 1, I look at the cross-section of 22 country-level ETFs traded in the U.S. and empirically investigate the ETF premiums/discounts[3]. The ETF arbitrage mechanism is a unique property of this type of investment products that differentiates them from open-ended and closed-ended mutual funds. It allows ETF dealers to continuously create and redeem fund shares while maintaining a liquid secondary market. This mechanism is usually efficient in maintaining ETF prices close to the net asset values (NAVs) with deviations being low on average. However, I observe that for country-level ETFs this is not always the case. On some days, deviations can be as high as 16%. Such differences constitute a violation of law of one price since ETFs represent the direct claim on the assets in the basket and therefore, they should be valued identically. Previous research relates such deviations to ETFs tracking countries with non-synchronised trading periods (e.g. Levy and Lieberman, 2013). When the U.S. market is open and the underlying stock market is closed the arbitrageur is unable to close the deviations directly. Following Delcoure and Zhong (2007) and using two NAV adjustment models proposed by Goetzmann et al. (2001) and by Engle and Sarkar (2006) I correct the NAV for stale pricing, but still observe significant deviations even for the U.S. and Canada where time-difference is not a concern. In the paper I provide a risk-based limits to arbitrage explanation of premiums. I argue that ETF arbitrage is exposed to financial frictions (factors

[3]Through out this thesis, depending on the context, I often use the term “premium” and “discount” interchangeably to represent any deviation of price from the NAV

that are often not accounted in basic asset-pricing models, but that interfere with trades and affect risk exposure of stocks) and that any deviation of prices from NAVs represent a reward the arbitrageurs demand for being exposed to such friction risk. I start with currency and equity market illiquidity risks and show that they are related to the magnitude of the deviation. However, liquidity can only explain up to around 24% of the variations in premiums. I then show that the absolute value of ETF premium is related to multiple dimensions of the financial frictions including credit risk, funding illiquidity and information uncertainty. I argue that country-ETFs are better than many other products used in the literature (e.g. ADRs, bonds etc.) in capturing frictions risk at the aggregate country-level due to cross-sectional data availability. Finally, I show that my financial friction proxy is priced internationally, implying the investor hedging demand against risks discussed above.

After investigating the risks affecting the arbitrage mechanism of ETFs at the country-level I proceed to study the consequences of arbitrage in the country-ETFs on international diversification. To the best of my knowledge, chapter 2 is the first paper looking at the impact of ETFs on the cross-country correlations. Using a broader sample of 41 developed and emerging country-ETFs I show that since the majority of ETFs are traded in the U.S. its market conditions impact the decisions of ETF investors when trading international funds. In particular, in order to understand such impact I proxy for economic uncertainty using U.S. and local VIX variables and compute order imbalances for different versions of ETF investors for every country in our sample at a daily level. Our first contribution is in showing that investors mainly react to changes in the U.S. uncertainty rather than the local one. In fact, the evidence suggest that they only react to significant bad developments affecting the local economy. I show that when U.S. VIX increases investors leave the foreign market and move money to cash ETFs. Secondly, I develop a shock transmission mechanism that via ETF arbitrage incorporates U.S. uncertainty into the local prices of foreign stocks. Using staggered introduction of ETFs I demonstrate that in line with wake-up call hypothesis significant increases in the U.S. VIX force investors to reassess the fundamentals of the local economies and via ETF arbitrage increase the equity return correlation of foreign countries with the U.S. market. If the arbitrage is one of the channels that is responsible for correlation increase, then any impediments to such mechanism should prevent the shock transmission. In the cross-sectional sort I confirm that countries with higher limits to arbitrage (proxied by liquidity mismatch of Pan and Zeng (2019), as well as by Amihud (2002) illiquidity measure of the underlying stock market) in the ETF market have a lower return correlated with the U.S.. Finally, I show that our shock transmission mechanism relies upon investors treating any shock coming from the U.S. as fundamental. If they consider such shocks as noise (i.e. the ones arising due to liquidity reasons) the effect on prices should be reverted and no impact on correlation should exist. Following Broman (2016) I measure the degree of price discovery in the ETF (i.e if fundamental information is generated in the ETF market) and show that countries where such measure is high are more correlated with the U.S. market. Overall, in this paper I demonstrate that such increase in return correlation of countries with the U.S. makes it more difficult to achieve international diversification for U.S. investors who use country-ETFs for such purposes.

After showing a new channel via which ETFs negatively impact the financial markets, I investigate whether there are positive consequences of ETF growth. In chapter 3, I look at the industry ETFs and show the positive impact of such funds on the price efficiency of the distressed stocks. The novelty of this paper is that I highlight the increasing heterogeneity of type of stocks traded in the ETF basket. As such, I show that in contrast to previous ETF research that often does not differentiate between securities, the impact of ETF trading should be assessed conditionally on the individual stock characteristics. I demonstrate that credit quality is one of them. It is often hard to negatively bet against stocks in the financial distress due to costly short-selling. This paper demonstrates that industry ETFs (funds with one of the highest level of short-selling) provide an alternative route to sell distressed stocks indirectly. While the industry ETF short-selling has a limited impact on the non-distressed stocks, it is more likely to reduce the overpricing of distressed stocks by creating a downwards price pressure. I show that the impact of ETFs on such stocks is fundamental irrespective of the trading motives. I directly test the effect of indirect short-selling on the efficiency of stocks and find the positive relationship. I then turn to the distressed anomaly (reflects under-performance of riskier stocks with low credit rating) and show that stocks with low S&P rating that are members of industry ETFs are less likely to be under-priced in comparison to non-member securities. This effect is robust to alternative explanations of the anomaly and different measures of distress. Overall, this paper contributes to the ETF literature by showing the positive consequences of the development of such products.

Chapter 1

Financial Frictions Risk and ETF Premium

1.1 Introduction

Markovitz's portfolios theory and many asset pricing models make a crucial assumption of no frictions in the financial markets. From one side, this means that there is no constraint on how much leverage investor can take and subsequently no limits to investor's position on capital allocation line. However, studies such as Boguth and Simutin (2018) demonstrate that in reality the ability to borrow is limited and leverage constraint tightness is a priced risk factor. From another side, no frictions means that theoretical market portfolio is combined of all the world assets. In real life, the substantial transaction costs, barriers to international investments, inability to short-sell, search costs and many other factors limit the value of international diversification and leave investor portfolios concentrated in their home market (home bias). Many attempts were made to identify some of the friction components discussed above and to determine their relationship to asset pricing. Usually, to measure an individual component, a proxy variable is used. The issue is with those frictions that are not easily quantifiable, such as degree of asymmetric information, or with generating a measure for non-U.S. frictions. In this paper I provide a new aggregate country-specific measure of financial frictions in 22 markets based on the absolute value of international exchange-traded fund (ETF) premiums/discounts, show that it is able to capture multiple dimensions of frictions risk and that it is priced internationally.

Many ETF studies investigate fund premiums, but with a limited success. A large portion of ETF premium is still left unexplained (e.g. Delcoure and Zhong, 2007). The deviation of fund price from its net asset value (NAV) is a violation of the law of one price. It represents the arbitrage opportunity, which in theory should be corrected immediately. However, due to numerous limits to arbitrage related to financial frictions this is not always the case with significant deviations observed across many ETFs. I provide a risk-based explanation of premium and identify its main driving forces. Since any authorised participant (AP) has a limited risk-bearing capacity, they may not be willing to engage in an ETF arbitrage when it is too risky (relative to the possible reward). In addition, studies show that while domestic ETF premiums are quickly corrected, international ETF premiums may last for days (e.g. Engle and Sarkar, 2006). In this paper, I focus on country-level ETFs which are defined as those funds that are tracking a single country index (e.g. iShares MSCI Germany). I argue that premiums/discounts are not pure arbitrage opportunities and represent a reward for being exposed to numerous risk factors. When such reward is high enough, the deviation is corrected via unique ETF arbitrage mechanism. Often ETF deviations are related to liquidity. I proxy for market illiquidity with global equity illiquid-minus-liquid ($MILLIQ$) risk factor based on Amihud, Hameed, Kang, and Zhang (2015) and currency illiquidity using IML factor of Mancini, Ranaldo, and Wrampelmeyer (2013). I show that while market illiquidity and currency illiquidity risks are important in explaining ETF premiums they are only able to explain up to 20% of daily variation in ETF premiums suggesting that there are other dimensions of friction risk that drive the deviations.

DeGennaro and Robotti (2007) define market frictions as “anything that interferes with trade”. They classify frictions into several categories: transaction costs (including cost of trade and the opportunity cost of time), taxes and regulations, asset indivisibility, non-traded assets, agency and information problems. I argue that the arbitrageur in the ETF market faces numerous of those frictions and being risk averse demands the compensation for being exposed to them. For example, in order to redeem the ETF shares (to close the arbitrage opportunity in case when ETF is traded at discount) the authorised participant (AP) has to deliver to the ETF sponsor at least 1 unit of the fund (typically 100,000 shares). Therefore, such arbitrage may be too expensive/risky to execute, which is the example of asset indivisibility friction.[1]

I use the absolute value of ETF premium/discount as a measure of aggregate country-specific financial friction risk. In order to deal with a stale pricing problem found in the ETFs where ETF shares and the underlying basket of securities are not traded at the same time I adjust NAV using methods of Goetzmann et al. (2001) and Engle and Sarkar (2006). I investigate if other known proxies of different dimensions of frictions are related to absolute ETF deviation. I show that while ETF premiums are related to closed-end fund discount the correlation is not very high. In contrast to mutual funds that suffer from the inability to arbitrage the deviations due to closed-end fund structure the unique ETF arbitrage mechanism ensures that ETF premiums are small. As such, the use of ETFs allows for a more precise estimation of frictions than mutual funds. I find that apart from equity illiquidity, the studied proxy is related to funding conditions (e.g. Ted Spread), information uncertainty (e.g. Aggregate Disagreement) and credit risk (e.g. Moody's Spread). As such, I provide the evidence that the absolute ETF premium is related to multiple dimensions of friction risk.

After developing the financial friction proxy I compute the sensitivity of stock returns to financial frictions for every security in my sample and study what stock characteristics drive such sensitivity. I find that returns of stocks with high

[1] While some of the funds allow partial basket creation/redemption this is not a standard practice and requires a special agreement with the fund

market beta (cyclical stocks) and small market capitalisation co-move with frictions risk. In addition, stocks with high information uncertainty (low analyst coverage and high dispersion in forecasts) and securities with high financial leverage are also positively related to frictions. Such relationship between friction sensitivity and stock characteristics is consistent with different dimensions of frictions related to the ETF-based proxy. I then study the cross-sectional pricing ability of the financial friction risk. I find that stocks with high friction sensitivity tend to under-perform stocks with low sensitivity. The high-minus-low portfolio generates approximately -3.6% per annum on average. This suggests that investors are willing to hedge against friction risk. The result is robust to using value and equity-weighted portfolios. In addition, I investigate this risk premium internationally and find that 14 out 21 countries in my sample have negative and significant financial friction risk premiums.

Several previous studies attempt to measure different components of frictions. My paper is related to Hou and Moskowitz (2005). They proxy for the financial frictions using the stock delay measure. Such variable attempts to capture the difficulty in the information diffusion, as market news are reflected in stock returns with a delay. Alternatively, this proxy reflects the illiquidity of a stock. They find that the delay measure is able to predict returns cross-sectionally. However, the disadvantage of this approach to measure frictions is the need to pre-define the asset pricing model and the number of lags included in the delay measure. In contrast, the measure based on country ETFs does not require such assumptions. Malkhozov, Mueller, Vedolin, and Venter (2018) also develop a proxy for a financial friction that captures funding illiquidity. Similarly to this paper the proxy is at the country level and is based on the deviation of the government bond yields from the fitted yield curve. They show that this funding illiquidity is priced internationally and affects the slope and intercept of the international security market line. While such measure is relatively easy to identify the main drawback of this method is the limited data availability cross-sectionally and the existence of substantial credit risk premium embedded in the yield for some countries. As a results authors only cover

6 countries in their paper. My financial friction proxy is also related to funding liquidity, but, in contrast, ETF data allows to have a much broader sample with 22 countries covered in this study. My work is also related to Pasquariello (2014) who combines the deviations of law of one price in different markets into a single market dislocation index. The global measure includes arbitrage parity violations from the FX market (Covered Interest Rate Parity and Triangular Arbitrage), as well as from equity market in a form of ADR price discrepancies. Such measure is highly dependent on specific events and is decreasing over time. In contrast, I do not observe the reduction in friction premium over my sample period. In addition, ETFs allow to clearly estimate the country-specific frictions. While ADR arbitrage is fundamentally similar to ETFs there are significant differences. First of all, there is large variability in ADRs (e.g. over-the-counter Level 1 vs exchange-listed Level 2 and 3, different bundling ratios etc.). In contrast, country-level ETFs are relatively more standardised. Secondly, an ADR represents a claim on one specific foreign company. As such, any price deviation is likely to be influenced by idiosyncratic factors. Gagnon and Karolyi (2010) show that the daily premium can be as high as 127.4%. The ETF deviations are much smaller, as being an index product, they represent claims on the diversified basket of securities. I find the maximum deviation of approximately 16.5% in my sample. Finally, to compute the deviations for a specific country researchers are required to average such parity violations across ADRs from this country. This is a significant data problem as often there is not enough ADRs on stocks listed from a particular country. For example, Gagnon and Karolyi (2010) have only one stock for Austrian market. This is not a concern for ETFs for reasons mentioned above, which makes country-level ETFs a particularly appealing product in measuring country-level frictions. My work is also related to Bandi, Moise, and Russell (2006) who use intra-day prices of SPIDERS to measure volatility of frictions. Since my study is international, I use the lower frequency data (daily and monthly). In addition, I study the aggregate financial frictions using the first moment, while they focus on the second moment. My proxy is also

related to studies on stock market illiquidity. Acharya and Pedersen (2005) develop liquidity-adjusted CAPM and show that aggregate liquidity risk is priced in the cross-section of stock returns.

Most of the ETF research is focused on the U.S. index funds. Only few studies, such as Levy and Lieberman (2013), consider the international funds and non-synchronised trading periods of ETF shares and underlying baskets. They show that while ETF prices are driven by NAV returns, during the periods when underlying stock market is closed, ETF price overreacts to S&P 500 index return. This mispricing is possible due to failure of arbitrage mechanism in those periods of a day. In this paper, I show that even after controlling for non-synchronised periods the significant mispricing exists and is related to financial friction risks. Petajisto (2017) shows that the creation and redemption occurs on a much larger scale for U.S. ETFs, rather than international ETFs, as the underlying assets of those funds are more difficult to trade. In addition, he documents that while on average the premium is close to zero it exhibits a significant time-series variation. Some of the factors cited by Petajisto that can explain the variation include the movement of investors into and out of funds (creates a price pressure) and the availability of arbitrage capital. Petajisto also expresses concern that when using the in-kind creation/redemption arbitrageur bears the risk of non-simultaneous trading and unpredictable transaction costs. The noise in stock prices is also an important issue to arbitrageur, who is restricted to trade while one of the markets is closed. AP bears the risk that if the price of underlying stocks is noisy it can move to unpredictable direction when the second market gets opened (the so-called price slippage of Malamud, 2016). One of the contributions of my paper is that it quantifies the risk factors affecting the ETF arbitrage and show the direct impact on the ETF premiums/discounts. A large part of ETF literature demonstrates how ETF trading transmits noise to the underlying stock prices making this risk significant. Israeli et al. (2017) examine long-term implications of ETF ownership on information efficiency of underlying stocks. They show that as ETF acquires more shares of a company, those shares become less price informative. The effect comes from two sources. The first one is that when a large portion of shares gets locked-up in the ETF basket, they become unavailable for informed agents willing to transact based on firm-specific information. Secondly, presence of ETFs shifts uninformed investors from the direct trading of the underlying stocks to ETF trading, as these index instruments minimise their losses to informed traders. Subsequently, this leads to an increase in the transaction costs for the underlying securities and, since the information acquisition is costly, reduces the incentive of informed traders to acquire such information. In particular, authors show that an increase in ETF ownership leads to widening of bid-ask spread, increase in Amihuds illiquidity measure and increase in firms return synchronicity with overall market. Similarly to studies cited above, I argue that noise in underlying stock prices (a serious concern during the non-synchronised trading periods) and transaction costs are parts of friction risk.

There is a growing body of literature that attempts to determine the driving factors of the deviation between ETF prices and NAVs. Chacko, Das, and Fan (2016) use premiums of bond ETFs to develop a new fixed-income illiquidity measure that, as they suggest, can potentially be further extended to other asset classes. The main problem of typical proxies for illiquidity risk is that the standard approach of going long on assets with low liquidity characteristics (such as bid-ask spread) and short on assets with high liquidity characteristics may not offset other possible systematic risk factors. Despite the ability to overcome this problem and the computational simplicity of their approach the authors ignore many other possible factors that may explain the price premium. In particular, I demonstrate how other factors beyond liquidity, such as credit spread and funding constraints, are also important in determining the price difference between ETF and the underlying portfolio. Delcoure and Zhong (2007) show that variables such as the degree of institutional ownership, bid-ask spread, trading volume, exchange rate volatility, dummies for economical and political events, as well as correlation between local market and the U.S. have the ability to explain the variation in premiums of country ETFs. Nevertheless, with so many variables included, there is still a large portion of variation that is left unexplained with R^2 of their model not exceeding 14%. In contrast, I demonstrate how the clearly identifiable risk factor proxies are more effective in explaining the deviations. Bertone, Paeglis, and Ravi (2015) consider the ETF premium/discount in the context of law of one price deviation. They show that the tracking error between Dow Jones Industrial Average index and ETF is related to liquidity, volatility and transactions costs. Similarly, to this stream of literature I also use liquidity to explain ETF premium/discount. However, I do not use liquidity proxies in my regression, but rather consider liquidity risk in a form of long-short portfolios for both currencies and equity markets.

The paper proceeds as follows: section 1.2 describes the ETF arbitrage mechanism, main players in this market and provides a risk-based explanation of price deviations; section 1.3 provides the details on the ETF, equity sample and risk factor construction; section 1.4 shows the empirical results; section 1.5 concludes.

1.2 ETF mechanics

Exchange-traded fund is a modern hybrid of closed-end and open-end mutual funds. Similarly to closed-end funds its shares are traded on exchange and its price is subject to supply and demand of market participants. At the same time, it overcomes one of the main drawbacks of closed-end fund structure, difficulty in raising additional funds, by incorporating the special feature of open-end funds – the ability to issue new shares at their net asset value. ETF sponsor makes a special creation/redemption agreement with several broker dealers (usually large financial institutions), called authorised participants, to ensure that funds are traded at a price close to its net asset value. APs also often act as market makers in the ETF market (although this is not always the case). The authorised participant can (but not obliged to) accumulate the basket of underlying shares that ETF tracks in the right proportion[2]

[2] Typically the size, weights and constituents of the basket are disseminated daily in the fund's portfolio composition file

and pack it in the appropriate amount to reach a necessary size of a unit creation. He then has a right to exchange this basket for newly created ETF shares. Usually, this exchange happens in-kind (i.e. there is a physical delivery of the basket), however in-cash transaction may also be possible, but at additional fee.[3]

This unit creation/redemption process creates a unique arbitrage mechanism in the ETF market, which allows to keep the price of ETF close to the net asset value of the basket that the fund tracks. When the temporary ETF price P_t is above its NAV_t (ETF is traded at a premium) the authorised participant has the incentive to short-sell the ETF shares and simultaneously purchase the basket of underlying stocks. Then authorised participant can deliver this basket to the ETF sponsor, exchange it for new ETF shares and close the short position at a profit of $P_t - NAV_t$ (ignoring the effect of sale on price), arbitraging away the initial difference. Similarly, when the price of ETF is below its NAV (ETF is traded at discount) the authorised participant buys the ETF shares (enough to reach a unit creation size) redeems them for the underlying stocks and sells them in the stock market generating a profit of $NAV_t - P_t$. The initial purchase of ETF shares pushes the price upwards until it reaches the NAV. Therefore, theoretically, to preclude the existence of arbitrage, the ETF price must be equal to the NAV of the basket of shares that it tracks. Engle and Sarkar (2006) show that ETF prices and NAVs are co-integrated and any deviation is corrected within minutes for domestic funds. In contrast, international funds can be traded at premium or discount for several days. Similarly, Madhavan and Sobczyk (2016) show that the half-life of the deviation for domestic funds is 0.43 days, while it is 6.56 days for international funds. Arbitrage opportunities may also arise between futures and ETFs (Richie, Daigler, and Gleason, 2008) and when different funds tracking the same or similar basket (e.g SPDR Trust and iShares IVV) have different prices (Marshall, Nguyen, and Visaltanachoti, 2013). In the latter

[3]In-kind creation/redemption is often preferred by ETF sponsors. Firstly, it reduces the operational complexity of buying the underlying securities (e.g. round lots, trading costs, liquidity etc.). Secondly, for U.S. funds the in-kind creation allows the fund manager to pass shares with high accumulated capital gains to APs. Since no money is exchanged, such transfer minimises tax liability of ETF shareholders

case, authors argue that although the risk of correcting such mispricing exists it is minimised, as both funds are highly liquid and convergence risk is low. In contrast, I consider arbitrage opportunities between international funds and their underlying basket and show that both liquidity and convergence risks are important. While the ETF shares and the underlying basket represent claims on the same future cashflows, any deviations in international funds, which are the focus of this study, are harder and riskier to arbitrage due to non-synchronised trading periods and other market frictions.

It is possible for non-AP investors to get engaged in the ETF arbitrage activity. For example, they can sell ETF shares, buy underlying basket of securities (in case of ETF premium) and wait for price convergence. However, this is not a pure riskless arbitrage opportunity in a traditional sense (Ben-David et al., 2018) and APs remain key players in eliminating arbitrage opportunities. Often, previous research assumes that the authorised participant has an unlimited risk bearing capacity. This assumption leads to a doubtful conclusion that any deviation of ETF price from NAV will be corrected by an arbitrageur, as it represents a profit opportunity. However, in real life the arbitrage mechanism can be risky in particular, as argued before, for international funds. As APs have limited capital, may not be able to specialise in both ETF and stock markets at the same time (Bhattacharya and O'Hara, 2018) and require appropriate compensation for any kind of risk they undertake it may not be optimal for them to correct a small deviation. Figure 1.1 illustrates this idea by showing a hypothetical variation in prices of an ETF. I argue that there exist an upper and lower bound around the NAV. This bound is a minimum reward required by the AP to compensate for the risk of ETF "arbitrage". Starting with price P_1 that is above upper bound, the AP will close the deviation by pushing the price closer to the net asset value and earn $P_1 - P_2$. Assume that over time there is a liquidity shock in the secondary market pushing the price up to P_3. At this point it is not optimal for the AP to intervene as the potential profit $P_3 - NAV$ is less than $UpperBound - NAV$. The AP will intervene only when the premium/discount is sufficiently high (price reaches P_4) and close the gap between NAV and price.

[*insert figure 1.1 here*]

ETF market makers face numerous direct and indirect costs when conducting the arbitrage trades. While the underlying bid-ask spread, creation/redemption fees, trading fees and stamp taxes are important, the arbitrageur is also exposed to hedging costs that create a natural "fair value band" around the NAV (Vanguard, 2016; WisdomTree, 2019). Such costs vary depending on the fund (it is more expensive for emerging market funds due to lack of hedging instruments) or during times of uncertainty (which affects the difficulty in estimating the cost of a hedge). Petajisto (2017) shows that the volatility of premiums/discounts is economically significant and can be as high as 130 bps for some of the international funds. He also argues that in certain cases market makers may need days to accumulate the positions to conduct the arbitrage trade which exposes them to the timing risk. Consequently arbitrageurs may choose to wait until premium/discount widens enough to correct the deviations (SEC, 2019). Mackintosh (2014) argues that 90% of US equity arbitrage opportunities is unprofitable. The AP may even choose to stop accepting redemption orders as did Citigroup in 2013 when hitting the internal risk limits.[4] I argue that significant deviations between ETF price and NAV observed for the international country funds are driven by the existence of arbitrage-efficient bounds and represent a risk premium that arbitrageurs demand for being exposed to risk of correcting the mispricing.

1.3 Data and Methodology

In this section I describe the procedures used in this study for the construction of the ETF sample, as well as different methodologies for dealing with a stale pricing problem present in the international funds. I also provide the details of the

[4]https://www.ipe.com/reports/special-reports/etfs-guide/global-regulators-take-another-look-at-etfs/10021549.article

construction of order imbalance, equity and foreign exchange liquidity risk factors. Finally, I describe the construction of stock sensitivity to country-level frictions and of the global equity sample that is used for the asset pricing tests.

1.3.1 ETF Sample

In this study I cover all currently available developed market single country ETFs sponsored by iShares (Blackrock Inc.) that are not hedged (ticker starts with H) or target only small or large market capitalization firms. The sample consists of 22 ETFs traded on NYSE Arca, BATS exchange or NASDAQ: 21 developed market MSCI-based ETFs and IVV that tracks the S&P 500 index to represent the U.S. market. Table 1.1 provides the description of each fund. All of the funds in the sample use "physical replication" and attempt to minimize the fund deviations from their benchmarks (country-level MSCI indices). I collect daily closing prices for each of the fund from Thomson Reuters Datastream and Bloomberg. When the closing price is not available I replace it with the average of bid-ask spread. Net asset values are obtained from individual fund sections on the iShares website. As in Ackert and Tian (2008) I begin the sample on June 1, 2002, as this is the first date when free-float adjustment was incorporated into MSCI index. Before this date, MSCI methodology was based on the number of shares outstanding, which could prevent ETF sponsor from closely tracking the index due to non-availability of privately held shares for investment. The sample is at the daily frequency from June 2002-June 2018.

1.3.2 Stale pricing problem

The correct computation of ETF premiums is complicated by difficulty in estimating the true net asset value of the fund. Every 15 seconds market data vendor provides the indicative net asset value (INAV) of the fund. However, those indicative values may not represent the true intrinsic value. First of all, instead of the current mid-quote the last trading price is used in the NAV computation. Therefore, the direction of the last trade determines whether the bid or ask price is used in valuation. In addition, according to Madhavan and Sobczyk (2016), when market is very illiquid the last trading price may represent the significantly delayed valuation of securities, as the trade could happen minutes ago. Ben-David et al. (2018) describe the mechanism of propagation of fundamental shock when price discovery occurs in the ETF market. In their example, due to ETF being more liquid than the underlying basket its price immediately reflects new fundamental information, but the prices of securities are temporally stale and only reach new equilibrium with a delay.

The stale pricing problem is especially relevant for country ETFs due to non-synchronized periods of trading. The challenge in accurately computing NAV arises as ETF shares are traded in the U.S. market, while the price of underlying basket of securities is determined in the foreign market. Asian markets do not have the overlapping trading hours with the U.S. market. NAV provided by the vendor when Asian market is closed reflects the closing price of the underlying securities adjusted for foreign exchange return. This elevates the stale pricing problem and, according to Levy and Lieberman (2013), also limits the ETF arbitrage mechanism.

Several solutions are proposed to correct the reported NAV. Valuation based on stale prices does not reflect all available information immediately. Following Goetzmann et al. (2001) I adjust today's NAV with the predictable component of tomorrow's NAV. This incorporates available value-relevant information into the valuation of the underlying basket and makes its return unpredictable. First, I regress NAV return on the instrument Z that helps to forecast NAV at time $t+1$:

$$R_{i,t+1}^{NAV} = \alpha_i + \beta_i Z_t + \varepsilon_t \tag{1.1}$$

I use S&P 500 index return as an instrument. Then I update NAV at time t with its predictable return by assuming that α is equal to zero. $\overline{NAV}$ shows the true (adjusted) net asset value.

$$\overline{NAV_{i,t}} = NAV_{i,t}(1 + \beta_i Z_t) \quad (1.2)$$

Engle and Sarkar (2006) propose the alternative adjustment to NAV in order to solve the stale pricing problem. They suggest that even in the presence of measurement errors in the long-run the ETF prices and NAV must be the same due to built-in self correcting arbitrage mechanism. Based on this cointegrating property they derive the following relationship:

$$p_{i,t} - NAV_{i,t} = \alpha_i \Delta NAV_{i,t} + \beta_i x_{i,t} + u_t \quad (1.3)$$

where $x_{i,t}$ is a set of variables that explain the difference between measured and true NAV and u_t is the true premium. Similarly to Delcoure and Zhong (2007) I use S&P 500 return and spot exchange rate returns as proxies for x_t.

In contrast to previous methodologies, Petajisto (2017) develops another method for NAV correction that does not require any assumptions about the price processes. He sorts funds that track the same or highly correlated basket of securities into groups and then computes the true NAV as the average of group prices. Similar funds must move together and any deviation from the group captures the idiosyncratic mispricing. While this method is computationally easier to use in comparison to the ones described above, it suffers from several problems and therefore is hard to apply in this study. First of all, it assumes that there is no systematic mispricing among all the ETFs in the group. In contrast, Broman (2016) demonstrates the significant systematic mispricing in international iShares ETFs. Secondly, MSCI index provider compiles the list of all ETFs tracking each of its country indices. According to this list, many countries have very few tracking ETFs. In addition, to have a sizeable group for each country, ETFs traded outside of the U.S. must be included. This would lead to the same time zone problem as with NAV computation. On top of that, different currency exposures must be taken into account. Considering above, I deem this method less appropriate for international country ETFs and, as such, I do not use it for this study.

1.3.3 Currency and Equity Market Illiquidity Risks

Liquidity risk is an important factor when considering the deviation of ETF prices from NAVs. As mentioned earlier, it is often considered a part of the financial friction risk. Pan and Zeng (2019) and other market reports argue that liquidity risk is significant for ETF APs, especially when the liquidity mismatch between ETF and underlying market is high (e.g. bonds, international funds etc.).

Although ETF and its basket of underlying securities represent a claim on the same assets there are two main differences between them. First, while ETF shares are traded in U.S. dollars (within the sample considered), the underlying basket is priced in the local currency. These currencies are more illiquid than U.S. dollars and, therefore, holding underlying basket exposes international investor to additional risk, which must be priced and reflected in the NAV. I obtain closing bid and ask exchange rates for each underlying currency in the ETF sample from Thomson Reuters Datastream. In order to construct currency illiquidity risk factor I follow Mancini et al. (2013) in their procedure of computing IML (illiquid-minus-liquid) factor. Each day I rank currencies based on their liquidity (using relative quoted bid-ask spread). I then construct a portfolio that is long in the tercile of the most illiquid currencies and short the tercile of the most liquid ones. The portfolio is rebalanced daily. The constructed IML factor represents the risk premium that investors demand for being exposed to currency illiquidity risk. The only fund that is not affected by currency risk in my sample is IVV that tracks the U.S. index.

Similarly, I construct equity market illiquidity risk factor. In the international sample, underlying basket of securities may be illiquid. Since the arbitrageur has to go long or short in the underlying market (depending if the fund is traded at a premium or discount) such risk is important to consider. I follow Amihud et al. (2015) in constructing the global illiquidity risk premium. For every stock j at time

t I compute the Amihud (2002) illiquidity measure as follows:

$$Illiq_{j,t} = \log\left(1 + \frac{|R_{j,t}|}{DVOL_{j,t}}\right)$$

where $R_{j,t}$ is the daily stock return, $DVOL_{j,t}$ is the daily dollar traded volume in local currency obtained by multiplying the closing stock price by the number of shares traded. Similarly, to many empirical papers (e.g Malkhozov et al., 2018) I reduce the impact of outliers by adding a constant and taking a logarithm. Daily equity market illiquidity risk premium ($MILLIQ_t^G$) at time t is calculated as a return on portfolio that goes long in the tercile of the most illiquid stocks and short the tercile of the most liquid ones.

1.3.4 Financial Frictions Risk Exposure

Using the risk-based limits to arbitrage explanation I consider the absolute deviation between ETF price and its NAV as a risk premium that arbitrageurs demand to get engaged in the arbitrage activities. Therefore, I use the absolute value of the ETF premium as a financial friction proxy (FFP, FFP^G, FFP^E for simple, Goetzmann et al. (2001) adjusted and Engle and Sarkar (2006) adjusted premiums). Since I use country-level ETFs, this proxy is country specific. As discussed before, I hypothesise that such financial friction measure for fund i has the ability to explain variations in excess returns for stocks of country i. Assuming that the average coefficient of risk aversion of authorised participants stays constant, a higher FFP represents a larger financial friction risk. I first find the sensitivity of each stock in country i to this proxy. Following Pasquariello (2014) every month t, for every available stock j in country i I estimate beta from the following regression:

$$R_{j,T} = \alpha + \beta_1 FFP_T^i + Factors_T^i + \varepsilon_{j,t} \qquad T \in \{t-59, t\} \tag{1.4}$$

where $R_{j,T}$ is a monthly excess return over the previous 60-month for stock j traded in country i. For each country, I first standardise FFP^i by its rolling historic mean and standard deviation (as in Pasquariello, 2014) to avoid forward looking bias. I then average standardised variables within each month to generate monthly series of friction proxies used in regression. $Factors^i$ are monthly country-specific MRP, HML, SMB and UMD factors from Carhart (1997) model obtained from the AQR website. I then sort stocks based on estimated betas into quintiles and compute the returns of portfolios over the next month. I construct a long-short portfolio by buying the most sensitive group (high β_1) and selling the least sensitive one (low β_1).

1.3.5 International Stock Data

I follow Asness, Frazzini, and Pedersen (2019) in the dataset construction. The universe of stocks that I consider in this study covers all constituents of Compustat Global, Compustat North America and CRSP databases for the period of June 2002-June 2018, whose primary listings are located in one of the 22 countries from the ETF sample. I only consider common stocks by filtering databases by the issue code (TPCI=0) or share code (SHRCD=10 or 11) and I exclude all preferred stocks, depository receipts, REITs, warrants etc. Each stock is allocated to a particular country based on the location of its primary exchange, rather than based on the country of domicile. Only primary listings are considered (identified using primary issue tag). As common in the literature, I include all dead stocks to limit the potential survivorship bias effect. I account for delisting returns of U.S. stocks by following Shumway and Warther (1999). After careful consideration of data structure and following cleaning procedures introduced in asset pricing literature I apply the following filters to the dataset:

1. Volume filter: I exclude all observations with non-positive trading volume
2. Liquidity filter: I exclude all penny stocks defined as stocks with mean price lower than 1 unit of exchange-local currency within the sample period
3. Trading day filter (e.g. Malkhozov et al., 2018): I remove those days on which more than 90% of return observations are zero

4. Staleness filter (e.g. Malkhozov et al., 2018): I remove a month of observations if at least 80% of returns within this month are zeros. Similarly, I remove a month of observations if 50% of observations within this month are not available

5. Survival bias (e.g Fama and French, 1993): I require stocks to have at least 2 years of return observations within a sample period

I then compute individual dollar returns in USD for each stock in excess of U.S. T-bill rate.

1.4 Empirical Results

In this section I provide the empirical evidence of the relationship of absolute value of ETF price deviations from NAV to aggregate financial frictions. I start by investigating the spikes in the time-series of ETF deviations and comparing the ETF premiums to closed-end fund discounts highlighting that they are different. I proceed by analysing the importance of currency and equity market illiquidity risks as driving factors of ETF mispricing. I then compare my proxy to other known dimensions of frictions risk, investigate what stock specific characteristics drive the return sensitivity to the aggregate level of frictions and study the pricing ability of frictions in the cross-section of stock returns.

1.4.1 Descriptive Statistics

Figure 1.2 shows the cross-sectional average of the absolute premiums for 22 ETFs in the sample. While the average daily deviation is below 1% there are numerous time periods when it is high. For example, at its peak during the financial crisis such deviation was more than 6%. The financial crisis is associated with a time of large uncertainty with many frictions such as the degree of asymmetric information spiking over this period. Petajisto (2017) documents a rise of the cross-sectional dispersion in ETF premiums at that time. In addition, noticeable jumps in the deviation occurred during historically important volatile events (i.e. "market dislocations" as they are called in Pasquariello, 2014). Deviations were high during the Flash Crash event on May 6th, 2010. According to BlackRock (2011) report on that day within minutes 25% of Russell 3000 stocks dropped by more than 10%. Due to the extreme fall in the values of U.S. shares it was difficult for ETF market makers to value funds and they had to discount their bid quotes. In addition, arbitrage became extremely risky as to hedge the long exposure in ETFs APs had to short falling stocks. The possibility of short-selling cancellation by exchange was high making the arbitrage mechanism non-functional. In addition, the situation was worsen by the inability of traders to route orders to NYSE Arca where many ETFs are traded. Overall, this situation resulted in many APs having to step away. This example highlights numerous technological, regulatory and other hard to quantify frictions that are present in the financial markets. Another example of such event occurred on February 5th, 2018 that resulted in the significant deviations of ETF prices from NAVs. On this day a jump of 115% in VIX resulted in market turbulence and extreme trading volumes in the ETF market (BlackRock, 2018). High volatility environment makes it hard to price the underlying assets and, on average, deviations from NAV spiked. Interestingly, the events described above affected most of the country-level ETFs in my sample, with the deviation being significant even after removing U.S. ETF from the sample. This suggests that there might be some friction factors at the global level that drive the common component of deviations. I test the impact of global illiquidity later in this paper. Different methods of premium adjustment produce a similar pattern of absolute deviations, with the main difference being the range of observations.

[*insert figure 1.2 here*]

Table 1.2 shows the summary statistics for premiums and the absolute premiums of ETFs in my sample. As can be seen from the table most of the international country-level ETFs trade at a small premium. This is in contrast to closed-end funds that are traded at discount (e.g. Lee and Ready, 1991). On average, daily mean deviation is 0.42% globally and it is less than 1% for every developed country in my sample. The mean absolute deviations exhibit a clear geographical pattern with ETFs tracking countries that have the highest time difference with the U.S. having the highest deviations (Australia, Japan, Hong Kong). Unsurprisingly, the smallest average deviations are for the U.S. and Canada, where the time difference is zero. On the regional level, North American funds have the lowest absolute deviations, followed by Europe and Asia Pacific. Such geographical pattern emphasizes the stale pricing problem discussed previously and the importance in adjusting the premiums. Tables 1.3 and 1.4 show the summary statistics for the adjusted premiums. The geographical pattern is much weaker with Ireland having the largest mean deviation of 0.78% and Australia showing a much lower mean deviation ranking despite having the highest time difference (average deviation is 0.48%). Asia Pacific has still the highest deviation on average, but the difference with Europe is much lower after the adjustment. While mean deviation remains below 1% the premiums vary significantly with minimum for Israel being -16.50% and the maximum for Canada (where one would not expect high deviations due to proximity to the U.S.) of 10.48%. Table A.1 in the appendix shows the correlation of absolute premiums across different methods of NAV adjustment. Tables A.2–A.4 show the correlation for different versions of absolute premiums across countries. Different methods of NAV adjustment produce positively correlated results. The adjustments tend to affect countries with the highest level of non-synchronicity. For example, the correlation between non-adjusted absolute premiums and the ones adjusted based on Engle and Sarkar (2006) is the lowest for Australia, Hong Kong and Japan. In contrast, it is 0.92 for Canada. Across countries all pairwise correlations are positive no matter what adjustment is used. This suggests that premiums and discounts tend to co-move across different funds. Interestingly, the magnitude of correlations is being diminished by adjustments with "no adjustment" being the highest on average and Engle and Sarkar (2006) adjustment being the lowest.

[*insert tables 1.2, 1.3 and 1.4 here*]

1.4.2 ETF Premium and Illiquidity Risks

Numerous market reports cite illiquidity as the main driver of ETF premiums. I test this by performing a panel regression with time and country fixed effects of the absolute value of ETF premiums on my proxies for currency and global market illiquidity risks:

$$|p_{i,t} - NAV_{i,t}| = \alpha + \beta_1 IML_t + \beta_2 MILLIQ_t^G + FE_i + FE_t + \varepsilon_{i,t} \tag{1.5}$$

In contrast to Delcoure and Zhong (2007) who uses 8 different variables to explain the deviations I attempt to explain these arbitrage opportunities from a risky arbitrage perspective using only tradable risk factors constructed as long-short portfolios. I run this regression conditionally on whether the fund is traded at a premium or discount. The reason for this separation is that the risk exposure is different depending on the direction of arbitrageur's trades. When the fund is traded at a premium the AP has to buy the underlying stocks and sell ETF. As a result he is long the foreign equities expressed in a foreign currency that are more likely to be illiquid (in comparison to the U.S. equity market and USD). In contrast, when the fund is traded at discount the opposite is true. Therefore, since the risk factors are constructed as illiquid-minus-liquid portfolios I expect different signs of the beta coefficients of the risk exposures for premiums and discounts.

Table 1.5 reports the results at a daily frequency. Results are provided for two methods of net asset value adjustment discussed previously. I find that both currency and equity market illiquidity risk increase the absolute premiums, which is in line with a risk-based limits to arbitrage explanation. As expected, coefficients are negative but still significant for the discount version of the regression.[5] The R^2 ranges from 17.8% to 23.6% which is larger than the ones reported in Delcoure and

[5]Table A.5 in the appendix presents the alternative specification without absolute values and conditional sorting. The conclusion remains the same

Zhong (2007) (13%-14%) despite only 2 regressors being used. Such R^2 suggests that while liquidity risk is an important driver of the deviations there is still a substantial portion of variation in absolute premiums that remains unexplained. I argue that other non-liquidity dimensions of frictions are also important determinants of premiums.

[*insert table 1.5 here*]

1.4.3 FFP and Individual Components of Frictions

I proceed by determining whether the obtained FFP indeed captures the financial frictions and if it is related to different types of frictions other than stock and currency illiquidity. I obtain numerous proxies for different dimensions of frictions and compute the correlation of those measures with FFP. Due to data availability the analysis is done at a monthly frequency for the U.S. version of frictions. I obtain U.S. average closed-end fund discount (CEFD) and investor sentiments proxy[6] from Baker and Wurgler (2006). Since FFP is constructed as the absolute value of the deviation I use the absolute value of CEFD. The results are almost numerically identical when a general version of CEFD is used. I use TED Spread (obtained from the Federal Reserve Bank of St. Louis database), measured as the difference between 3-month Libor rate and U.S. Treasury bill rate, as a proxy for funding liquidity. Similarly, I use Moody's Spread, measured as a difference between BAA and AAA corporate bond rates, to proxy for the counter-party credit risk. I obtain VIX from CBOE and Consumer Confidence Index (CCI) from OECD to proxy for the general economic uncertainty. I use Leverage Constraint Tightness (LCT) of Boguth and Simutin (2018) to proxy for the difficulty in accessing leverage[7]. I test the importance of liquidity in determining ETF premium by using Pástor and Stambaugh (2003) aggregate liquidity measure. Finally, I follow Hong and Sraer (2016) to construct the

[6] Although Sentiments variable is constructed using closed-end fund discount, the variable is based on the first principle component of numerous other proxies for sentiments

[7] Following Boguth and Simutin (2018) I measure correlation between changes in LCT and AR(1) residuals of other variables

Aggregate Disagreement measure, as a beta weighted average of dispersion in analyst forecasts of the EPS long-term growth rate (obtained from IBES database) for every stock in the country. This measure proxies the degree of information uncertainty in the economy.

Table 1.6 shows the pairwise correlation of individual proxies for financial frictions with FFP. Although the correlation with CEFD is positive it is one of the lowest among other variables (0.18). This suggests that these measures are different. The correlation with Investor Sentiments is negative but higher (in absolute values) than with CEFD. The negative coefficient is unsurprising since negative CEFD is one of the components of this measure. Higher value of correlation implies that FFP is more correlated with other components of Sentiments. Most importantly, FFP is highly correlated with TED and Moody's spreads (coefficients are 0.66 and 0.81), which means that funding illiquidity and credit risk are very important determinants of frictions. The arbitrageur needs funds to buy stocks/ETFs, as well as for posting margin as the collateral and for rebate fees for the short side of the trade. Interestingly, the TED spread is uncorrelated with CEFD. VIX and CCI are also strongly related to FFP. As mentioned previously, days with VIX jumps correspond to turbulent events and non-functional arbitrage mechanism. FFP is also correlated with Aggregate Disagreements measure since informational uncertainty affects stock prices. Finally, there is a strong negative correlation with Pástor and Stambaugh (2003) aggregate liquidity, as illiquidity complicates the arbitrage mechanism. Overall, my proxy for the financial frictions risk performs well at capturing different dimensions of frictions such as funding constraints, illiquidity and information uncertainty.

[*insert table 1.6 here*]

I also investigate the relative importance of each factor by regressing the FFP^{US} on individual proxies unconditionally (at a monthly level) and conditionally on whether the fund is traded at a premium or discount (at a daily level[8]). Similarly

[8] Not all proxies are available at a daily frequency. When daily data is not available I assume that the variable is constant throughout a month

to correlation results, I find that funding illiquidity, credit risk and information uncertainty are the main drivers of premiums/discounts. When considered jointly with other proxies, I do not find the significance for aggregate disagreement and liquidity. At a daily level I find that TED spread, Moody's spread and VIX are "symmetric" and similar in magnitude in their effect on premiums and discounts. Interestingly, I find that high investor sentiments negatively affect the premium and do not affect the discount.

[*insert table 1.7 here*]

1.4.4 Individual Stock Determinants of Exposure to Frictions

After showing that my proxy captures the financial friction risk at the aggregate level I proceed by measuring the exposure of individual stock returns to FFP. I analyse the stock-specific determinants of such exposure to see if they are consistent with different dimensions of frictions discussed above. As described in section 1.3.4 I compute the monthly exposure of stocks using rolling 60-month regression of stock returns on FFP and other factors. The obtained β_{FFP} captures the sensitivity of excess returns to frictions controlling for other risk factors. In order to understand what drives the risk exposure to frictions I then regress the time-series of obtained beta exposures for each stock on the individual stock characteristics. Due to data availability I focus on the FFP^{US}. Table 1.8 shows the results of such regressions. The panel regression includes time and industry fixed effects (based on 49 Fama-French industry classification). I follow Green, Hand, and Zhang (2017) to construct the explanatory variables. The data is obtained from CRSP, Compustat and I/B/E/S. I first test the relationship between friction sensitivity and risk sensitivity measures from Fama-French 3 factor model: market beta, log of market value of equity and book to market ratio. I then include variables that capture the information uncertainty in stock prices: accruals, number of analysts following a stock, dispersion in analyst forecasts and idiosyncratic volatility. Finally, I add financial leverage and Amihud (2002) illiquidity measure to proxy for stock liquidity.

[*insert table 1.8 here*]

I find that cyclical stocks are more sensitive to frictions as there is a positive coefficient for market beta. Unsurprisingly, smaller stocks tend to have a higher β_{FFP}, as such firms are more likely to be financially constraint and it is relatively more costly for them to obtain new financing. I do not find significant relationship between friction exposure and growth/value stocks in the full regression specification, which is in contrast to Boguth and Simutin (2018) who relate their leverage constraint tightness proxy to leverage embedded in growth options in some stocks (in the short version of the regression the negative coefficient for book to market ratio is consistent with their findings). I find that stocks with higher price uncertainty (low number of analysts and high dispersion of forecasts) tend to have a higher exposure to frictions. Stocks with high leverage are more sensitive to frictions since, as shown before, FFP is related to cost of borrowing. Surprisingly, I do not find the relationship with Amihud (2002) illiquidity which could be due to noise in the estimation of this variable at the stock level and at the monthly frequency. Overall, the results are consistent with previous evidence that FFP is related to different dimensions of frictions.

1.4.5 Cross-Sectional Stock Returns and Financial Frictions Risk

I proceed by investigating if the financial frictions risk is priced in the cross-section of U.S. and international stocks. Every month I sort stocks into quintiles based on their ex-ante sensitivity to the financial frictions proxy and then compute a return in excess of the risk-free rate over the next month of the value-weighted portfolios in each quintile. Table 1.9 shows the results of such sorting for the U.S. stocks. I report excess returns and intercepts based on CAPM, Fama-French 3 factor and Carhart (1997) 4-factor models. The estimated excess returns decline almost monotonically from Low to High. The High minus Low strategy generates the significant excess return of -0.299%. Similar pattern is observed for alphas. The negative premium is consistent with findings of Boguth and Simutin (2018), Malkhozov et al. (2018) and

Pasquariello (2014). The most sensitive to frictions group has a higher return when the level of frictions risk is high (i.e in the bad state of the world: when funding conditions, market illiquidity and other dimension of frictions are tight). In contrast, Low group returns negatively covary with frictions (i.e. it pays low in the bad state of the world). As a result investors demand the compensation to hold such riskier securities. In other words, the negative premium represents a demand by investors to hedge against friction risk. I find that the risk premium estimated based on FFP is relatively smaller than the one reported in Pasquariello (2014).[9] Panel B of table 1.9 shows that the results are similar for equally-weighted portfolios.

[*insert table 1.9 here*]

I also perform the same sorting exercise for every country in my sample. Table 1.10 reports the results based on FFP constructed using Goetzmann et al. (2001) adjustment. The results remain quantitatively similar for Engle and Sarkar (2006) adjustment due to high correlation between two types of premiums. The table reports the excess returns of portfolio 5 over portfolio 1, as well as 4-factor alpha based on Carhart (1997) model. More than half of the countries have the significant and negative average excess returns. Only 3 countries have insignificant positive returns. The highest premium that investors require for being exposed to frictions is in Denmark (12.32% per annum). The UK has one of the lowest estimates of 1.22% per annum. I do not find the significant results for some of the countries with low number of stocks in the cross-section (e.g. Finland, Ireland, Israel). Surprisingly, Japan, despite containing a relatively large number of stocks, is not significant over the sample considered in this study. This is consistent with Malkhozov et al. (2018) who also find that funding illiquidity is not priced in Japan. As shown previously, funding illiquidity is one of the main components of financial frictions that explains the ETF premium and therefore, it could be the reason for the absence of significant friction premium. When considering intercepts from the 4-factor model I find that

[9] It is 3.6% per annum in my sample vs 8.76% in Pasquariello (2014) for 1994-2009 sample. Of course, some dimensions of frictions, such as trading costs, are expected to diminish over time

14 out of 21 countries have the significantly negative coefficients. Overall, such result confirms that the findings based on the U.S. sort holds internationally and investors demand a risk premium for holding stocks sensitive to financial frictions.

[*insert table 1.10 here*]

1.5 Conclusion

I study the deviations of ETF prices from NAVs for 22 international country-level ETFs and provide a risk based explanation for premiums/discounts. I find that on average country ETFs are traded at a premium of less than 1%, but significant deviations in the range of -16% to +10% are also possible. Such high deviations represent a violation of law of one price. I provide a risk-based limits to arbitrage explanation of why they can persist in ETF market. While ETF arbitrage mechanism is efficient in quickly eliminating such profit opportunities, I argue that APs would only engage in creation/redemption when the reward (i.e. deviation of prices from NAV) is high enough to compensate for numerous risks associate with such trades. This creates upper and lower trading bounds around the NAV that vary with the magnitude of risk.

I argue that the absolute deviation of ETF prices from NAVs is a good proxy for country-specific aggregate financial friction risk. I demonstrate that since ETF shares and underlying basket of securities are traded in different countries and denominated in different currencies, market and currency illiquidity are important factors in explaining international ETF premiums. However, they are only able to explain up to 20% of premium variation at a daily level. Consequently, I relate my proxy to numerous other known dimensions of frictions. I find that apart from liquidity FFP is related to funding conditions, information uncertainty and credit risk.

I investigate the stock sensitivity to aggregate frictions and find that securities with high market beta and small capitalisation tend to co-move with frictions.

Similarly, stocks with high price uncertainty proxied by low analyst coverage and high dispersion of EPS forecasts are also more sensitive to friction risk. Finally, stocks with high leverage are more dependent on aggregate frictions.

Finally, I study the cross-sectional ability of financial friction risk to explain stock returns. I find that investors are willing to hedge against such risk. Stocks with high sensitivity to frictions are under-performing securities with low sensitivity generating a negative risk premium. Overall, this paper offers a risk based explanation of observed ETF premiums and provides a novel method for computing aggregate financial friction proxy at a country level using the absolute ETF premiums/discounts that can be useful in the future international studies.

Figures and Tables

Figure 1.1: ETF Trading Bounds

This diagram shows a hypothetical path of ETF price (from P1 to P5) relative to NAV in the presence of upper and lower trading bounds.

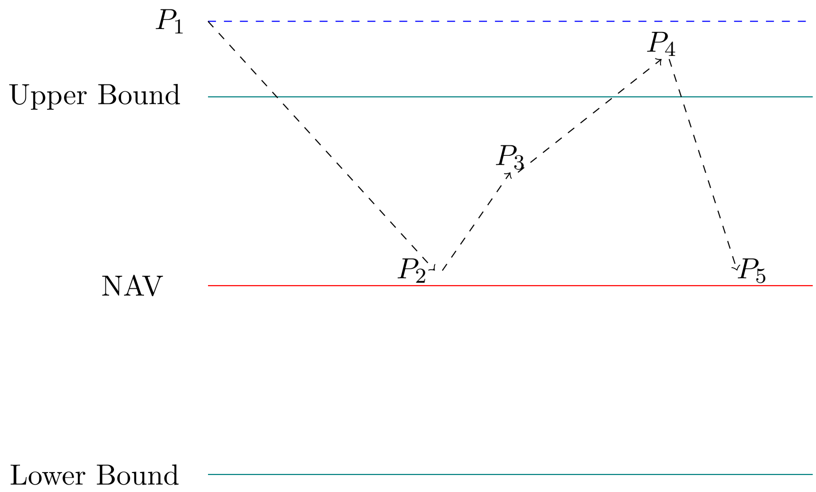


Figure 1.2: Average Absolute Premium

This figure shows the cross-sectional daily average of absolute premiums of 22 ETFs in the sample from June 2002- June 2018. The absolute premium is measured as the absolute value of the difference between log price (p) and log of net asset value (NAV). Top figure shows the simple version, the middle figure shows the version based on Goetzmann et al. (2001) adjustment, and the bottom figure shows the version based on Engle and Sarkar (2006) adjustment.

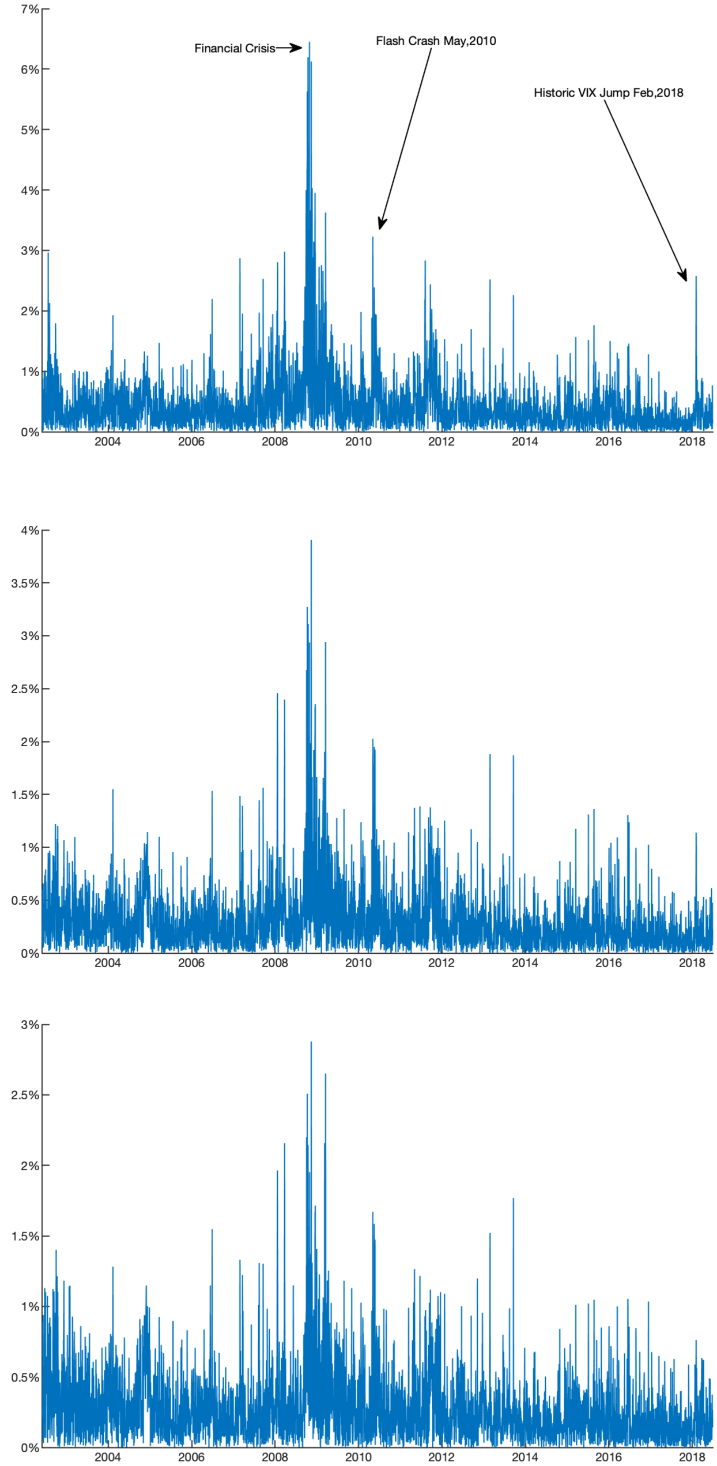


Table 1.1: Sample of iShares ETFs considered in the study

This table shows the list of international country-level iShares ETFs used in this study.

Country	ETF Name: iShares	Ticker	Basket Currency	Tracking Index
Australia	MSCI Australia ETF	EWA	AUD	MSCI Australia Index
Austria	MSCI Austria Capped ETF	EWO	EUR	MSCI Austria IMI 25/50
Belgium	MSCI Belgium Capped ETF	EWK	EUR	MSCI Belgium IMI 25/50
Canada	MSCI Canada ETF	EWC	CAD	MSCI Canada Index
Denmark	MSCI Denmark Capped ETF	EDEN	DKK	MSCI Denmark IMI 25/50 Index
Finland	MSCI Finland Capped ETF	EFNL	EUR	MSCI Finland IMI 25/50 Index
France	MSCI France ETF	EWQ	EUR	MSCI France Index
Germany	MSCI Germany ETF	EWG	EUR	MSCI Germany Index
Hong Kong	MSCI Hong Kong ETF	EWH	HKD	MSCI Hong Kong Index
Ireland	MSCI Ireland Capped ETF	EIRL	EUR	MSCI All Ireland Capped Index
Israel	MSCI Israel Capped ETF	EIS	ILS	MSCI Israel Capped IMI
Italy	MSCI Italy Capped ETF	EWI	EUR	MSCI Italy 25/50
Japan	MSCI Japan ETF	EWJ	JPY	MSCI Japan Index
Netherlands	MSCI Netherlands ETF	EWN	EUR	MSCI Netherlands IMI
New Zealand	MSCI New Zealand Capped	ENZL	NZD	MSCI New Zealand IMI 25/50
Norway	MSCI Norway Capped ETF	ENOR	NOK	MSCI Norway IMI 25-50 Index
Singapore	MSCI Singapore ETF	EWS	SGD	MSCI Singapore Index
Spain	MSCI Spain Capped ETF	EWP	EUR	MSCI Spain 25/50
Sweden	MSCI Sweden ETF	EWD	SEK	MSCI Sweden Index
Switzerland	MSCI Switzerland Capped ETF	EWL	CHF	MSCI Switzerland 25/50
United Kingdom	MSCI United Kingdom ETF	EWU	GBP	MSCI United Kingdom Index
USA	Core S&P 500 ETF	IVV	USD	S&P 500

Table 1.2: Summary Statistics: ETF Premium- no adjustment

This table shows the summary statistics for the daily ETF premium and for absolute value of the premium (Mean Deviation). The absolute deviation is measured as $\sum_{t=1}^{T} |p_{i,t} - NAV_{i,t}|$, where $p_{i,t}$ and $NAV_{i,t}$ are the log of price and log of net asset value of fund i at time t. The standard deviation, minimum, maximum as well as skewness and kurtosis are shown for premium. The bottom of the table shows the cross-sectional average of the summary statistics aggregated by the geographical areas: Global, North America, Europe, Asia Pacific. The data period is from June 2002 until June 2018.

Countries	Mean Premium (%)	Mean Deviation (%)	Standard Deviation (%)	Min (%)	Max (%)	Skewness	Kurtosis
Australia	0.114	0.745	1.102	-8.902	10.920	-0.355	14.208
Austria	-0.002	0.571	0.862	-8.889	8.419	-0.790	17.277
Belgium	0.057	0.517	0.785	-7.601	7.668	-0.215	17.100
Canada	0.063	0.317	0.519	-5.348	12.453	3.943	91.089
Denmark	0.119	0.368	0.474	-2.270	2.892	-0.011	5.350
Finland	0.128	0.393	0.557	-7.734	4.030	-1.682	32.975
France	0.056	0.480	0.726	-5.028	8.836	0.627	19.633
Germany	0.042	0.470	0.702	-5.827	7.354	0.247	14.478
Hong Kong	0.025	0.734	1.104	-9.347	7.917	-0.538	11.493
Ireland	0.539	0.787	0.863	-3.230	5.201	0.182	3.532
Israel	-0.048	0.567	0.899	-16.645	4.575	-3.362	55.170
Italy	0.049	0.517	0.760	-6.413	7.690	-0.057	14.221
Japan	0.071	0.780	1.127	-12.666	12.155	-0.383	14.489
Netherlands	0.045	0.486	0.729	-6.055	6.833	-0.147	14.961
New Zealand	0.055	0.464	0.636	-3.390	4.387	-0.102	6.405
Norway	0.079	0.460	0.635	-3.410	6.414	0.213	11.391
Singapore	0.031	0.732	1.085	-8.646	7.520	-0.794	9.328
Spain	0.044	0.518	0.774	-6.248	8.380	0.279	15.552
Sweden	0.096	0.615	0.935	-7.759	9.437	0.122	15.822
Switzerland	0.136	0.485	0.700	-4.525	5.844	0.060	10.205
UK	0.307	0.594	0.785	-6.313	9.876	0.258	16.320
U.S.	0.005	0.063	0.108	-1.046	1.514	1.046	30.405
Global	0.083	0.420	0.623	-6.196	6.451	-0.428	18.382
North America	0.034	0.163	0.267	-2.715	6.410	3.999	92.346
Europe	0.099	0.445	0.663	-5.951	7.105	-0.180	18.767
Asia & Pacific	0.062	0.651	0.971	-9.890	8.903	-0.679	14.057

Table 1.3: Summary Statistics: ETF Premium- Goetzmann et al. (2001) adjustment

This table shows the summary statistics for the daily ETF premium and for absolute value of the premium (Mean Deviation). The absolute deviation is measured as $\sum_{t=1}^{T} |p_{i,t} - NAV_{i,t}|$, where $p_{i,t}$ and $NAV_{i,t}$ are the log of price and log of net asset value of fund i at time t. NAV is adjusted based on Goetzmann et al. (2001) as in equation 1.2. The standard deviation, minimum, maximum as well as skewness and kurtosis are shown for premium. The bottom of the table shows the cross-sectional average of the summary statistics aggregated by the geographical areas: Global, North America, Europe, Asia Pacific. The data period is from June 2002 until June 2018.

Countries	Mean Premium (%)	Mean Deviation (%)	Standard Deviation (%)	Min (%)	Max (%)	Skewness	Kurtosis
Australia	0.100	0.481	0.681	-5.708	6.060	0.065	10.378
Austria	-0.010	0.507	0.737	-8.280	5.824	-0.716	12.727
Belgium	0.049	0.463	0.683	-6.403	5.487	-0.404	13.068
Canada	0.058	0.323	0.521	-5.537	10.628	2.854	54.941
Denmark	0.105	0.331	0.430	-1.345	2.624	0.491	5.686
Finland	0.114	0.372	0.510	-6.419	4.070	-0.927	24.269
France	0.049	0.410	0.592	-3.591	5.907	0.459	11.336
Germany	0.036	0.406	0.589	-3.087	5.517	0.370	10.314
Hong Kong	0.016	0.531	0.789	-5.450	8.422	-0.050	11.642
Ireland	0.530	0.777	0.852	-2.662	4.972	0.253	3.285
Israel	-0.054	0.489	0.774	-16.551	4.178	-4.403	87.953
Italy	0.042	0.446	0.633	-5.447	4.870	0.100	9.710
Japan	0.061	0.538	0.769	-7.490	7.145	0.115	13.630
Netherlands	0.038	0.414	0.593	-3.513	4.590	0.055	8.807
New Zealand	0.039	0.372	0.500	-2.229	2.723	0.118	5.291
Norway	0.061	0.413	0.572	-2.642	6.537	0.819	14.671
Singapore	0.024	0.580	0.870	-5.406	5.436	-0.893	8.973
Spain	0.037	0.437	0.631	-5.106	5.105	0.320	10.730
Sweden	0.087	0.520	0.757	-6.323	7.147	0.233	11.520
Switzerland	0.129	0.421	0.593	-3.679	3.710	0.267	7.090
UK	0.300	0.524	0.648	-3.470	6.274	0.339	8.195
U.S.	0.005	0.063	0.108	-1.046	1.514	1.046	30.405
Global	0.076	0.291	0.408	-3.272	3.907	-0.085	12.038
North America	0.032	0.166	0.269	-2.810	5.498	2.890	55.608
Europe	0.092	0.359	0.508	-3.683	4.625	0.048	11.421
Asia & Pacific	0.052	0.406	0.585	-4.974	6.107	-0.224	11.355

Table 1.4: Summary Statistics: ETF Premium- Engle and Sarkar (2006) adjustment

This table shows the summary statistics for the daily ETF premium and for absolute value of the premium (Mean Deviation). The absolute deviation is measured as $\sum_{t=1}^{T} |p_{i,t} - NAV_{i,t}|$, where $p_{i,t}$ and $NAV_{i,t}$ are the log of price and log of net asset value of fund i at time t. NAV is adjusted based on Engle and Sarkar (2006) as in equation 1.3. The standard deviation, minimum, maximum as well as skewness and kurtosis are shown for premium. The bottom of the table shows the cross-sectional average of the summary statistics aggregated by the geographical areas: Global, North America, Europe, Asia Pacific. The data period is from June 2002 until June 2018.

Countries	Mean Premium (%)	Mean Deviation (%)	Standard Deviation (%)	Min (%)	Max (%)	Skewness	Kurtosis
Australia	0.097	0.472	0.663	-5.320	4.908	-0.127	8.674
Austria	-0.011	0.491	0.715	-7.371	4.995	-0.885	12.177
Belgium	0.048	0.439	0.652	-6.549	4.234	-0.726	14.498
Canada	0.061	0.306	0.488	-4.267	10.477	3.276	61.915
Denmark	0.111	0.322	0.413	-1.340	2.618	0.400	5.468
Finland	0.117	0.347	0.479	-6.227	4.144	-0.995	27.523
France	0.045	0.363	0.510	-3.189	5.131	0.279	9.623
Germany	0.032	0.355	0.499	-2.446	4.285	0.299	8.170
Hong Kong	0.009	0.491	0.707	-4.484	8.064	0.048	11.188
Ireland	0.534	0.762	0.830	-2.237	4.823	0.250	3.086
Israel	-0.061	0.462	0.726	-16.501	4.503	-4.840	107.938
Italy	0.034	0.408	0.566	-4.826	4.008	-0.068	8.019
Japan	0.057	0.483	0.689	-5.599	6.522	0.230	12.741
Netherlands	0.035	0.378	0.533	-3.420	3.840	-0.231	7.529
New Zealand	0.037	0.362	0.480	-2.034	2.205	0.070	4.814
Norway	0.063	0.392	0.540	-2.398	6.275	0.864	15.299
Singapore	0.017	0.542	0.818	-5.926	4.948	-1.080	9.903
Spain	0.032	0.409	0.578	-4.565	4.155	0.137	8.466
Sweden	0.085	0.480	0.685	-5.511	5.141	0.110	9.491
Switzerland	0.128	0.401	0.564	-3.566	4.029	0.310	7.157
UK	0.296	0.492	0.589	-3.415	4.930	0.226	6.604
U.S.	0.005	0.063	0.108	-1.046	1.514	1.046	30.405
Global	0.074	0.274	0.371	-2.507	2.878	-0.225	8.090
North America	0.033	0.157	0.250	-2.175	5.422	3.308	63.434
Europe	0.089	0.324	0.445	-2.866	3.593	-0.154	8.873
Asia & Pacific	0.048	0.375	0.532	-3.156	5.538	-0.272	9.408

Table 1.5: Absolute Premium and Illiquidity Risks

This table shows the results of a panel regression of the absolute value of ETF premiums on the currency illiquidity (IML) and global market illiquidity ($MILLIQ^G$), as well as monthly fixed effects and country fixed effects run conditionally on whether the fund is traded at the premium or discount.

$$|p_{i,t} - NAV_{i,t}| = \alpha + \beta_1 IML_t + \beta_2 MILLIQ_t^G + FE_i + FE_t + \varepsilon_{i,t}$$

The results are presented for 2 versions of ETF premiums: the one based on Goetzmann et al. (2001) adjustment (FFP^G) and the one based on Engle and Sarkar (2006) adjustment (FFP^E). The IML ($MILLIQ^G$) is measured as a long-short portfolio of currency pairs (stocks) sorted by bid-ask spread (Amihud (2002) illiquidity ratio) and rebalanced daily. The regression is performed at a daily level. The sample is from June 2002- June 2018. ***,**,* show the significance at 1%, 5% and 10%.

	Panel A: Premium		Panel B: Discount	
Variables	FFP^G	FFP^E	FFP^G	FFP^E
	(1)	(2)	(3)	(4)
IML	0.050***	0.038***	-0.049***	-0.015*
	(5.13)	(6.19)	(-3.22)	(-2.06)
$MILLIQ^G$	0.026**	0.014**	-0.049***	-0.031***
	(2.13)	(2.21)	(-3.61)	(-7.15)
$Constant$	0.008	0.009	0.007***	0.007***
	(7.84)	(9.47)	(14.72)	(18.05)
Country FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Observations	42,561	42,563	33,354	33,352
Countries	22	22	22	22
$Adjusted\ R^2$	0.178	0.175	0.236	0.222

Table 1.6: Correlation of U.S. FFP and Proxies for Individual Components of Frictions

This table shows the correlation of financial friction proxy FFP for the U.S. with various proxies for individual components of frictions such as: absolute value of closed-end fund discount, investor sentiments (Baker and Wurgler (2006)),the TED spread, Moody's spread, VIX, change in leverage constraint tightness of Boguth and Simutin (2018), Consumer Confidence Index , Aggregate disagreement of Hong and Sraer (2016) and Pástor and Stambaugh (2003) liquidity proxy. AR(1) residuals of all variables are used when computing the pairwise correlation with changes in LCT. The data is at the monthly frequency and is from June 2002-June 2018. The only exception is Leverage Constraint Tightness which is from June 2002- December 2014 due to data availability. ***,**,* shows the significant at 1%, 5% and 10%.

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) FFP^{US}	1.00									
(2) |Closed-End Fund Discount|	0.18**	1.00								
(3) Investor Sentiments	-0.28***	-0.41***	1.00							
(4) TED Spread	0.66***	0.05	0.12	1.00						
(5) Moody's Spread	0.81***	0.39***	-0.39***	0.60***	1.00					
(6) VIX	0.65***	0.27***	-0.44***	0.47***	0.83***	1.00				
(7) Consumer Confidence Index	-0.55***	-0.30***	0.24***	-0.36***	-0.61***	-0.64***	1.00			
(8) Δ Leverage Constraint Tightness	0.07	0.06	-0.10	0.13	0.04	-0.02	-0.07	1.00		
(9) Aggregate Disagreement	0.24***	0.66***	-0.55***	0.09	0.40***	0.35***	-0.40***	-0.04	1.00	
(10) Aggregate PS Liquidity	-0.41***	0.09	0.10	-0.33***	-0.35***	-0.41***	0.21***	-0.05	0.03	1.00

Table 1.7: U.S. FFP and Proxies for Individual Components of Frictions

This table shows the result of a time-series regression of the financial friction proxy FFP for the U.S. on various proxies for individual components of frictions such as: investor sentiments ($Sent$) of Baker and Wurgler (2006), TED spread (TED), Moody's spread ($Moodys$), ΔVIX, change in Consumer Confidence Index (ΔCCI) , Aggregate disagreement (AD) of Hong and Sraer (2016) and Pástor and Stambaugh (2003) liquidity proxy (PS).

$$FFP^{US} = \alpha + \beta_1 Sent + \beta_2 TED + \beta_3 Moodys + \beta_4 \Delta VIX + \beta_5 \Delta CCI + \beta_6 AD + \beta_7 PS + \varepsilon_t \quad (1.6)$$

The standard errors are based on Newey and West (1987). The data is at the monthly and daily frequency and is from June 2002-June 2018. $Sent$, AD and PS are assumed to be constant during a month for daily regression. ***,**,* shows the significance at 1%, 5% and 10%.

Variables	Monthly	Daily		
	Total	Total	Premium	Discount
$Sent$	-0.002*	-0.001	-0.002**	-0.001
	(-1.76)	(-1.06)	(-2.12)	(-0.69)
TED	0.003***	0.005***	0.005***	0.004***
	(3.10)	(5.38)	(5.07)	(3.90)
$Moodys$	0.006***	0.261***	0.214***	0.326***
	(5.18)	(7.09)	(6.41)	(7.03)
ΔVIX	0.002**	0.048***	0.041***	0.051***
	(2.33)	(5.68)	(9.85)	(9.64)
ΔCCI	-0.171	-	-	-
	(-1.42)	-	-	-
AD	-0.007	-0.015	-0.019	-0.020
	(-1.23)	(-0.16)	(-0.78)	(-0.28)
PS	-0.006	-0.003	-0.002	-0.004
	(-1.42)	(-0.66)	(-0.37)	(-0.78)
$Constant$	0.001	0.006***	0.005***	0.004***
	(0.49)	(7.98)	(7.33)	(4.40)
Observations	189	3,806	2,338	1,468
Adjusted R^2	0.731	0.252	0.315	0.455

Table 1.8: Stock Determinants of FFP Exposure

This table shows the results of a panel regression of FFP betas $\beta_{FFP^{US}_{i,t}}$ of stock i (from the U.S.) at time t on the market beta, log of market value of equity, book to market ratio, number of analysts covering the stock, dispersion of analyst forecasts of EPS, idiosyncratic volatility of residuals from market model, stock leverage measured as a ratio of the value of total liabilities to market capitalization and Amihud (2002) illiquidity ratio (coefficient is shown in 10^2). $\beta_{FFP_{i,t}}$ is computed from equation 1.4 . The regression includes time and industry fixed effects based on 49 Fama-French industry classification. The sample is from May 2007 until June 2018. ***,**,* show the significance at 1%, 5% and 10%.

Variables	(1)	(2)	(3)
Market beta	0.028***	0.036***	0.035***
	(5.10)	(5.91)	(5.83)
Market Equity	-0.016***	-0.016***	-0.016***
	(-4.72)	(-4.27)	(-4.26)
Book to Market ratio	-0.007**	0.002	-0.002
	(-2.38)	(0.45)	(-0.54)
Number of Analysts		-0.002***	-0.002***
		(-4.83)	(-4.86)
Dispersion in Analyst Forecasts		0.003*	0.003*
		(1.81)	(1.71)
Idiosyncratic volatility		-0.223	-0.295
		(-1.63)	(-1.12)
Financial leverage			0.001***
			(3.39)
Amihud Illiquidity			-1.00
			(-1.61)
Constant	0.143	0.087	0.089
	(1.42)	(0.82)	(0.85)
Industry FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Observations	416,723	293,278	292,450
Adjusted R^2	0.12	0.18	0.19

Table 1.9: Stock Portfolios and Financial Frictions: the U.S.

This table shows the average excess returns over the risk-free rate of portfolios sorted by the exposure to financial friction proxy (FFP^{US}). The sample covers stocks with primary listing in the U.S. At the beginning of every month stocks are sorted into quintiles based on ex ante sensitivity to FFP^{US} (measured using 60-month rolling window). Excess returns, intercept from CAPM model, Fama-French 3-factor model and Carhart (1997) 4-factor model (includes MRP, HML, SMB and UMD factors) are reported. Panel A shows the results based on value-weighted portfolios. Panel B shows the results based on equally-weighted portfolios. Returns and alphas are in percent per month. T-statistics is based on Newey and West (1987) standard errors. The sample is from May 2007 until June 2018. ***,**,* show the significance at 1%, 5% and 10%.

Panel A: Value-weighted portfolios						
Portfolio	Low	2	3	4	High	High-Low
Excess Returns	0.871***	0.723*	0.589**	0.653**	0.572*	-0.299**
	(2.81)	(1.70)	(2.19)	(2.04)	(1.91)	(-2.24)
CAPM α	0.322**	0.421*	0.264*	0.085*	0.009	-0.313**
	(2.54)	(1.82)	(1.95)	(1.73)	(1.35)	(-2.41)
FF 3-factor α	0.301**	0.409	0.257**	-0.034*	-0.084	-0.385**
	(2.25)	(1.62)	(1.98)	(-1.78)	(1.14)	(-2.33)
4-factor α	0.289**	0.402*	0.211**	-0.088*	-0.181	-0.470**
	(2.12)	(1.65)	(2.13)	(-1.75)	(-1.62)	(-2.21)
Panel B: Equally-weighted portfolios						
Excess Returns	0.748***	0.616*	0.519*	0.569**	0.477**	-0.271**
	(2.41)	(1.69)	(1.79)	(2.13)	(2.02)	(-2.07)
4-factor α	0.333**	0.454*	0.143**	0.091	-0.119*	-0.452**
	(2.32)	(1.73)	(2.01)	(0.32)	(-1.71)	(-1.97)

Table 1.10: Stock Portfolios and Financial Frictions: International

This table shows the average excess returns (H-L) of portfolios with high exposure to financial friction proxy (FFP) over the portfolios with low exposure to FFP for 21 country in the development market sample. At the beginning of every month stocks are sorted into quintiles based on ex ante sensitivity to FFP (measured based on Goetzmann et al. (2001) adjustment using 60-month rolling window). The reported excess return is the average difference between value-weighted portfolio 5 and portfolio 1 (where 5 is the highest). 4-factor alpha is the intercept from the regression of the long-short portfolio on Carhart (1997) 4 factor model that includes MRP, HML, SMB and UMD factors. Returns and alphas are in percent per month. T-statistics is based on Newey and West (1987) standard errors. The sample is from May 2007 until June 2018. ***,**,* show the significance at 1%, 5% and 10%.

Based on Goetzmann et al. (2001) adjustment				
Country	H-L	t-stat	4-factor α	t-stat
Australia	-0.562***	(-2.79)	-0.811**	(-2.10)
Austria	-0.657**	(-2.01)	-0.546***	(-2.99)
Belgium	-0.292	(-0.75)	-0.301	(-0.71)
Canada	-0.621**	(-2.27)	-0.484**	(-2.52)
Denmark	-1.027**	(-2.33)	-0.873***	(-2.60)
Finland	0.344	(0.80)	0.124	(0.22)
France	-0.502**	(-2.37)	-0.326*	(-1.79)
Germany	-0.281**	(-2.41)	-0.304***	(-2.66)
Hong Kong	-0.487*	(-1.93)	-0.724**	(-2.28)
Ireland	0.186	(0.20)	0.062	(0.02)
Israel	-0.040	(-0.09)	-0.058	(-0.10)
Italy	-0.162**	(-2.52)	-0.237***	(-2.83)
Japan	-0.042	(-0.16)	-0.012	(-0.40)
Netherlands	0.865	(1.50)	0.689	(1.22)
New Zealand	-0.259*	(-1.78)	-0.342**	(-2.26)
Norway	-0.205**	(-2.32)	-0.218**	(-2.17)
Singapore	0.217	(0.66)	0.197	(0.45)
Spain	-0.544*	(-1.73)	-0.631*	(-1.94)
Sweden	-0.221***	(-2.64)	-0.334***	(-2.89)
Switzerland	-0.248**	(-2.30)	-0.214**	(-2.16)
UK	-0.102**	(-2.11)	-0.121**	(-2.35)

Chapter 2

ETF Arbitrage and International Diversification

2.1 Introduction

Significant innovations in financial products made international investments increasingly possible. Over the recent years, exchange-traded funds experienced a double-digit growth in assets under management. Low management fees allow ETFs to compete for market share with more expensive mutual funds and future contracts (e.g., Ben-David, Franzoni, and Moussawi, 2017).[1] Yet, trading across major country ETFs, and its association with local and global uncertainty remains understudied. Country ETFs are a low-cost vehicle for foreign investments in benchmark country indices across the world, and hence provide access to foreign markets, in particular for retail investors. Many exchange-traded fund providers refer to international diversification as one of the key advantages of investing in this type of products.[2] While the majority of earlier studies focus on the effects of ETFs on the underlying securities in the basket that it tracks, I propose a transmission mechanism of U.S. market shocks to foreign country equity markets via ETF trading.

I provide a view that as the U.S. accommodates the largest share of ETF global trading volume, its market conditions directly impact the decisions of country

[1] Roll costs are important when trading futures.

[2] Blackrock: https://www.ishares.com/us/strategies/invest-internationally.

ETF investors.[3] I show that international ETF market participants trade based on shocks related to U.S. fundamentals rather than local ones, and propagate those shocks to local markets. The shock transmission is performed via ETF arbitrage. I argue that such arbitrage activity is one of the few mechanisms responsible for increasing correlation between the U.S. market and the rest of the world. This high cross-country correlation limits the ability of investors to cheaply diversify U.S. risk via international ETF investments. Country ETFs often provide an easier access to less integrated emerging markets or to countries where direct investments are costly (e.g., Brazil). Such ETFs become increasingly popular with iShares Emerging Markets ETF being the second largest ETF by trading volume in the world.[4] However, the transmission of U.S. shocks to those markets limits the diversification benefits of emerging market strategies.

I first test the hypothesis whether country ETF investors react to changes in the U.S. rather than local economic uncertainty, as measured by CBOE Volatility Index (VIX). To this end, I compute order imbalances of different market participants, based on identification of Boehmer, Jones, and Zhang (e.g., 2017) and trade size (e.g., Barber, Odean, and Zhu, 2009; Peress and Schmidt, 2017), with a particular focus on retail activity. Using a large cross-section of 41 countries, I find strong association between ETF order imbalances and U.S. VIX, indicating that international investment decisions are mainly driven by the latter measure, rather than its local counterparts. For example, an increase in the U.S. VIX results in a selling pressure in the country ETF market. Such result is robust to different volatility regimes and is consistent across different types of investors. Asymmetric response analysis confirms that country ETF investors only react to positive changes in local VIX, which correspond to negative news in the local markets. Moreover, I observe that, when reacting to an increase in U.S. uncertainty, investors switch to safer assets such as cash equivalent ETFs. I also decompose the VIX into the economic uncertainty and risk aversion

[3]According to Deutsche Bank's ETF Annual Review & Outlook just for equity ETFs turnover in U.S. was $16.38 trillion out of $18.26 trillion globally in 2016.

[4]http://etfdb.com/compare/volume/.

components and find that traders react more strongly to changes in U.S. economic uncertainty rather than the risk aversion component of the VIX.

I also find that investors respond to changes in U.S. political uncertainty differently than to economic uncertainty- they leave the U.S. stock market and buy international country-level ETFs. However, they do not react to local political uncertainty and the economic effect of political risk is much smaller than of changes in U.S. VIX. My result is related to Levy and Lieberman (2013) who show the overreaction to U.S. returns during non-synchronised trading hours. They observe that since ETF and local market (especially Asian) are open during different market-hours, intra-day price formation is often driven by S&P 500 returns, rather than changes in net asset value of the fund. In contrast to their study, I focus on order imbalance rather than returns, as it allows me to assess the trading decisions of different types of investors. My analysis is both on the daily and monthly level, and as such I alleviate non-synchronicity effects. Furthermore, I utilize a much broader cross-sectional sample. In addition, my result is in line with a recent study of Converse, Levy-Yeyati, and Williams (2018) who show that ETF fund flows are much more (less) sensitive to global (local) risk factors than mutual funds. Authors relate this effect to ETFs attracting uninformed investors. In contrast, my analysis shows the reaction to U.S. risk is common to all investor types.

A large set of literature focuses on the effect of one central economy, the U.S., on the rest of the world. Rapach, Strauss, and Zhou (2013) highlight the leading role of the U.S. market, and show the predictability of country-level returns by U.S. returns. Miranda-Agrippino and Rey (2015) examine the effect of U.S. economy on global financial variables (e.g., cross-border credit flows, leverage etc.). They highlight the role of a global factor that can explain a large portion of the variation in global asset returns, and is related to global risk aversion and aggregate volatility. Atanasov (2014) shows that a single "global consumption factor" can explain more than 70% of cross-sectional variation in stock returns. Rey (2015) documents the existence of a global financial cycle. A "central country" has an impact on leverage of banks, growth and availability of credit across the world. CBOE Volatility index (VIX), that is, implied volatility of options written on S&P 500, is often used as a measure of uncertainty and is generally perceived as an indicator of market fear. VIX is significantly correlated with a global risk factor affecting international stock returns. Rey (2015) shows that when VIX is low for a long time, there is a boom in the global financial cycle and inflation of stock prices. In contrast, high values of VIX are negatively associated with capital inflow, credit growth and leverage in all of the main financial centres across the world. Forbes and Warnock (2012) highlights the relationship between VIX, as a proxy for global risk, and international capital flows. Therefore, I use VIX as a key variable in my empirical analysis in the context of passive international investment via ETFs.

The central role in earlier models is often allocated to large international banks (via leverage and risk appetite) that use U.S. dollars as a funding currency and provide credit across borders and foreign direct investments. Another set of literature studies the role of mutual funds that are being affected by investor redemptions during crisis periods on international transmission of shocks (e.g., Jotikasthira, Lundblad, and Ramadorai, 2012; Raddatz and Schmukler, 2012). However, the role of indirect investment via ETFs is often overlooked. In contrast, I look at the ETF arbitrage that propagates both fundamental and non-fundamental shocks to the underlying economies. I argue that due to ETF arbitrage mechanism, U.S. fundamentals get incorporated into local market returns, which in turn results in a high positive correlation of local market returns with the U.S. market.

Another strand of the literature studies the role of increased correlation during periods of high volatility (e.g., Solnik, Boucrelle, and Le Fur, 1996). In particularly, an important concern is the comovement of countries during crisis periods, but there is a disagreement on what can be classified as contagion. Forbes (2012) treats it more generally, as a strong negative shock that is transmitted to other countries. He suggests that there are four important and often interlinked channels for contagion: trade, banks, portfolio investors, and wake-up calls. In contrast, Bekaert, Ehrmann, Fratzscher, and Mehl (2014) define contagion as an increase in correlation across stocks beyond what can be explained by fundamentals. They explore different types of contagion during recent financial crisis and find the support for wake-up call hypothesis. According to wake-up call hypothesis (first proposed by Goldstein, 1998), the crisis originated in one market provides a new information about true value of fundamentals to other markets. I complement this literature by showing that ETF arbitrage mechanism is an important channel that transmits U.S. shocks to individual countries and hence increases cross-country equity market correlations beyond crises.

In order to access the impact of ETF arbitrage on correlation of country returns with the U.S. market, I regress monthly innovations in such correlation both on a dummy variable capturing staggered introduction of ETF markets across countries, and on a proxy for ETF arbitrage during different volatility regimes. I provide time-series evidence that during periods of high volatility in the U.S., the introduction of the ETF market, and more importantly, an increase in the arbitrage activity by the authorized participant (AP) (as measured by net share creation/redemption) in the ETF market results in an increase in innovation in such monthly correlation of the underlying country stock market indices. The latter result is consistent with the literature on global contagion and, in particular, wake-up call hypothesis. I argue that during periods of high volatility in the U.S. market, it is harder for investors to distinguish between noise and fundamental component of the order flow. Consequently, based on wake-up call hypothesis investor may treat U.S. shocks as relevant to their own country and consume such shocks via ETF arbitrage.

I also explain cross-country variation in return correlation with the U.S. market. According to Ben-David et al. (2018), non-fundamental shocks must be reversed over time. This suggests that if all shocks transmitted from ETF market to local economies were non-fundamental, ETF arbitrage would not contribute towards increased correlation. In contrast, if the price deviation from the NAV is due to faster incorporation of fundamental information in ETF market, then arbitrage should affect returns of underlying index, and such effect should not be reverted. If such fundamental information is common both to U.S. and local market, one should observe a higher correlation between them. Section 2.2 provides the details of this mechanism which leads to increased correlation. Consistent with the literature, I argue that ETF transmits both fundamental and noise shocks to the underlying economies. I show that countries that have a higher degree of price discovery in their ETFs have on average a higher correlation (integration) with the U.S. market. In these markets fundamental information gets incorporated into ETF prices faster than in the Net Asset Value (NAV), and therefore, market makers closely follow and learn from changes in ETF prices. This is the case when derivative securities price the underlying assets, rather than the other way around. In addition, in order for fundamental shocks to get transmitted to underlying markets, the authorised participants (AP) must engage in arbitrage activity. I find that the lower the limits to ETF arbitrage, the higher is the correlation (integration) between a country and the U.S. market. Neither the international trade channel nor the business cycles alter this result. This is in line with Rozental (2019) who argues that the AP may not engage in the arbitrage activity if the reward for facing ETF arbitrage risks is not high enough.

Most of ETF research focuses on evaluation of negative and positive consequences of ETF trading on underlying markets (see Ben-David et al. (2017) for a survey of recent ETF literature). Malamud (2016) theoretically shows that demand shocks can be propagated to the underlying markets. A strong debate is about whether such shocks reflect fundamental information that is incorporated into ETF faster than to NAV or reflect non-fundamental liquidity shocks that diminish information efficiency of underlying stock prices. There is mixed evidence of both effects. Glosten, Nallareddy, and Zou (2016) show that ETF trading can partially transmit information about systematic fundamentals to the underlying stocks leading to information efficiency improvement. Lettau and Madhavan (2018) and Wermers and Xue (2015) argue for the existence of price discovery in ETF market. Madhavan and Sobczyk (2016) make a similar argument and develop a theoretical model that incorporates both noise and price discovery in ETF prices. Marshall et al. (2013) find that when underlying liquidity of stock market is low, ETF prices adjust faster than NAV. In contrast, Ben-David et al. (2018) show that ETFs increase the volatility of underlying assets due to propagation of noise via arbitrage mechanism. They show that such additional layer of volatility is non-fundamental. Israeli et al. (2017) show that increase in ETF ownership leads to a rise in trading costs of the underlying markets and a potential shift of retail traders to ETF market leading to a decline in information efficiency over a longer term (due to less analyst coverage). Brown, Davies, and Ringgenberg (2019) demonstrate that arbitrage activity in ETF market negatively predicts future stock returns suggesting non-fundamental based view. Da and Shive (2018) show that arbitrage in U.S. domestic ETFs can cause an excessive comovement amount stocks in the underlying basket of securities. They show that shocks propagated from ETF markets also include non-fundamental ones due to price pressure, and are reflected in negative autocorrelation in stocks and ETF returns. My study complements this literature and studies the consequences of price discovery process in ETF market on the cross-country correlation.

Section 2.2 describes the link between ETF arbitrage and cross-country correlation. Section 2.3 introduces the data sources and the construction of key variables. Section 2.4 provides the empirical results. Section 2.5 shows the results of my robustness tests. Section 2.6 concludes.

2.2 ETF Arbitrage and Correlation: Mechanism

ETF arbitrage mechanism is a unique feature of the market that theoretically allows prices to continuously track underlying stocks. A fund is traded at a premium (discount) when ETF price is higher (lower) than the NAV. The AP (designated dealer in the ETF market) has an incentive to correct the emergence of arbitrage by placing opposing trades in local and ETF markets. For example, to correct the

ETF premium AP can sell ETF shares and buy underlying basket of securities. The constituents of this basket are published daily by ETF sponsor in the portfolio composition file (PCF). Then, at the end of the day AP is able to deliver and exchange such basket of securities to ETF sponsor for newly created ETF shares ("in-kind" creation). As a result of such arbitrage activity the gap between ETF price and NAV should be closed.

Despite the existence of such mechanism deviations of prices from the NAV of the fund are common. Pan and Zeng (2019) show that when there is a liquidity mismatch between ETF and underlying market, APs may not be willing to correct the existing deviations. Petajisto (2017) highlights the existence of limits to arbitrage in ETF market especially for international funds. He shows that such deviations can last for days. I am interested in the consequences of such arbitrage incidents on the return of the underlying index. As argued by Ben-David et al. (2018) when arbitrage mechanism transmits non-fundamental shocks to underlying stocks, over time, stock and ETF prices move back to fundamental levels. In contrast to such view, Madhavan and Sobczyk (2016) argue that although the indicative NAV is published throughout the day (every 15 seconds) the "true" NAV is often hard to estimate. U.S. market often trades when underlying markets are closed, in which case NAV is a closing value of a previous day corrected by foreign exchange return. International equity ETFs specifically suffer from such a problem. In addition, for a basket with a large composition the correct estimation of the total value of assets is often complicated. As such, any deviation between price and NAV can be either due to transitory liquidity shocks or due to price discovery in ETF market.[5]

Similarly to this view, I argue that any deviation between ETF price and NAV reflects a mix of noise and fundamental information. The mechanism is illustrated in Figure 2.1. When U.S. investors experience an increase in VIX (e.g., bad news about future U.S. fundamentals arrives in the market) the following happens:

1. Both retail and institutional U.S. investors negatively react to increase in

[5]Appendix B.1 provides a description of the model of Madhavan and Sobczyk (2016).

market uncertainty and sell ETF of country A. I show such response in section 2.4.2. Investors also sell S&P constituents (negative return of U.S. market).

2. A sell-off of ETF leads to a decrease in its market price below the NAV of the fund. When the decrease is significant enough and limits to arbitrage are low, AP intervenes exploiting the ETF arbitrage mechanism outlined above.

3. AP buys ETF shares and short-sells (or reduces his inventory) the underlying stocks of country A (in a correct proportion in line with portfolio composition file). As a result, the prices of underlying stocks are reduced. ETF shares are delivered to ETF sponsor and get redeemed for the underlying stocks. AP closes his short position.

4. If local dealer uses ETF market to price the underlying market (i.e., price discovery happens in the ETF market), the decrease in prices of underlying stocks is permanent (negative return of county's A market). This results in positive correlation between the U.S. market and country A. If decrease in ETF price is considered as noise, that is, if VIX changes are not fundamental news for country A, both ETF and underlying stock prices will be moved back to fundamental level (as in Ben-David et al., 2018) and therefore, there should not be any positive effect on correlation. This mechanism will be impaired, if the price discovery in the ETF market is distorted due to noise trading.

[insert figure 2.1 here]

2.3 Data and Methodology

In this section I discuss my sample construction, different measures of order imbalances used in this paper, as well as the methodology for obtaining proxies for economic and political risk sentiments. I begin by describing the type of funds that are considered in this paper and proceed with a discussion of methods for capturing trading activities of different types of investors, as well as proxies for economic and political uncertainty.

2.3.1 MSCI Country Indices and iShare ETFs

My focus is on exchange-traded funds provided by iShares (Blackrock, Inc.) that track a general MSCI index of a single country[6], do not hedge their currency exposure and are traded on one of the U.S. exchanges. The final sample consists of 41 funds traded on NYSE Arca, NASDAQ or CBOE BZX (Bats) (see table B.1 in the appendix for details of ETFs). U.S. ETF market is one of the most developed and represents a significant portion of the world ETF trading volume. My sample covers developed and emerging economies and has a wide geographical reach: 22 ETFs are from Europe, Middle East or African; 13 ETFs are from Asia and Pacific regions; 6 ETFs are from Latin and North America. The majority of my analysis is on a daily level and covers the period of January 2006 - June 2018. I obtain ETF prices, MSCI daily index (in USD) and its turnover from Thomson Reuters Datastream and Bloomberg. Officially published end of day net asset values (NAV) of funds are available directly from iShares website.

2.3.2 Total Order Imbalance

I obtain intra-day quote and trade data for 41 ETFs from TAQ database. Total ETF order flow (TOI) is constructed by matching quote and trade data from TAQ database using Holden and Jacobsen (2014) time interpolation method. I use Lee and Ready (1991) algorithm to sign the trades (see the appendix B.2 for more details on these two methods). The order imbalance is calculated as a difference between buyer and seller dollar volume scaled by the total dollar-volume on a particular day.

$$TOI_t = \frac{buyers_t - sellers_t}{buyers_t + sellers_t} \tag{2.1}$$

[6]Many of single country ETFs in my sample are on the list of top 100 funds by traded volume on etfdb.com. In my sample Brazil , Japan, China, Taiwan, India, Hong Kong, Mexico, Germany and South Korea are the most popular.

I also consider order imbalances of different trader types based on trade size and retail investor identification (Boehmer et al., 2017). Such differentiation allows me to conduct a deeper analysis of similarities and differences between these types of investors in the ways they react to new information.

Order Imbalance: Small vs. Large Trades

ETFs are designed for retail investors. Lack of access to foreign markets and inability to invest directly into underlying securities due to significant cost barriers (e.g., trading costs) are only a few reasons why individual investors use ETFs to invest in general market indices. Due to several institutional factors (e.g., commission based advisory practice) retail participation in European ETF markets is still low. In contrast, participation of U.S. retail investors in this market is relatively higher. As such, in contrast to other markets the analysis of this type of investors in the context of U.S. ETF trading is important.

I first consider a measure of retail trading activity based on trade size (Peress and Schmidt, 2017). Using equation 2.1, I obtain order imbalance for small trades (OI_s). Retail trades are usually identified as the smallest trades of less than or equal to $\$5,000$ (e.g., Barber et al., 2009). The limitation of this method is that with the rise of high-frequency algorithmic trading orders are often sliced into small quantities (e.g., Hendershott, Jones, and Menkveld, 2011) and therefore, small trades are likely to be a noisy measure of retail trading activity.

High-frequency traders often submit a large number of quotes that do not result in trades in order to uncover the direction of the market. Some exchanges introduced a fee to deter such activity.[7] Hangströmer and Lars (2013) distinguish between two types of high-frequency trading: market making strategies and opportunistic trading. They show that order-to-trade ratio (OTR) is much higher for the former group. High-frequency market makers tend to have zero inventory on average.

[7] For example see Friederich and Payne (2015) on regulatory fees in Borsa Italiana, Malinova, Park, and Riordan (2018) in Canada and Jørgensen, Skjeltorp, and Ødegaard (2017) in Oslo Stock Exchange.

Therefore, their trades might only reflect temporary inventory adjustments and do not contain any additional information about the direction of the market.

I follow Skjeltorp, Sojli, and Tham (2015) to construct the OTR. The variable is based on the number of daily quote updates in TAQ database relative to the total number of executed trades. My measure of quote updates includes any changes in the best bid or ask prices across all exchanges, as well as changes in quantities at such prices.

In order to clean the order imbalance measure from the effect of high frequency market making activity, I regress the raw measure of small trades (OI_s) on the OTR:

$$OI_{s,i,t} = \alpha + \beta_{1,i} OTR_{i,t} + \varepsilon_{i,t} \tag{2.2}$$

I take a residual from equation 2.2 to find an order imbalance that is uncorrelated with a measure of high frequency market making activity. Thus, I denote as $smallOI_{i,t}$ the innovation of equation 2.2, which serves as a cleaner proxy for small retail trades. However, I acknowledge that small trades are also likely to capture the activity of institutional investors who use high frequency algorithms to minimise the impact of their trades (through smaller trade size and by routing orders to more liquid trading venues). Therefore, I capture the retail trading activity with an alternative measure following Boehmer et al. (2017).

I also compute order imbalance for trades over a threshold of $\$20,000$ ($largeOI$) in a similar fashion to Barber et al. (2009). This measure presumably captures the trading activity of institutions that are less likely to use sophisticated high-frequency algorithms to slice the orders (e.g., pension funds etc.).

Retail Order Imbalance

I use an alternative method suggested by Boehmer et al. (2017) to sign the retail trades. The authors recognise that in contrast to institutional orders many retail trades are happening off-exchange and are internalized or executed by a wholesaler.

Such trades are reported to FINRA Trade Reporting Facility (TRF), marked with exchange code D in the TAQ trade database and usually executed at prices slightly above National Best Bid or Offer. The retail seller initiated transactions receive a small price improvement and are identified by prices with a fraction of a penny in a range of (0, 0.4). The retail buyer transactions receive a price improvement as a discount and are identified by prices with a fraction of a penny in a range of (0.6,1). The Boehmer et al. (2017) type of retail order imbalance ($retailOI$) is then calculated in the same way as in equation 2.1.

One of the limitations of this measure is that it only incorporates market orders, while retail traders often use limit orders. Nonetheless, Boehmer et al. (2017) suggest that more than half of trades on NYSE are captured by this methodology. Interestingly, Boehmer et al. (2017) find that this measure captures informed retail order flow rather than noise trading.

2.3.3 Economic Uncertainty

In order to see how the order imbalance of different groups of investors react to new information I measure changes in economic uncertainty in the U.S. and in the underlying economies. I use CBOE Volatility Index (VIX) as a proxy for U.S. economic uncertainty. The data is available from CBOE website. I also obtain local alternative of VIX (LVIX) from Bloomberg. Not all countries in my sample have a local version of VIX. For a few European countries I have to substitute LVIX with a general European index (VSTOXX Volatility Index). Table B.2 in the appendix provides the summary statistics of VIX and LVIX. Changes in VIX and LVIX are positively correlated for all countries, but such correlation is not high. Nevertheless, I orthogonalise changes in LVIX to changes in VIX and conduct the analysis using uncorrelated variable $\Delta LVIX^o$.

2.3.4 Political Uncertainty

In order to control for other sources of uncertainty, I follow Da, Engelberg, and Gao (2011, 2015) and Filippou and Li (2018) to construct a daily political uncertainty measure. The methodology is based on changes of frequency of word searches in Google. I first obtain the list of words from Harvard dictionary that is classified as "political". I then download the search volume index (SVI) for each of the word on the list, as well as for top searches that include these words. I take a view of U.S. investor and, as such, I only obtain the data for U.S.-based searches.[8] I compute the difference in SVI across time. I first winsorize then remove seasonality and standardise the data. Finally, similarly to Da et al. (2015), I select the most relevant words by performing an expanding backwards rolling regression of ΔSVI_i for word i on country j MSCI index return.

$$\Delta SVI_{i,t} = \alpha + \beta_i \Delta MSCI_{j,t} + \varepsilon \tag{2.3}$$

The political uncertainty measure is then constructed as a cross-sectional average of ΔSVI of the words with the most negative t-statistics for β_i. Intuitively, an increase of the political uncertainty measure is associated with increasing "fears" of the households regarding the political conditions of the country.

2.4 Empirical Results

In this section I evaluate the effects of changes in local and U.S. economic and political uncertainty on the trading decision of international ETF investors. I highlight the similarities and differences in the reaction of different investor types (as outlined above). I further show the implication of such decisions on the correlation of countries with the U.S. market and emphasize the role of ETF trading as a transmission mechanism of U.S. uncertainty to local markets.

[8] Data is available from https://trends.google.com/trends/.

2.4.1 Descriptive Statistics

Figure 2.2 shows the 36-month rolling correlation between returns on S&P 500 index and the rest of the world (ROW, proxied by returns on MSCI EAFE index). Over the last 14 years the value of correlation was volatile but high on average. I observe the minimum correlation of around 0.7 right before the financial crisis in 2007 and in the middle of 2015. During financial crisis the correlation significantly increases and reaches its peak of more than 0.9 in 2009. Such high correlation is in line with the evidence that cross-country correlation increases at times of high market volatility. Figure 2.2 also shows the 36-month moving average of VIX level. The rolling correlation between the U.S. market and the ROW closely follows the slow-moving fluctuations in VIX. Overall, correlation with the U.S. market experiences significant time variation, but remains high during my sample period.

[*insert figure 2.2 here*]

Table 2.1 provides the summary statistics of each of the four measures of order imbalance. On average, order imbalance for all type of traders is small but positive, suggesting that over my sample these investors are net buyers of ETFs. Among all countries the largest average order imbalance is for Colombia (0.16) and Saudi Arabia (0.12), while the most negative is for Peru (-0.03.). The reason why the first two countries mentioned above are in the top of the ranking can be due to the fact that the ETF markets for Colombia and Saudi Arabia are very undeveloped and not very liquid. Over my sample, there has been a number of days with net order imbalance of 1 for such countries (likely due to only a few buy trades per day).[9] For most of the countries standard deviation of *smallOI* is the lowest. Table 2.2 shows the correlation across different types of order imbalance for each country. All of these measures tend to be positively correlated, but the correlations are relatively low, on average. This suggests that I capture the trading activity of different set of investors with these measures.

[9]This illiquidity problem is not significant in my sample with most of the countries (even for Colombia and Saudi Arabia) having a significant number of buys and sells per day.

[*insert tables 2.1 and 2.2 here*]

2.4.2 The Role of VIX: Country-level and Panel Results

I first analyse how ETF traders make their investment decisions. My main interest is to investigate the key risk factors affecting investors' order imbalances. In order to understand how an increase in U.S. uncertainty affects market for foreign country ETFs I regress daily total ETF order imbalance at time t on the percentage change in VIX at time $t-1$. Table 2.3 shows individual country-level results of such regression. I control for local market effect by including changes in local version of VIX ($LVIX^o$) and changes in local interest rates (IR).[10] I also control for autocorrelation including lagged order imbalance. I select the optimal number of lags of ΔVIX and $\Delta LVIX^o$ using the combination of Akaike and Bayesian information criteria (e.g., AIC and BIC). My model takes the following form:

$$OI_{i,t} = \alpha + \sum_{k=1}^{5} \beta_k \Delta VIX_{t-k} + \sum_{k=1}^{2} \gamma_k \Delta LVIX^o_{i,t-k} + \delta_1 \Delta IR_{i,t-1} + \delta_2 OI_{i,t-1} + \varepsilon \quad (2.4)$$

If the local market drives the order imbalance in the exchange-traded fund market then I would expect the significance of the local VIX to be over and above the significance of the U.S. VIX. In contrast, Table 2.3 shows that U.S. VIX coefficients are significant for almost every country. More importantly, every significant coefficient is negative implying that investors are more likely to increase their selling pressure on country ETFs in anticipation of higher U.S. economic uncertainty. In contrast to U.S. volatility index, its local alternative is almost never significant when included in my model, suggesting that it is the U.S. uncertainty that has the first-order effect on international country investments, rather than the uncertainty of target countries. The significant results of my predictive regression suggest that there is a delay in reacting to past information. Very few countries have significant local VIX coefficients. Among those are Canada, Finland, Germany, Sweden and the UK.

[10] I use 3-months deposit rate available at a daily frequency from Datastream. If $LVIX$ or IR is not available it is omitted from the regression.

These are relatively big regional centres and, as such, investors in those countries may also pay attention to the information generated in the local market. Overall, the predictive results suggest that for most of the countries in my sample, an increase in U.S. uncertainty leads, on average, to a sell-off of country ETFs in the secondary market.

[*insert table 2.3 here*]

I also run a predictive panel regression with random effects for all countries in my sample. In addition to variables used in previous regressions I add two extra controls: dummy variable (L) that takes the value of 1 if a country has a common language with the U.S. and 0 otherwise; and a constant that represents the geographical distance of the country from the U.S. (G). Both of these variables capture the difficulty of the information generated in the foreign country to flow to the U.S. market. I expect that in the presence of information frictions, lack of common language and long distance between countries make it harder to acquire news about local fundamentals.

$$\begin{aligned} OI_{i,t} = \alpha + \sum_{k=1}^{5} \beta_k \Delta VIX_{t-k} + \sum_{k=1}^{2} \gamma_k \Delta LVIX^{o}_{i,t-k} + \\ + \delta_1 \Delta IR_{i,t-1} + \delta_2 L_i + \delta_3 G_i + \delta_4 OI_{i,t-1} + \varepsilon \end{aligned} \tag{2.5}$$

Table 2.4 shows the panel results of such regression. I provide the results for all versions of the dependent variable. In addition, I perform the difference in coefficients test for every version of order imbalance. On its own, local VIX is significant in every case while the inclusion of control variables does not reduce its significance. In addition, the consideration of two lags exhibit strong predictive ability for every type of order imbalance. However, when U.S. VIX is added in the model (always negative and highly significant) the loading on $\Delta LVIX$ becomes smaller demonstrating the economic significance of VIX relatively to the local measure. The loadings on local VIX are still negative and significant for TOI and $smallOI$, but not for retail and large order imbalance measures. In full specification, the loading on ΔVIX

is always at least 3 times larger than for $\Delta LVIX$. The results suggest that one standard deviation increase in ΔVIX implies a reduction in TOI by \$0.87 million on average[11]. I also report the results of the contemporaneous regression in table B.4 of the appendix. The results are quantitatively similar with even stronger reaction to ΔVIX.

[*insert table 2.4 here*]

Overall, the results suggest that traders mainly react to changes in U.S. uncertainty. The effect of local news that relates to future uncertainty has only a second-order effect on order imbalance of country ETF investors.

2.4.3 VIX and Cash ETFs

My results suggest that when observing an increase in U.S. volatility index investors tend to sell international country-level ETFs. In this section, I investigate whether investors switch to other types of ETFs after reducing their positions in international ETFs. U.S. VIX is a widely used proxy for investors' fear. An increase in this measure results in an increase in aggregate risk aversion of investors (e.g., Adrian, Stackman, and Vogt, 2019; Rey, 2015). I hypothesize that investors would move their funds to more safe-heaven assets when they expect high volatility in the U.S. stock market.

ETF.com is a website that lists all ETFs and groups them under different asset classes and segments. This is a useful tool to obtain a list of all ETFs that are classified as "cash ETFs". I select ETFs under "Fixed Income: U.S. Government Treasury Cash Equivalents" category. These funds invest in U.S. treasuries with less that 1 year maturity. I compute order imbalance for such funds and take a cross-sectional average. Table 2.5 shows the individual country-level results of U.S. cash order imbalance ($cashOI$) on ΔVIX_{t-1}, $\Delta LVIX^{o}_{t-1}$ and other control variables.[12]

[11] This is based on the average dollar trading volume of \$60 million per day.

[12] Since my depended variable $cashOI$ is the same for every country, I only show those countries for which local VIX is available. I use Italy as an example of countries for which VSTOXX is used as a substitute for local volatility. The full table B.14 is reported in the appendix.

Unlike $\Delta LVIX$, ΔVIX is significant for every country. In contrast to previous results the loading is positive, which suggests that investors buy cash equivalent ETFs when facing an increase in U.S. economic uncertainty. The increase in risk aversion and future volatility makes the expected returns of investing in international ETFs less attractive to U.S. investors forcing them to find safer investment opportunities.[13]

[*insert table 2.5 here*]

2.4.4 Asymmetric Reaction to Changes in VIX

Previous results do not consider the direction of changes in my proxy for economic uncertainty. In this section I investigate if ETF traders react symmetrically to positive and negative changes in future uncertainty (in the U.S. via ΔVIX and locally via $\Delta LVIX$). I split changes in VIX and LVIX into quintiles by size. I define five variables $\Delta VIX^{Q1} \dots \Delta VIX^{Q5}$ as follows:

$$\Delta VIX^{Qn} = \begin{cases} \Delta VIX, & \text{if } \Delta VIX \in \text{quintile } n \\ 0, & \text{otherwise} \end{cases} \tag{2.6}$$

Group 1 represents the most negative changes (good news), while group 5 includes only the most positive changes (bad news). Similarly I define five variables for local volatility index $\Delta LVIX^{Q1} \dots \Delta LVIX^{Q2}$. Table 2.6 shows the result of considering asymmetric response in predictive regression. I include total lagged changes in U.S. and local versions of VIX, as well as variables for groups 2 to 5 (as defined above):

$$\begin{aligned} OI_{i,t} = \alpha + \sum_{k=2}^{5} \beta_k \Delta VIX_{t-1}^{Qk} + \sum_{k=2}^{5} \gamma_k \Delta LVIX_{t-1}^{o,Qk} + \\ + \sum_{I=1}^{5} \eta_I \Delta VIX_{t-I} + \sum_{I=1}^{2} \delta_I \Delta LVIX_{t-I}^{o} + \boldsymbol{\mu} Controls + \varepsilon_{i,t} \end{aligned} \tag{2.7}$$

[13]Interestingly, I find that local or U.S. VIX cannot explain the variation of gold ETF.

where the set of $Controls$ includes changes in interest rates (ΔIR), order imbalances (OI) at time $t-1$ as well as a dummy variable for a common language (L) and a distance measure to the U.S. (G). In such specification, the coefficients η_k (δ_k) represent the general reaction of investors to changes in VIX (LVIX). I see the reaction to U.S. VIX for every type of order imbalance, except for *smallOI*. On average, investors tend to react more to significant bad U.S. news (β_5 is significant for TOI), except for *retailOI*. Both small and retail investors exhibit a significant reaction to good U.S. news (β_2).

When splitting local VIX, I can see that the result is different to the previous one. On average, investors react only to large negative news (i.e., positive β_5). One of the possible reasons why investors react to increases to local VIX could be difficult or costly short-selling of the underlying foreign stocks during periods of high volatility. Li and Zhu (2018) argue that ETFs are often used for "synthetic short-selling". In order to short-sell a stock speculators can go short on the ETF and simultaneously take a long position in a group of other underlying stocks. Such approach allows trading stocks at times when the direct short-selling may be difficult. Shorting via ETF was particularly easy in the past, as before the introduction of alternative up-tick rule in 2011, ETFs were exempt from such price test (Miffre, 2007). Another potential explanation of such asymmetric reaction to negative news could be barriers to information that may make it more difficult to acquire foreign data. As such, only significant foreign negative news are likely to capture the attention of U.S. based investors. This explanation is consistent with salience theory of Bordalo, Gennaioli, and Shleifer (2012). Investors have a limited attention and are attracted only by significant changes in VIX.

[*insert table 2.6 here*]

2.4.5 Political Uncertainty

While VIX represents the economic uncertainty and general risk aversion of market participants, I also test the effect of political uncertainty on investment decisions of U.S. investors. Filippou, Gozluklu, and Taylor (2018) show that political (that is, changes in monthly ICRG policy risk index) and economic risks (measured by VIX) are different. However, Boutchkova, Doshi, Durnev, and Molchanov (2012) suggest that political risk is one of the drivers of stock volatility. As such, changes in VIX may reflect both economic and political uncertainty. I include my proxy for political uncertainty (see section 2.3.4) to measure them separately, and to investigate if investors react differently to different sources of uncertainty. Table B.3 in the appendix shows the summary statistics for political measure and its correlation with changes in LVIX. I also observe low correlation between two proxies for political and economic uncertainty in my sample.

Table 2.7 provides the results of regression of different versions of order imbalance on my proxy for changes in U.S. and local political sentiments, economic uncertainty and controls. I find that investors react to lagged U.S. political risk differently than to ΔVIX. While all types of investors react to an increase in U.S. economic uncertainty by exiting the international stock market (via selling country-level ETFs), investors react to U.S. political risk in a different way. They tend to exit the U.S. stock market and move their capital to international stock markets (reflected in positive loadings on $USPU_{t-1}$). Such different response seems to suggest that investors treat the U.S. economic uncertainty as a proxy for a "global risk" and political uncertainty as the U.S.- specific risk. While investors with small trades (*smallOI*) respond faster to $USPU$, others react with a delay. I also observe that *retailOI* is less sensitive to this type of risk. I do not find any strong evidence that investors react to my proxy for local political uncertainty.

As expected, the loadings on ΔVIX still remain significant both in statistical and economic terms above and beyond the political uncertainty measure, suggesting that my previous results are mainly driven by economic uncertainty which is the key determinant shaping the investment decisions of country ETF investors.

2.4.6 Correlation with the U.S. Market and ETF Arbitrage

I show that U.S. investors only marginally consider foreign risks when making investment decisions. U.S. uncertainty affects all types of ETF investors and determines the direction of their trades. If such risk is transmitted via ETF market the returns of countries whose ETFs are actively traded in the U.S. are likely to be correlated with U.S. market returns.

In order to avoid microstructural noise due to desynchronised trading hours I conduct the analysis using monthly returns. I measure market connectedness to the U.S. by computing 36-month rolling correlation ρ between returns of S&P 500 index that represent the U.S. market and local MSCI index that is tracked by a corresponding country-level iShares ETF. Figure 2.3 shows the map of correlations for 41 countries. My sample is geographically dispersed and covers developed and emerging economies. The choice of monthly frequency for computation of correlation eliminates the time-zone effect: Australia is highly correlated with the U.S. despite having the largest time difference.

[*insert figure 2.3 here*]

Many papers such as Solnik et al. (1996) show that international correlation tends to be higher during periods of high volatility. My focus is on innovations in correlation ($\Delta\rho$), which are computed as residuals from first-order auto-regressive model for ρ. I create a dummy variable ($D_{US|L,t}$) that takes the value of 1 when U.S. volatility, as measured by VIX, is high (larger than its mean plus 1 standard deviation), but local volatility index is not high:

$$D_{US|L,t} = \begin{cases} 1, & \text{if } VIX \geq \mu(VIX) + \sigma(VIX) \text{ and } LVIX < \mu(LVIX) + \sigma(LVIX) \\ 0, & \text{otherwise} \end{cases} \tag{2.8}$$

Firstly, I exploit the fact that ETFs were introduced at different times throughout the sample (see table B.1 in the appendix for fund inception dates). I create the dummy variable $Intro^{ETF}$ that takes the value of 1 at fund inception date and throughout its life. In other words, such dummy reflects if the ETF is traded in the market. Table 2.8 shows the result of regressing the innovation in correlation on the introduction dummy, volatility dummy, as well as the interaction of the two variables:

$$\Delta\rho_t = \alpha + \beta_1 Intro_t^{ETF} + \beta_2 Intro_t^{ETF} D_{US|L,t} + \beta_3 D_{US|L,t} + \beta_4 \Delta\rho_{t-1} + \varepsilon_{i,t}$$

My sample for this regression is extended from January 1988 until June 2018 to ensure that I have at least 8 years of data before the introduction of the first ETF.[14] The results show that on its own the high volatility environment does not lead to an increase in the correlation. However, the interaction term is positive and significant suggesting that the there is an effect conditional on the existence of the ETF. Such finding highlights the role of the country-level ETFs in the U.S. shock transmission to foreign countries.

[*insert table 2.8 here*]

I explore the role of ETF further by testing if such shock propagation is happening via the arbitrage activity. Similarly to Davies (2019) I use an absolute value of changes in shares outstanding as a proxy of arbitrage activity in the ETF market.

[14] As a consequences of extended sample for this test I use the realised volatility (RV), measured as a squared MSCI return, instead of VIX. Most of the subsequent analysis continues to use VIX.

Creation/redemption results in a change of number of shares in the ETF and therefore, this measure captures the frequency and significance of arbitrage trades.

Table 2.9 shows the result of regressing innovation in correlation on measures of arbitrage and volatility dummy.[15] In this regression I also control for the illiquidity of underlying market $ILLIQ_{MSCI}$ using Amihud (2002) illiquidity measure.

$$\begin{aligned} \Delta\rho_t = \alpha + \beta_1|\Delta SO_t| + \beta_2|\Delta SO_t|D_{US|L,t} + \beta_3 D_{US|L,t} + \\ + \beta_4\Delta\rho_{t-1} + \beta_5 ILLIQ_{MSCI,t} + \varepsilon_t \end{aligned} \tag{2.9}$$

In general, without controlling for volatility, I find only a weak evidence of the relationship between level of arbitrage and innovation. However, once considering full specification of regression and accounting for periods when VIX is high and LVIX is not high (column 2) I show that ETF arbitrage activity affects the innovation in correlation only during periods of high U.S. volatility (while LVIX is not high). In these periods U.S. VIX is at its peak. Taking the argument that VIX is a proxy for a global risk aversion I show that during periods of high uncertainty (e.g., crisis periods) there is a propagation of U.S. shocks to local underlying markets via ETFs. Observing high volatility in the U.S., investors treat order flow in ETF market as the one reflecting fundamentals about local economies. As argued before, this causes a comovement of these markets with the U.S. market. Such finding is in line with the idea that during periods of high volatility it is harder to distinguish if movements in ETF prices reflect liquidity shocks or new information about fundamentals. Consistent with wake-up call hypothesis a significant fall in ETF prices (due to a large increase in VIX) can force investors to reassess the value of local fundamentals.

[*insert table 2.9 here*]

After showing the effect of arbitrage on country correlation with the U.S. over time, I proceed to explore the cross-country differences in such correlation.

[15] I aggregate the daily data into monthly frequency by averaging variables within a month.

Bhattacharya and O'Hara (2018) show that when the underlying assets are hard to trade (e.g., fixed income) price discovery happens in the ETF market. "Hard to trade" situation also arises in a case of international funds with non overlapping trading hours. As mentioned before, Madhavan and Sobczyk (2016) show that a price discovery component of ETF premium is negatively related to variance of transitory liquidity shocks and positively related to the efficiency of arbitrage. My hypothesis is that countries with funds that have a high price discovery component (and therefore, for which market makers are closely following ETFs to price the underlying assets) have a higher correlation with the U.S.. Since all types of U.S. investors trade based on U.S. risks (economic or political) they are affecting the ETF price with U.S. fundamentals. Then the ETF arbitrage works as a transmission mechanism of U.S. risk to foreign countries. However, if the noise in the ETF market (e.g., via retail participation) clouds the price discovery process, I expect the transmission mechanism to be weaker. Overall, I expect the magnitude of correlation to be related to price discovery, noise in the ETF market, and the ease of arbitrage.

In order to test this I follow Broman (2016) to determine the degree of price discovery in ETF. If a demand shock related to fundamentals increases price of ETF above NAV (premium) such faster incorporation of prices reflects a price discovery. Next period, such premium (discount) should translate into a positive (negative) NAV return, as new information gets to the underlying market. In order to compare the extend to which price discovery happens in ETF I regress NAV return on the past ETF premium. Higher β loading represents a stronger adjustment to NAV and higher price discovery in ETF.

$$R_{i,t}^{NAV} = \alpha + \beta_i \left(\frac{P_{i,t-1} - NAV_{i,t-1}}{NAV_{i,t-1}} \right) + \varepsilon \tag{2.10}$$

Table 2.10 shows the average correlations of 3 portfolios formed based on degree of price discovery (panel A) and on a measure of limits to arbitrage (panel B). I also perform an independent double-sort, where I combine price discovery with proxy for limits to arbitrage (panel C). I pre-sort countries into portfolios 1 month before computing correlations. The low group correlation (ρ_t^L) is defined as the average correlation across all counties within this group:

$$\rho_t^L = \frac{1}{N_t^L} \sum_{i=1}^{N_t^L} Corr(\Delta MSCI_i, \Delta S\&P500) \tag{2.11}$$

where N_t^L is the number of countries in the group L in month $t-1$. The correlation of returns is computed using 36-month rolling window (I test the alternative measures in section 2.5). The correlation for medium and high groups is defined in the same manner.

As expected, the first sort based on β from equation 2.10 shows a significant increase in correlation for funds with a high price discovery. The difference in average correlation across low and high groups is 0.17. Panel B shows the sort based on a proxy for limits to arbitrage. I use Amihud's illiquidity ratio (Amihud, 2002) of underlying markets. As can be seen from this sort, increase in illiquidity measure leads to a relatively lower correlation with U.S. on average (by 0.11). Lower liquidity of the underlying market limits the profitability of the arbitrage (due to higher price impact) and, as such, U.S. shocks are less likely to get propagated to local economies. The double sorting confirms my previous findings: the average difference between a country with low price discovery and high limits to arbitrage in comparison to the opposite one (low limits and high price discovery) is around 0.16. I find that Australia, France, Germany, Spain and the UK are the countries that have the highest chance of being in the low illiquidity - high price discovery group. In contrast, Brazil, China, Hong Kong, India and Taiwan appear in low illiquidity - low price discovery group more often than any other country.

Berger, Pukthuanthong, and Yang (2011); Pukthuanthong and Roll (2009) argue that the correlation is an *imperfect* measure of market integration and propose a measure that is based on the first principal component of equity returns. To this end, table 2.10 also offers average adjusted R-squares that are estimated based on 36-month rolling contemporaneous regressions of MSCI index returns on the first principal component (PC1) of all available stock returns in my sample. The principal components are estimated based on a 36-month rolling window. I find that the average adjusted R-squares follow the same direction with the average rolling correlations implying higher (lower) levels of market integration as the NAV sensitivity to premium (stock illiquidity) increases.

Pan and Zeng (2019) show that for corporate bond ETF market due to existence of liquidity mismatch between fund and the underlying index the inventory management effect[16] may be dominant making APs reluctant to close the price deviations. I extend this argument to ETF market for international equities. While there may not be a significant mismatch for highly liquid developed market ETFs, my sample also includes small emerging market economies, where the underlying index is much more illiquid than the corresponding ETF fund. In contrast to Pan and Zeng (2019), as I have a broad range of different countries in my sample I consider ETF liquidity to have enough variability to be included in my measure. I compute the liquidity mismatch as a percentage difference in Amihud's illiquidity measures for ETF and local index.

$$Mismatch = \frac{ILLIQ_{ETF} - ILLIQ_{index}}{ILLIQ_{index}} \tag{2.12}$$

I use this mismatch measure for a limits to arbitrage proxy (in my definition lower the value of mismatch - stronger the limits). In contrast to previous measures, this variable combines both ETF and local market. Figure B.1 plots the mismatch together with an average GNI per capita (as a proxy to identify developed markets) obtained from World Bank database.[17] Countries with higher GNI per capita tend to have lower liquidity mismatches (more positive). Panel D of table 2.10 shows the result of double-sorting the funds by price discovery and this version of limits to arbitrage. As can be seen from the sort, countries where the difference in liquidity is

[16]In this market APs maintain inventory of illiquid corporate bonds.

[17]MSCI uses GNI per capita as one of many criteria for developed market classification.

the largest (the most negative mismatch), have on average lower correlation with U.S. market. The difference between extreme portfolios (high price discovery and high mismatch vs. low price discovery and low mismatch) is around 0.16. This is consistent with previous evidence of limits to arbitrage preventing the propagation of U.S. market shocks to local economies.

[*insert table 2.10 here*]

Finally, I test how the impediments to price discovery, say lack of price efficiency and/or noise trading activity in the ETF markets affects the correlation mechanism via ETF arbitrage. I form monthly portfolios based on one of the three proxies that affect price discovery. My first proxy is the variance ratio (VR), specifically $|VR - 1|$, where VR is the ratio of 15-sec returns over three times of 5-sec return variances. If the prices follow random walk, I expect this measure to be equal to zero. Deviations from zero indicate lack of price efficiency in the ETF market (Ben-David et al., 2018; O'Hara and Ye, 2011). My second proxy is based on the ratio of dollar volume of *smallOI* and of TOI. Small trades arguably capture either uninformed retail or high frequency trading (HFT) activity. While the former is likely to introduce noise in the market, the role of the latter on price efficiency is less clear with mixed evidence (Brogaard, Hendershott, and Riordan, 2014; Zhang, 2010). My third proxy is the ratio of dollar volume of *smallOI* over the (informed) $retailOI$. If the denominator is a measure of informed retail activity as argued in Boehmer et al. (2017), this proxy can be used to obtain the relative noise (over signal) due to retail participation. Admittedly, none of these proxies are perfect measures of noise. In each panel of table 2.11, I report both the single sorts and the double sort taking into account Amihud's illiquidity ratio (Amihud, 2002) of the underlying markets (*ILLIQ*).

The single-sort based on variance ratios in Panel A of table 2.11 shows that as the price efficiency of the ETF market declines, the correlation with the U.S. market declines as well. This is in line with the mechanism I described above: since the price discovery in the ETF market is more difficult, the correlation due to arbitrage mechanism is lower. However, this effect disappears in the double-sorts. This is not surprising, since the high-frequency variance ratios and illiquidity are inherently linked. The sorting results in Panel B and Panel C confirm the single-sort result that noisy ETF markets make it difficult for the arbitrage mechanism to induce higher correlations between U.S. and foreign markets, however, this time regardless of the illiquidity of the underlying markets. The evidence is particularly strong in Panel C. Overall, the evidence from table 2.11 lends support to the idea that noise in the ETF market is an impediment for the arbitrageurs, but at the same time brings good news for uninformed retail participants in the context of international diversification via ETFs.

[*insert table 2.11 here*]

2.5 Robustness and other Specification Tests

In this section I perform numerous robustness tests to ensure that results demonstrated in previous sections are not sensitive to my choice of methodologies. I start with the analysis of my cross-sectional sample choice, then I control for alternative types of risks that may affect the result, I follow by investigating how different volatility conditions affect my results and conclude by computing alternative correlation measures.

2.5.1 Sub-sample analysis

As shown in Figure 2.3, my choice of 41 counties has a wide geographical dispersion. I test if my results are robust to my sample selection. It may be possible that the dominant effect of VIX comes from countries that are more integrated with the U.S.. I split my sample by the level of economic development based on MSCI classification:

into developed and emerging markets. My sample is not dominated by developed countries, as almost half of the sample (20 countries) consists of emerging countries. Tables B.5 and B.6 in the appendix show the results of such split.

Developed markets results show that ΔVIX is significant with large negative coefficients for every type of investor. The reaction to changes in local VIX is much smaller or insignificant. This is consistent with my previous evidence despite the presence of major financial centres (e.g., UK, Germany, Switzerland etc.) in the region.

Emerging markets results also suggest that every type of investor mainly reacts to U.S. VIX. The loadings on ΔVIX are relatively smaller. I see the evidence of reaction to local VIX on average, however, it is mainly due to small trades. Overall, I show that my results are not driven by the choice of my sample, and that U.S. VIX remains the key variable to which U.S. investors react when trading country ETFs.

2.5.2 Foreign Exchange risk

None of ETFs in my sample include derivatives (usually futures) to hedge the foreign exchange risk. The NAV of the fund is the sum of its holdings expressed in USD. Therefore, NAV and price of ETF is subject to FX fluctuations. I obtain daily spot exchange rates for each country versus USD from Datastream. I test my results by including spot return of local currency (Δexr) and absolute value of such return ($|\Delta exr|$) to account for latent FX volatility (Menkhoff, Sarno, Schmeling, and Schrimpf, 2012).

Table 2.12 shows the results of such regression. Foreign exchange return and volatility variables are significant and negative for almost every type of order imbalance. When local currency depreciates relative to USD (positive Δexr) the NAV of the fund becomes smaller (as the value of securities owned by the fund expressed in USD becomes smaller). The negative loading on spot return reflects the sale of ETFs by investors, as the price temporarily becomes higher than NAV. The negative reaction to FX volatility is consistent with investors' fear to hold ETFs, as during volatile FX environment it is harder to price the basket of underlying securities. Despite significant reaction of investors to FX risk proxies, the response to changes in VIX and LVIX remains the same, suggesting that my results are not driven by fluctuations in exchange rates.

2.5.3 Volatility regimes and Recession Period

I also study the effect of volatility in the U.S. market on my results. As argued by Drechsler, Moreira, and Savov (2018) liquidity providers are exposed to volatility risk. When volatility is high one can expect the sensitivity of order imbalance to changes in U.S. VIX to be high. It is possible that the dominance of U.S. VIX is mainly due to a period of high volatility present in my sample (e.g., financial crisis). I split my sample into 3 periods: low, medium and high VIX. The split is based on terciles of historic VIX level from 1990-2018. Table 2.13 shows the results of such split. As can be seen, U.S. VIX remains significant in every period for almost every type of investor. The coefficient is insignificant only for *smallOI* in the low period. Increase in U.S. uncertainty when the general level of volatility is low may not be a strong signal for investors with small trades. Overall, I can see that the high exposure to changes in U.S. VIX remains in any volatility regime.

The sell-off of international ETFs could be particularly strong during recession periods in local economies. I obtain the recession indicators from OECD database for countries in my sample. I create a dummy variable D_R that takes the value of 1 if there is a recession period during a month in a local country and 0 otherwise. Table B.7 shows the results of including the recession periods in my regression of order imbalance on U.S. and local VIX. The result shows that all types of investors tend to sell international country-level ETFs more when there is a recession in a local country. For most of the investor types $\Delta LVIX$ becomes insignificant. This is consistent with my previous evidence that investors only react to large negative news (that are more likely to happen during recession). In contrast, the loadings on changes in U.S. VIX remain significant, suggesting that my result are not driven by recession periods.

2.5.4 Correlation: Alternative Explanations and Different Measures

One of the possible explanations of a high correlation of a country with the U.S. can be the importance of U.S. as a country's trading partner (Chen and Zhang, 1997). In order to control for this channel, I obtain monthly exports and imports between U.S. and my sample countries from U.S. census website. I scale the result by total amount of exports and imports of those countries obtained from OECD database. When computing partial correlations I use 36-month rolling window and control for export and import ratios:

$$\rho_t^L = \frac{1}{N_t^L} \sum_{i=1}^{N_t^L} partialCorr(\Delta MSCI_i, \Delta S\&P500 \quad | \frac{m_{US,t}}{m_{Total,t}}, \frac{x_{US,t}}{x_{Total,t}}) \tag{2.13}$$

where $m_{US,t}$ ($x_{US,t}$) is the amount of imports (exports) of a country from (to) U.S. in month t and $m_{Total,t}$ ($x_{Total,t}$) is the total amount of imports (exports) with all of its trading partners. The correlation for medium and high groups is defined in the same manner. Table B.9 in the appendix shows the sort using partial correlation measures. The result is robust to imports and exports controls.

The alternative explanation of high correlation between countries is related to business cycles. As discussed before, the correlation tends to be higher during recession periods. I obtain monthly industrial production index (a proxy for business cycles in monthly frequency) for most of the countries in my sample from the OECD and Global Financial Data databases. I control for annual changes in industrial production when computing partial correlation, which is defined as follows:

$$\rho_t^L = \frac{1}{N_t^L} \sum_{i=1}^{N_t^L} partialCorr(\Delta MSCI_i, \Delta S\&P500 \quad |\Delta IP_{i,t}, \Delta IP_{US,t}) \tag{2.14}$$

where $\Delta IP_{i,t}$ ($\Delta IP_{US,t}$) is the change in the industrial production for country i

(U.S.) from month $t-12$ to t. Table B.10 shows that my results are unchanged, and therefore robust to controlling for business cycles.

Most of the countries that appear in the low illiquidity- high price discovery group are developed financial centres. Therefore it is important to check if the result is driven by the level of country development. In table B.11 in the appendix I repeat the sorting conditional on countries being in G10 or in the MSCI Developed Country universe. I find that the price discovery channel remains significant even within the developed countries. In addition, I perform the sorting using partial correlations controlling for the level of country's financial development (FD). FD is computed as a ratio of total stock market capitalization (obtained from Global Financial Data) and GDP (obtained from OECD database). The results are robust to such control.

I also study the effect of my choice of correlation measure. In the main results I compute correlation using 36-month rolling window. The benefit of such method is that it does not rely on an assumption about the data generating process and is simple to compute. However, there is no clear way to choose the appropriate length of the rolling window. I first test the validity of my cross-sectional results using a longer 100-month period. Table B.12 in the appendix shows that the results of such sort produce identical outcome as before.

In order to overcome the need to choose the length of the rolling window (longer length may result in a smoother correlation estimates), I employ an alternative measure of correlation - Dynamic Conditional Correlation (DCC) of Engle (2002). The assumption is that returns conditional on prior available information is normally distributed with mean 0 and time-varying covariance matrix H_t: $r_t|\mathcal{F}_{t-1} \sim (0, H_t)$. Then covariance matrix can be represented as:

$$H_t = D_t R_t D_t \tag{2.15}$$

D_t is the square root of diagonal matrix of H_t and R_t is the time-varying correlation matrix. I model the volatility of returns for each country using GARCH(1,1) process.

Matrix R_t can be further decomposed into:

$$R_t = diag(Q_t)^{-1/2} Q_t diag(Q_t)^{-1/2} \tag{2.16}$$

Then auxiliary variable Q_t can in turn be represented using GARCH(1,1) process as:

$$Q_{ij,t} = \bar{\rho}_{ij} + \alpha(\varepsilon_{i,t-1}\varepsilon_{j,t-1} - \bar{\rho}_{ij}) + \beta(Q_{ij,t-1} - \bar{\rho}_{ij}) \tag{2.17}$$

In this equation, $\varepsilon_{i,t-1} = D_{t-1}^{-1} r_{t-1}$ and $\bar{\rho} = \mathbb{E}[\varepsilon_{i,t}\varepsilon_{j,t}]$.

Table B.13 in the appendix shows the results using DCC correlation. Both single and double sorts produce very similar results to the ones using 36 and 100-month correlations. The outcomes of these robustness tests suggest that my results are not sensitive to the modelling choice of correlation.

2.5.5 VIX Decomposition

Here, I examine which part of the variation of the VIX drives its predictive ability for ETF order imbalances. To this end, I extract the *economic uncertainty* and *risk aversion* components of VIX following Bekaert and Hoerova (2014); Bekaert, Hoerova, and Duca (2013).[18] Specifically, the VIX component that serves as a proxy of *economic uncertainty* is the forecast of the following month stock market variance (e.g., 22 trading days) that is estimated based on a model that includes as independent variables a squared value of VIX that is annualized and expressed in percentage terms as well as the continuous component of the daily quadratic variation of stock returns (e.g., $C_t = RV_t - J_t$, where RV is the daily realized variance of the stock market return that is computed as the sums of squared five-minute returns as well as the squared close-to-open return and J_t represents the jump component) over the previous day, week and month. Thus, I denote the conditional (physical) variance by VIX^{CV}. The *risk aversion* (e.g., VIX^{RA}) component of VIX reflects the variance risk premium which is measured as the difference between the implied and

[18] I would like to thank the authors for making the data available on their webpage.

conditional variance (e.g., $VIX_t^2 - E_t[RV_{t+1}^{(22)}]$, where $RV_{t+1}^{(22)}$ is the realized variance of the S&P 500 over the next month).

I run a predictive panel regression with random effects for all countries in my sample. My model also includes lagged percentage changes of the components of VIX, the local VIX and a number of controls: a dummy variable (L) that takes the value of 1 if a country has a common language with the U.S. and 0 otherwise; and a constant that represents the geographical distance of the country from the U.S. (G). I also consider lagged order imbalance, lagged interest rates and lagged specifications of components of VIX. My model includes both components of VIX as well as nested specifications of the model below:

$$
\begin{aligned}
OI_{i,t} = \alpha + \sum_{k=1}^{5} \beta_k^{CV} \Delta VIX_{t-k}^{CV} + \sum_{k=1}^{5} \beta_k^{RA} \Delta VIX_{t-k}^{RA} + \sum_{k=1}^{2} \gamma_k \Delta LVIX_{i,t-k}^{o} + \\
+ \delta_1 \Delta IR_{i,t-1} + \delta_2 L_i + \delta_3 G_i + \delta_4 OI_{i,t-1} + \varepsilon
\end{aligned}
\tag{2.18}
$$

Table B.15 of the appendix shows the panel results of such regression. I consider different order imbalance measures as independent variables (e.g., TOI, smallOI, retailOI and largeOI). I find that the economic uncertainty component of VIX is a strong predictor of total ETF order flow over and above the risk aversion component. However, I find that both components of VIX render a similar contribution to small and retail order imbalances while economic uncertainty drives the variation of order imbalances of institutional investors. Overall, the results suggest that traders mainly react to changes in U.S. economic uncertainty rather than the risk aversion part of the VIX.

2.5.6 Central Bank ETF Purchases: The Case of Japan

There is a policy discussion whether the central banks should extend their quantitative easing (QE) programmes to stock markets. While previous Fed Chair Yellen hinted at the possibility in a testimony before the House Financial Services Committee (Reuters, 2016)[19], the Bank of Japan (BOJ) has been purchasing stocks through its ETF Purchase programme since 2010, and has gradually increased the intensity of ETF purchases in 2014 (3 trillion yen per year) and 2016 (6 trillion yen per year)(Barbon and Gianinazzi, 2019; Charoenwong, Morck, and Wiwattanakantang, 2019). These interventions in the (TOPIX and Nikkei 225) ETF market aim at reducing the equity risk premia and improving underlying equity valuations, however, they happen to be distortionary (Barbon and Gianinazzi, 2019) and relatively predictable, that is, BOJ purchases are triggered when indices go down due to market uncertainty (FT, 2018)[20].

In figure B.2 I show the actual BOJ ETF purchases (in billion yen) and the intraday variance ratios for both EWJ and SPY ETFs since the beginning of the BOJ ETF purchase programme, over the period 2010-2018. These ETFs that track MSCI Japan and S&P 500, respectively, and both are traded on NYSE Arca platform. My proxy for intraday price distortions for the ETF market suggests that the average level of distortions increases as BOJ moves to a new regime of ETF purchases with larger amounts, while I do not observe any effect on SPY trading. In fact, as figure B.3 indicates, the price distortions are significantly larger on the days of BOJ purchases which coincide with an increase in uncertainty measured by U.S. VIX.

The impact of BOJ's actions on Japanese equity market is asymmetric, as the central bank acts only in one direction, i.e., buys ETFs. As a result, BOJ is likely to create an upward bias on the valuation of the underlying stocks (Charoenwong et al., 2019), which translates into noisy overvalued NAV of EWJ. Realising the probability of BOJ's intervention, market participants in the U.S. are then likely to sell EWJ pushing the price downwards until the fundamental level is reached. Figure B.4 shows that on average returns on the EWJ ETF are much more negative on BOJ ETF purchase days than on non-intervention days. The selling pressure

[19]https://www.reuters.com/article/us-usa-fed-yellen-purchases-idUSKCN11Z2WI.
[20]https://www.ft.com/content/8f472648-a783-11e8-8ecf-a7ae1beff35b.

results in a positive intra-day autocorrelation captured by the variance ratio. At the same time, figure B.4 shows that SPY returns tend to be negative on the BOJ purchase days, presumably through the increase in VIX. Finally, I demonstrate that negative returns of SPY and EWJ on the interventions days results in a higher (the difference is statistically significant at 1% level) intraday correlation of these two ETF products traded on the U.S. market. As a result, the impact of BOJ's purchase program is likely to have adverse effect on the ability of U.S. investors to diversify globally by purchasing Japanese country-level ETFs.

2.6 Conclusion

Overall, I investigate how different groups of country ETF traders make investment decisions. I show that order imbalance of country ETF trades mainly reflects changes in U.S. implied volatility index rather than local VIX. Such result is robust to different volatility regimes and a sub-sample analysis. Asymmetric response analysis shows that investors react only to negative foreign news, as measured by positive changes in LVIX. I find that when reducing positions in international ETFs, investors switch to a safe asset such as cash ETFs, reflecting a risk based explanation of ETFs sell-off.

I use these results to investigate the mechanism of transmission of U.S. shocks to foreign countries that results in high cross-country correlation. I argue that such shocks are propagated to different countries via ETF arbitrage mechanism. Consistent with this argument my time-series analysis shows that both the introduction of the ETF market, and significant arbitrage activity in the ETF market at times of high U.S. volatility results in a positive innovation in country's stock market correlation with U.S. market. My finding is in line with a wake-up call hypothesis of contagion.

Finally, I investigate the cross-sectional differences in countries' correlations with the U.S. market. APs engage in arbitrage activity to correct the deviations between ETF price and the NAV. If such deviation is caused due to transitory liquidity shock, the adjustment to ETF price and NAV should be reverted. In contrast, I argue, that if such deviation is a result of a faster incorporation of fundamentals in ETF price, arbitrage should lead to higher correlation of a country with the U.S. equity market. In support of this hypothesis, I find that countries with higher price discovery and lower limits to arbitrage have a higher correlation with the U.S. market. By the same token, countries with lower price discovery and higher limits to arbitrage due to noise in the ETF market have a lower correlation with the U.S. market. The latter finding implies that an increase in (uninformed) retail participation is likely to have a positive effect on international diversification via country ETFs.

I also consider the response to local and global political uncertainty and its association with economic uncertainty (e.g., the VIX). I find that investors tend to respond to past movements of U.S. political uncertainty in a different fashion than to economic uncertainty. Specifically, investors react to such signals by exiting the U.S. stock market and moving their capital to international stock markets. I do not find any evidence that investors react to my local political uncertainty proxy. Overall, I find that the effect of U.S. VIX on ETF order flow is stronger compared to local and global political uncertainty.

My work is important for international investors seeking to diversify their U.S. exposure by investing in international ETFs. My analysis suggests that even emerging countries with low integration are still significantly affected by the U.S. stock market. While previous research focuses on the role of global banks and the U.S. as the central economy on cross-country correlation, the novelty of my study is that I discover a new channel of country connectedness that is via ETF arbitrage.

Figures and Tables

Figure 2.1: Correlation Mechanism

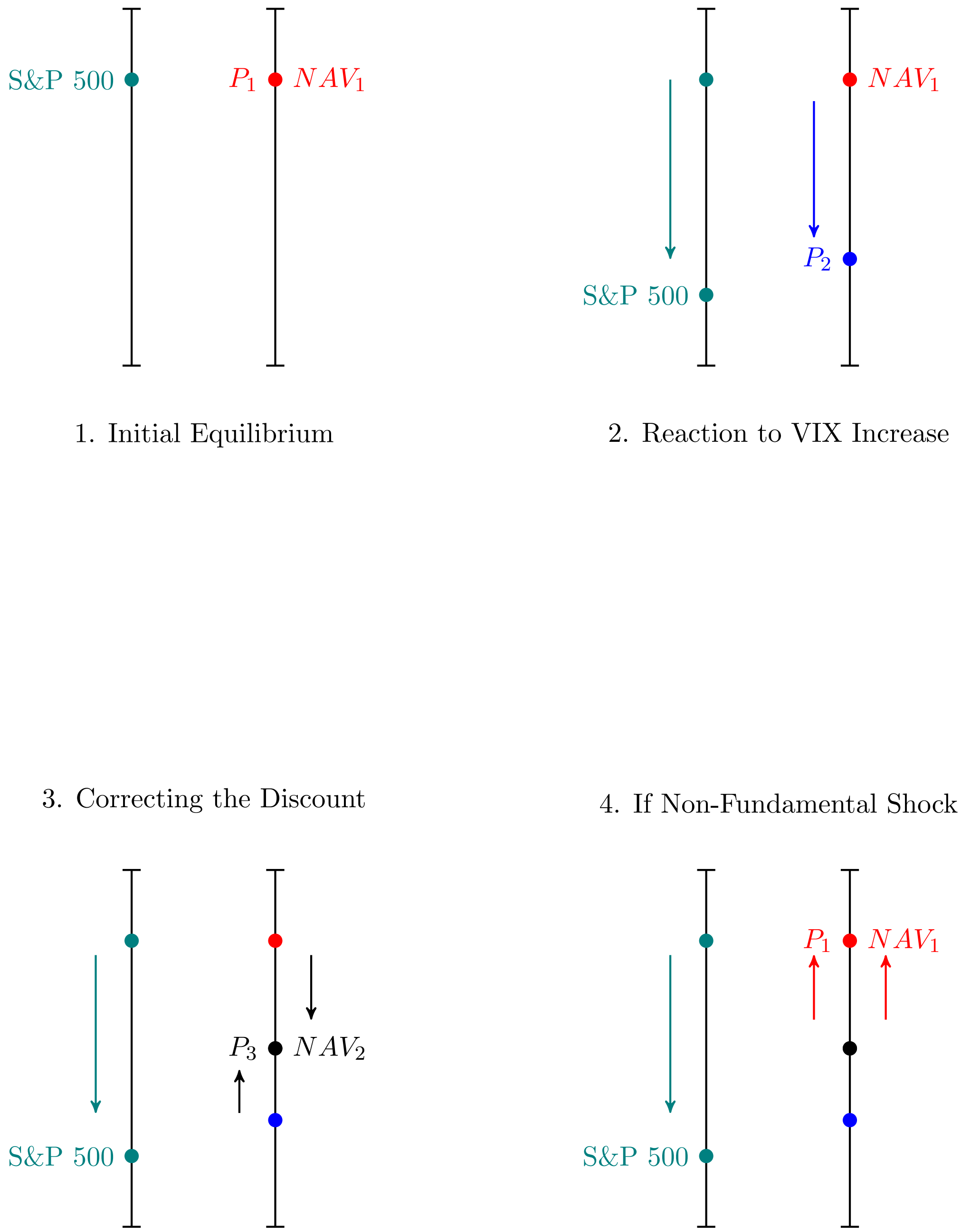


Figure 2.2: Correlation between U.S. Market and the Rest of the World

Solid line shows the rolling 36-month correlation between S&P 500 and MSCI EAFE total returns ($\rho = Corr(\Delta S\&P500, \Delta MSCI_{EAFE})$). MSCI EAFE represents the portfolio of more than 900 stocks from Europe, Australia, Asia, and the Far East. Index returns are measured at a monthly frequency. Data sample is 2001-2018. Dashed line shows the 36-month moving average of CBOE VIX level at a monthly frequency. Shaded area is official NBER recession period of December 2007-June 2009.

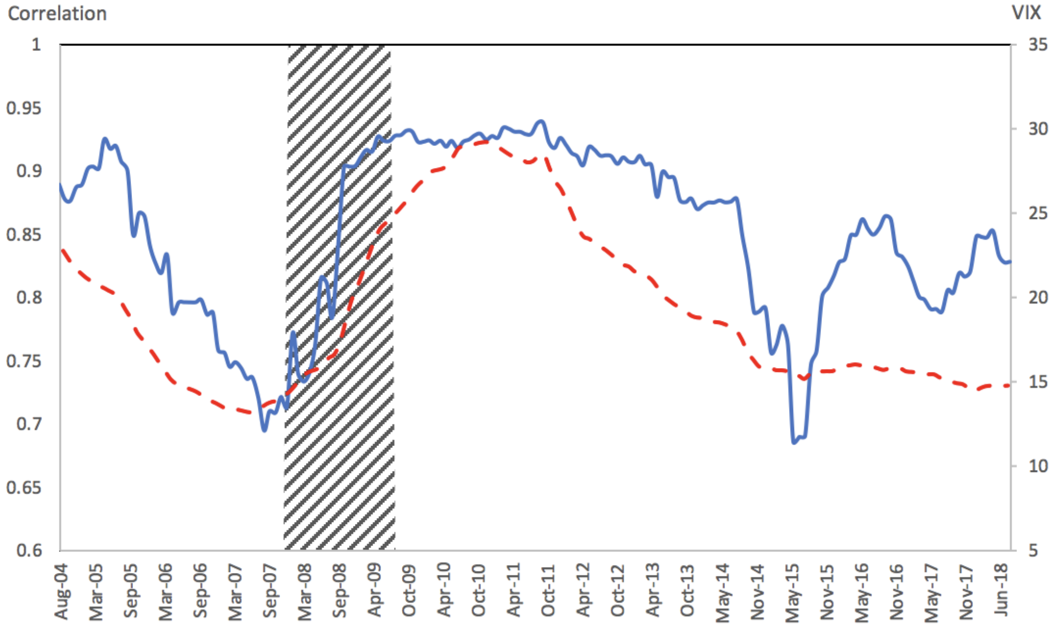


Figure 2.3: Correlation between the U.S. and Other Countries

This map shows the 36-month correlation between return of S&P 500 and returns of MSCI indices ($\rho_i = Corr(\Delta S\&P500, \Delta MSCI_i)$) of 41 countries used in the sample (see table B.8 for the full list of ETFs and corresponding indices). The correlation varies from light blue (the lowest) to dark blue (the largest) .

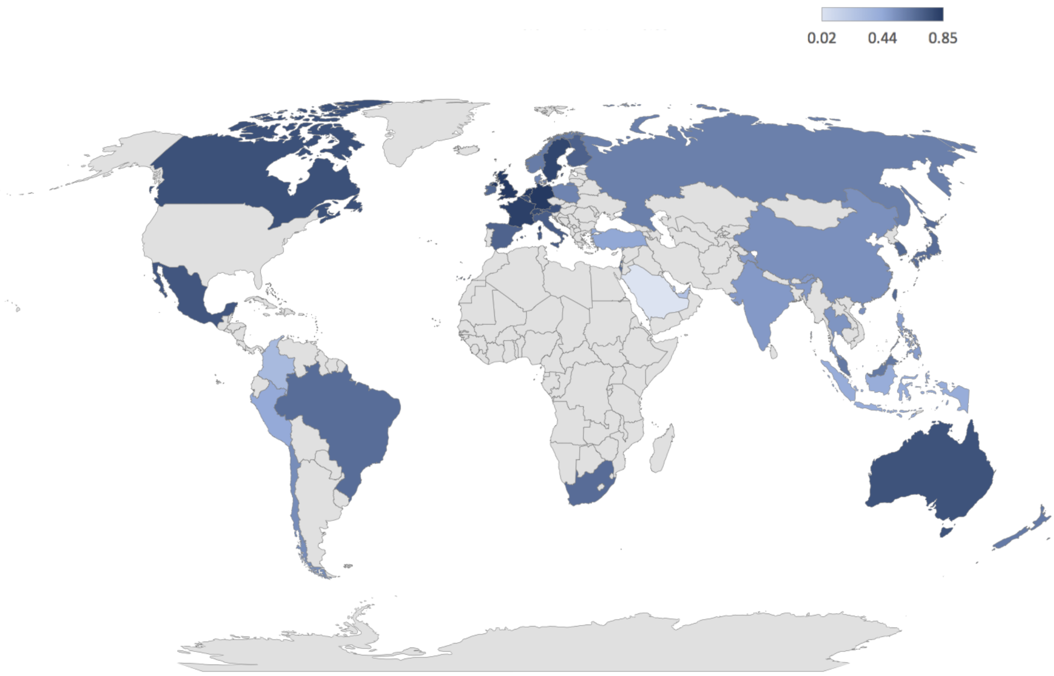


Table 2.1: Summary Statistics of Order Imbalances

Summary statistics of daily total, small, retail and large order imbalances. *retailOI* is based on Boehmer et al. (2017) and *largeOI* is based on Barber et al. (2009). *smallOI* is the residual from equation 2.2 and expressed in 10^{-16}. All order imbalances are calculated as defined in equation 2.1. Mean and standard deviation is based on the sample period from 2006 or the first trading day of ETF (whichever is later) until the end of June 2018.

	TOI		*smallOI*		*retailOI*		*largeOI*	
Country	Mean	Std	Mean	Std	Mean	Std	Mean	Std
AUS	0.02	17.33%	0.06	14.05%	−0.01	32.98%	0.02	27.22%
AUT	−0.02	42.38%	−0.57	35.78%	−0.04	46.84%	−0.02	60.27%
BEL	0.01	43.58%	0.68	38.61%	−0.05	51.68%	0.02	61.97%
BRA	0.00	10.08%	0.02	24.61%	0.01	24.98%	0.00	16.75%
CAN	0.02	20.51%	0.06	14.67%	0.04	30.76%	0.02	30.12%
CHL	0.02	30.61%	−0.41	39.84%	0.01	41.24%	0.02	41.12%
CHN	0.06	35.80%	−0.01	30.21%	0.10	50.33%	0.06	45.20%
COL	0.16	66.25%	0.01	64.58%	0.16	81.51%	0.22	91.22%
DNK	0.07	52.70%	0.00	50.83%	0.07	61.32%	0.04	70.49%
FIN	0.08	57.64%	0.86	57.87%	0.16	73.26%	0.06	82.82%
FRA	0.01	32.06%	0.08	26.07%	0.01	44.91%	0.01	39.97%
DEU	0.02	24.30%	0.20	15.57%	0.02	36.47%	0.02	33.36%
HKG	0.01	18.02%	−0.11	13.85%	−0.04	35.93%	0.01	26.26%
IND	0.08	38.36%	0.22	34.98%	0.08	47.70%	0.07	46.33%
IDN	0.04	27.53%	0.14	25.95%	0.01	41.87%	0.03	40.05%
IRL	0.03	51.89%	−0.72	50.05%	0.06	59.85%	−0.02	73.03%
ISR	−0.01	45.21%	−0.10	43.94%	−0.02	53.32%	−0.02	63.44%
ITA	0.02	32.12%	−0.19	28.19%	0.03	46.89%	0.02	41.22%
JPN	0.01	13.92%	0.03	14.46%	−0.03	33.01%	0.01	16.64%
MYS	0.00	20.11%	0.08	17.54%	−0.03	40.22%	−0.01	32.02%
MEX	0.01	12.85%	0.10	26.64%	0.01	33.15%	0.00	22.75%
NLD	0.00	39.58%	−0.16	34.33%	−0.02	46.76%	0.00	51.98%
NZL	0.02	40.06%	0.74	37.10%	0.00	46.62%	0.03	59.52%
NOR	0.09	54.95%	0.06	55.27%	0.08	75.97%	0.10	84.22%
PER	−0.03	38.45%	0.03	33.69%	−0.01	45.32%	−0.05	54.98%
PHL	0.02	30.44%	0.34	27.24%	0.02	45.43%	0.01	46.30%
POL	0.03	33.46%	−0.11	27.33%	0.02	46.89%	0.03	51.55%
QAT	0.07	71.15%	0.90	69.15%	0.00	80.32%	0.10	90.79%
RUS	0.01	34.43%	0.01	26.59%	0.03	47.64%	0.00	50.82%
SAU	0.12	59.70%	0.64	59.54%	0.14	73.77%	0.13	77.67%
SGP	0.01	19.53%	0.15	17.31%	−0.04	32.82%	0.01	30.09%
ZAF	0.00	23.14%	0.09	36.20%	−0.03	39.31%	0.00	30.91%
KOR	0.00	12.93%	−0.01	25.34%	−0.01	34.40%	−0.01	21.26%
ESP	0.01	29.84%	−0.07	23.52%	0.01	37.94%	0.02	37.77%
SWE	0.01	32.01%	0.06	27.35%	0.02	35.05%	0.01	41.47%
CHE	0.03	28.30%	0.05	23.76%	0.05	32.61%	0.02	37.58%
TWN	0.00	13.13%	−0.01	12.88%	−0.03	37.43%	0.00	19.16%
THA	0.02	29.96%	0.30	37.35%	0.01	43.57%	0.02	40.99%
TUR	0.00	26.88%	0.29	34.03%	−0.01	35.07%	0.00	38.68%
ARE	0.01	63.45%	0.13	63.12%	0.02	80.17%	0.04	89.04%
GBR	0.03	25.40%	−0.18	20.86%	0.07	34.60%	0.04	34.59%

Table 2.2: Correlation of Order Imbalances

Per country correlations of daily retail, small and large order imbalances based on Barber et al. (2009) and retail order imbalances based on Boehmer et al. (2017). Construction of variables is defined in sections 2.3.2 and 2.3.2. Sample period begins from 2006 or the first trading day of ETF (whichever is later) until the end of June 2018.

Country	$\rho(smallOI, retailOI)$	p-val	$\rho(smallOI, largeOI)$	p-val	$\rho(retailOI, largeOI)$	p-val
AUS	0.199	< 0.01	0.280	< 0.01	0.243	< 0.01
AUT	0.277	< 0.01	0.361	< 0.01	0.296	< 0.01
BEL	0.306	< 0.01	0.342	< 0.01	0.241	< 0.01
BRA	0.053	0.03	0.005	0.84	0.027	0.13
CAN	0.241	< 0.01	0.321	< 0.01	0.332	< 0.01
CHL	0.088	< 0.01	0.245	< 0.01	0.216	< 0.01
CHN	0.166	< 0.01	0.199	< 0.01	0.168	< 0.01
COL	0.196	< 0.01	0.330	< 0.01	0.099	0.10
DNK	0.232	< 0.01	0.237	< 0.01	0.174	< 0.01
FIN	0.198	< 0.01	0.280	< 0.01	0.117	< 0.01
FRA	0.202	< 0.01	0.282	< 0.01	0.176	< 0.01
DEU	0.192	< 0.01	0.325	< 0.01	0.250	< 0.01
HKG	0.119	< 0.01	0.232	< 0.01	0.205	< 0.01
IND	0.154	< 0.01	0.324	< 0.01	0.171	< 0.01
IDN	0.175	< 0.01	0.314	< 0.01	0.238	< 0.01
IRL	0.278	< 0.01	0.274	< 0.01	0.205	< 0.01
ISR	0.193	< 0.01	0.279	< 0.01	0.234	< 0.01
ITA	0.195	< 0.01	0.329	< 0.01	0.203	< 0.01
JPN	0.214	< 0.01	0.278	< 0.01	0.224	< 0.01
MYS	0.224	< 0.01	0.262	< 0.01	0.269	< 0.01
MEX	0.007	0.77	0.031	0.17	0.157	< 0.01
NLD	0.218	< 0.01	0.336	< 0.01	0.237	< 0.01
NZL	0.188	< 0.01	0.307	< 0.01	0.196	< 0.01
NOR	0.213	< 0.01	0.274	< 0.01	0.180	< 0.01
PER	0.284	< 0.01	0.392	< 0.01	0.258	< 0.01
PHL	0.222	< 0.01	0.329	< 0.01	0.239	< 0.01
POL	0.195	< 0.01	0.298	< 0.01	0.196	< 0.01
QAT	0.255	< 0.01	0.394	< 0.01	0.126	0.02
RUS	0.178	< 0.01	0.250	< 0.01	0.219	< 0.01
SAU	0.293	< 0.01	0.199	< 0.01	0.103	0.08
SGP	0.303	< 0.01	0.350	< 0.01	0.302	< 0.01
ZAF	0.040	0.15	0.108	< 0.01	0.205	< 0.01
KOR	0.032	0.15	0.086	< 0.01	0.179	< 0.01
ESP	0.193	< 0.01	0.320	< 0.01	0.274	< 0.01
SWE	0.233	< 0.01	0.365	< 0.01	0.275	< 0.01
CHE	0.206	< 0.01	0.284	< 0.01	0.224	< 0.01
TWN	0.124	< 0.01	0.278	< 0.01	0.189	< 0.01
THA	0.124	< 0.01	0.233	< 0.01	0.207	< 0.01
TUR	0.139	< 0.01	0.166	< 0.01	0.220	< 0.01
ARE	0.169	< 0.01	0.323	< 0.01	0.018	0.73
GBR	0.168	< 0.01	0.318	< 0.01	0.191	< 0.01

Table 2.3: Individual Country Regressions

Predictive regression of order imbalances for 41 country-level MSCI based iShares ETFs on percentage change in volatility index (ΔVIX), orthogonalised percentage change in local volatility index ($\Delta LVIX^o$), change in local interest rates (IR), lagged order imbalance (OI_{t-1}), as well as lags of ΔVIX and $\Delta LVIX^o$. Specifically, I estimate the model below:

$$OI_{i,t} = \alpha + \sum_{k=1}^{5} \beta_k \Delta VIX_{t-k} + \sum_{k=1}^{2} \gamma_k \Delta LVIX^o_{i,t-k} + \delta_1 \Delta IR_{i,t-1} + \delta_2 OI_{i,t-1} + \varepsilon \tag{2.19}$$

The number of lags in the model is determined using Akaike Information Criterion and Bayesian information criterion jointly. Coefficients and t-statistics (based on Newey and West (1987) standard errors) are presented. The lag for predictive regression is 1 day. Data is at a daily frequency and covers the period from 2006 or the first trading day of ETF (whichever is later) until the end of June 2018. ***,**,* denote significance at the 1%, 5% and 10% level.

Country OI	Constant	ΔVIX_{t-1}	$\Delta LVIX^o_{t-1}$	ΔIR_{t-1}	OI_{t-1}	ΔVIX_{t-2}	$\Delta LVIX^o_{t-2}$	ΔVIX_{t-3}	ΔVIX_{t-4}	ΔVIX_{t-5}	Obs.	R^2
AUS	0.001	−0.114***	−0.015	−0.092	0.075***	−0.027	−0.071	−0.012	−0.056	0.032	2,640	0.009
	(0.25)	(−3.46)	(−0.30)	(−0.90)	(2.71)	(−0.65)	(−1.56)	(−0.32)	(−1.62)	(0.95)		
AUT	−0.015	−0.302***	0.054	−0.039	0.183***	−0.163**	0.032	−0.072	−0.104	−0.135	3,140	0.038
	(−1.49)	(−3.29)	(0.40)	(−0.63)	(8.64)	(−2.01)	(0.23)	(−0.78)	(−1.16)	(−1.57)		
BEL	0.030*	−0.417**	0.267	−0.146	0.353***	−0.107	−0.258	0.039	−0.072	−0.138	1,220	0.129
	(1.78)	(−2.48)	(1.14)	(−0.60)	(10.30)	(−0.58)	(−1.19)	(0.22)	(−0.42)	(−0.80)		
BRA	0.002	0.022	–	−0.074	0.067***	0.015	–	0.004	0.013	0.004	3,140	0.003
	(1.08)	(1.14)	–	(−0.84)	(2.76)	(0.72)	–	(0.17)	(0.62)	(0.21)		
CAN	0.004	−0.090**	−0.087**	0.029	0.177***	−0.013	−0.046	−0.079*	−0.088**	−0.028	2,199	0.036
	(0.71)	(−2.18)	(−2.23)	(0.71)	(7.02)	(−0.32)	(−1.12)	(−1.85)	(−2.36)	(−0.72)		
CHL	0.016**	−0.066	–	−0.054	0.229***	0.034	–	−0.057	0.065	−0.086	2,672	0.052
	(2.13)	(−0.82)	–	(−1.54)	(9.63)	(0.44)	–	(−0.85)	(0.91)	(−1.17)		
CHN	0.051***	−0.196**	−0.085	−0.003	0.208***	0.130	−0.165	−0.065	−0.028	−0.143	1,649	0.044
	(4.74)	(−2.07)	(−0.46)	(−0.03)	(5.41)	(1.02)	(−0.87)	(−0.65)	(−0.26)	(−1.42)		
COL	0.155***	−0.111	–	0.229	0.067*	−0.015	–	−0.101	0.201	−0.288	1,122	0.002
	(6.46)	(−0.56)	–	(0.83)	(1.95)	(−0.08)	–	(−0.46)	(0.95)	(−1.42)		
DNK	0.065***	−0.342**	–	0.017	0.148***	0.003	–	−0.260*	−0.106	−0.030	1,567	0.024
	(3.96)	(−2.43)	–	(1.22)	(5.11)	(0.018)	–	(−1.83)	(−0.74)	(−0.17)		
FIN	0.074***	−0.514***	−0.520**	−0.018	0.143***	0.130	−0.217	−0.392**	−0.038	−0.019	1,609	0.027
	(4.41)	(−3.14)	(−2.16)	(−0.80)	(4.80)	(0.67)	(−0.97)	(−2.37)	(−0.23)	(−0.11)		
FRA	0.014*	−0.100	0.010	0.004	0.105***	−0.053	0.013	−0.096	−0.156**	−0.041	3,132	0.011
	(1.91)	(−1.49)	(0.43)	(0.61)	(4.28)	(−0.88)	(0.74)	(−1.57)	(−2.47)	(−0.72)		
DEU	0.020***	−0.111**	−0.144*	−0.002	0.151***	−0.035	−0.094	0.002	0.017	−0.088*	3,132	0.024
	(3.57)	(−2.03)	(−1.69)	(−0.23)	(6.69)	(−0.73)	(−1.03)	(0.03)	(0.32)	(−1.66)		
HKG	0.010***	−0.179***	−0.043	0.056	0.060***	−0.016	−0.036	−0.060	−0.056	0.030	3,140	0.010
	(2.79)	(−3.91)	(−0.78)	(0.89)	(2.92)	(−0.34)	(−0.64)	(−1.44)	(−1.51)	(0.85)		
IND	0.065***	−0.091	−0.126	0.419	0.179***	0.025	0.319	−0.146	−0.069	0.030	1,611	0.031
	(5.14)	(−1.10)	(−0.70)	(0.91)	(4.24)	(0.28)	(1.61)	(−1.45)	(−0.77)	(0.32)		
IDN	0.026***	−0.159**	–	0.275	0.360***	0.072	–	−0.056	0.000	0.091	2,051	0.133
	(3.56)	(−2.35)	–	(1.27)	(9.36)	(1.18)	–	(−0.91)	(0.00)	(1.63)		
IRL	0.026*	−0.380***	−0.129	−0.100	0.113***	−0.167	−0.010	−0.155	−0.145	−0.106	1,982	0.015
	(1.76)	(−3.03)	(−0.67)	(−1.26)	(4.13)	(−1.15)	(−0.05)	(−1.10)	(−1.12)	(−0.83)		

Continue on the next page

Table continued from previous page

Country OI	$Constant$	ΔVIX_{t-1}	$\Delta LVIX^o_{t-1}$	ΔIR_{t-1}	OI_{t-1}	ΔVIX_{t-2}	$\Delta LVIX^o_{t-2}$	ΔVIX_{t-3}	ΔVIX_{t-4}	ΔVIX_{t-5}	Obs.	R^2
ISR	−0.007	−0.202	–	−0.072	0.097***	−0.167	–	−0.131	−0.030	−0.127	2,583	0.010
	(−0.70)	(−1.59)	–	(−0.27)	(4.56)	(−1.25)	–	(−1.27)	(−0.28)	(−0.88)		
ITA	0.017**	−0.142**	−0.003	−0.012**	0.161***	−0.050	0.005	−0.019	−0.029	−0.014	3,132	0.026
	(2.34)	(−2.28)	(−0.03)	(−2.20)	(6.09)	(−0.78)	(0.06)	(−0.28)	(−0.45)	(−0.21)		
JPN	0.006*	−0.016	−0.017	−0.003	0.152***	−0.006	0.015	0.033	−0.005	0.048*	3,083	0.022
	(1.67)	(−0.53)	(−0.41)	(−1.05)	(6.29)	(−0.18)	(0.44)	(1.02)	(−0.18)	(1.65)		
MYS	0.004	−0.172***	–	−0.276	0.117***	−0.005	–	−0.149***	0.035	−0.047	3,140	0.022
	(1.01)	(−4.09)	–	(−1.17)	(5.03)	(−0.12)	–	(−3.64)	(0.93)	(−1.25)		
MEX	0.005*	−0.036	−0.001	0.513	0.067**	−0.009	−0.017	0.027	0.012	−0.055**	2,870	0.004
	(1.90)	(−1.31)	(−0.01)	(1.41)	(2.55)	(−0.34)	(−0.34)	(0.98)	(0.40)	(−1.99)		
NLD	0.003	−0.333***	−0.061	−0.019	0.155***	0.006	−0.142	0.038	−0.177**	−0.035	3,132	0.029
	(0.33)	(−3.72)	(−0.69)	(−1.38)	(5.99)	(0.06)	(−1.35)	(0.48)	(−2.27)	(−0.40)		
NZL	0.020*	−0.037	–	−0.229	0.167***	−0.031	–	−0.184*	−0.056	−0.074	1,969	0.027
	(1.68)	(−0.32)	–	(−0.92)	(6.65)	(−0.30)	–	(−1.73)	(−0.62)	(−0.78)		
NOR	0.087***	−0.359**	–	−0.742	0.065**	−0.297	–	−0.137	−0.208	−0.258	1,619	0.008
	(5.34)	(−2.50)	–	(−0.79)	(2.39)	(−1.99)	–	(−1.00)	(−1.58)	(−1.67)		
PER	−0.018*	−0.314***	–	0.003***	0.299***	−0.039	–	−0.095	−0.271***	0.143	2,272	0.098
	(−1.73)	(−3.61)	–	(18.36)	(10.94)	(−0.44)	–	(−1.05)	(−3.07)	(1.30)		
PHL	0.069***	−0.088	–	−0.024*	0.178***	−0.146	–	−0.322*	−0.364*	−0.040	638	0.034
	(3.77)	(−0.47)	–	(−1.73)	(4.01)	(−0.68)	–	(−1.71)	(−1.76)	(−0.22)		
POL	0.027***	−0.329***	–	−0.543	0.150***	−0.116	–	−0.233***	−0.038	0.081	2,038	0.033
	(2.97)	(−4.07)	–	(−0.25)	(5.72)	(−1.29)	–	(−2.72)	(−0.45)	(0.91)		
QAT	0.066***	−0.568**	–	0.318	0.079**	0.193	–	−0.048	−0.083	0.344	790	0.006
	(2.76)	(−2.28)	–	(0.20)	(2.42)	(0.69)	–	(−0.24)	(−0.31)	(1.22)		
RUS	0.012	−0.265***	−0.042	0.549	0.157***	−0.116	−0.124	−0.037	0.054	0.106	1,535	0.026
	(1.08)	(−2.82)	(−0.39)	(1.01)	(5.29)	(−1.05)	(−1.10)	(−0.32)	(0.44)	(0.85)		
SAU	0.087***	−0.370*	–	16.635***	0.057	−0.025	–	−0.106	−0.075	0.288	638	0.020
	(3.44)	(−1.65)	–	(3.27)	(1.38)	(−0.13)	–	(−0.52)	(−0.28)	(1.37)		
SGP	0.011**	−0.207***	–	−0.146	0.117***	−0.070*	–	−0.036	−0.061	0.007	3,140	0.022
	(2.31)	(−4.42)	–	(−1.32)	(3.57)	(−1.84)	–	(−0.99)	(−1.34)	(0.20)		
ZAF	0.002	−0.161***	−0.120	−0.005	0.081***	0.012	−0.334***	−0.024	−0.018	−0.045	2,871	0.011
	(0.57)	(−3.25)	(−0.76)	(−0.03)	(4.11)	(0.24)	(−2.73)	(−0.42)	(−0.35)	(−0.91)		
KOR	0.000	−0.097***	−0.044	−0.477	0.077***	0.003	−0.046	−0.033	−0.033	0.019	3,140	0.010
	(0.01)	(−3.85)	(−0.99)	(−1.64)	(2.99)	(0.09)	(−1.19)	(−1.22)	(−1.26)	(0.76)		
ESP	0.011	−0.138**	−0.019	−0.007	0.268***	0.023	−0.067	−0.018	−0.164***	−0.094	3,132	0.075
	(1.55)	(−2.32)	(−0.23)	(−1.08)	(9.59)	(0.43)	(−0.76)	(−0.33)	(−2.64)	(−1.47)		
SWE	0.011	−0.334***	−0.167**	0.007	0.216***	−0.041	−0.240***	−0.105	−0.201***	−0.199***	3,139	0.065
	(1.43)	(−5.27)	(−2.08)	(0.97)	(8.96)	(−0.56)	(−2.60)	(−1.48)	(−2.67)	(−2.83)		
CHE	0.024***	−0.198***	0.146	−0.001	0.138***	−0.098	−0.070	−0.119*	−0.048	0.001	3,055	0.022
	(3.68)	(−3.33)	(1.42)	(−0.47)	(5.68)	(−1.54)	(−0.64)	(−1.83)	(−0.87)	(0.01)		
TWN	0.005*	−0.074***	–	−0.267**	0.063***	−0.023	–	−0.029	−0.044*	0.001	3,140	0.006
	(1.83)	(−2.66)	–	(−2.12)	(2.80)	(−0.94)	–	(−1.04)	(−1.80)	(0.04)		
THA	0.021***	−0.438***	–	−0.020***	0.194***	−0.107	–	−0.167**	−0.064	−0.105*	2,583	0.057
	(2.73)	(−5.60)	–	(−3.53)	(6.03)	(−1.50)	–	(−2.25)	(−1.01)	(−1.78)		
TUR	0.004	−0.417***	–	0.065	0.150***	−0.046	–	−0.124**	−0.046	0.023	2,581	0.041
	(0.58)	(−6.39)	–	(0.13)	(6.17)	(−0.77)	–	(−2.15)	(−0.94)	(0.38)		
ARE	0.019	0.106	–	−0.048	0.119***	−0.244	–	−0.251	−0.133	−0.131	1,029	0.012
	(0.94)	(0.61)	–	(−1.13)	(3.93)	(−1.35)	–	(−1.43)	(−0.67)	(−0.68)		
GBR	0.027***	−0.153***	−0.117*	−0.037	0.161***	−0.052	−0.190***	0.000	−0.081	−0.081	3,140	0.032
	(5.00)	(−2.86)	(−1.88)	(−0.85)	(5.87)	(−0.97)	(−2.65)	(−0.01)	(−1.65)	(−1.46)		

Table 2.4: Panel Results

Panel regressions with random effects for predictive regression of order imbalance for 41 country-level MSCI based iShares ETFs on percentage change in volatility index (ΔVIX), percentage change in orthogonalised local volatility index ($\Delta LVIX^o$), change in local interest rates (IR), lagged order imbalance (OI_{t-1}), dummy variable for a common language (L) and a distance to the U.S. (G), as well as lags of ΔVIX and $\Delta LVIX^o$. Specifically, I estimate the model below:

$$OI_{i,t} = \alpha + \sum_{k=1}^{5} \beta_k \Delta VIX_{t-k} + \sum_{k=1}^{2} \gamma_k \Delta LVIX^o_{i,t-k} + \delta_1 \Delta IR_{i,t-1} + \delta_2 L_i + \delta_3 G_i + \delta_4 TOI_{i,t-1} + \varepsilon$$

The number of lags in the model is determined using Akaike Information Criterion and Bayesian information criterion jointly. Coefficients and z-statistics (based on robust standard errors) are presented. The lag for predictive regression is 1 day. Data is at a daily frequency and covers the period from 2006 or the first trading day of ETF (whichever is later) until the end of June 2018. Z-statistics is reported for difference in coefficients test $H_o : \beta_1 \geq \gamma_1$.***,**,* denote significance at the 1%, 5% and 10% level.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
VARIABLES	*TOI*	*TOI*	*TOI*	*smallOI*	*smallOI*	*smallOI*	*retailOI*	*retailOI*	*retailOI*	*largeOI*	*largeOI*	*largeOI*
ΔVIX_{t-1}			-0.183***			-0.117***			-0.261***			-0.243***
			(-7.25)			(-4.80)			(-9.67)			(-8.10)
$\Delta LVIX^o_{t-1}$	-0.060***	-0.059***	-0.044**	-0.049***	-0.047***	-0.035**	-0.046*	-0.047*	-0.007	-0.062**	-0.062**	-0.043
	(-2.82)	(-2.81)	(-2.11)	(-3.03)	(-2.99)	(-2.29)	(-1.80)	(-1.82)	(-0.25)	(-2.21)	(-2.21)	(-1.51)
ΔIR_{t-1}		-0.004**	-0.004**		-0.002	-0.002		-0.006*	-0.006**		-0.002	-0.002
		(-2.18)	(-2.29)		(-1.53)	(-1.59)		(-1.95)	(-1.96)		(-0.76)	(-0.85)
OI_{t-1}	0.174***	0.174***	0.170***	0.197***	0.197***	0.194***	0.134***	0.133***	0.129***	0.117***	0.117***	0.113***
	(12.42)	(12.38)	(12.30)	(8.24)	(8.27)	(8.22)	(8.80)	(8.97)	(8.86)	(10.99)	(10.91)	(10.56)
L		0.003	0.003		-0.005**	-0.005**		0.013	0.013		0.001	0.001
		(0.36)	(0.36)		(-2.25)	(-2.24)		(0.86)	(0.86)		(0.13)	(0.13)
G		-0.000	-0.000		-0.000	-0.000		-0.000**	-0.000**		0.000	0.000
		(-0.20)	(-0.20)		(-0.63)	(-0.62)		(-2.52)	(-2.52)		(0.15)	(0.15)

Continue on the next page

Table continued from previous page

VARIABLES	(1) *TOI*	(2) *TOI*	(3) *TOI*	(4) *smallOI*	(5) *smallOI*	(6) *smallOI*	(7) *retailOI*	(8) *retailOI*	(9) *retailOI*	(10) *largeOI*	(11) *largeOI*	(12) *largeOI*
ΔVIX_{t-2}			-0.031**			-0.026**			-0.103***			-0.044*
			(-2.23)			(-2.06)			(-4.59)			(-1.75)
$\Delta LVIX^{o}_{t-2}$	-0.075***	-0.074***	-0.054**	-0.036	-0.036	-0.026	-0.076***	-0.077***	-0.052*	-0.093***	-0.093***	-0.071***
	(-2.93)	(-2.93)	(-2.28)	(-1.61)	(-1.61)	(-1.20)	(-2.66)	(-2.66)	(-1.92)	(-3.30)	(-3.31)	(-2.63)
ΔVIX_{t-3}			-0.056***			-0.028***			-0.062***			-0.062***
			(-3.26)			(-2.78)			(-2.96)			(-2.70)
ΔVIX_{t-4}			-0.072***			-0.058***			-0.084***			-0.086***
			(-4.68)			(-3.69)			(-5.09)			(-4.15)
ΔVIX_{t-5}			-0.045***			-0.031**			-0.028			-0.057**
			(-2.88)			(-2.26)			(-1.50)			(-2.29)
Constant	0.014***	0.014*	0.015**	-0.000	0.003	0.004	0.009	0.038***	0.040***	0.010***	0.008	0.009
	(3.87)	(1.94)	(2.09)	(-0.09)	(1.00)	(1.17)	(1.24)	(3.04)	(3.14)	(2.87)	(1.15)	(1.38)
Observations	57,916	57,725	57,683	52,811	52,620	52,578	54,642	54,455	54,455	55,572	55,388	55,346
Adjusted R^2	0.029	0.029	0.032	0.039	0.039	0.040	0.016	0.016	0.019	0.014	0.014	0.016
$H_o : \beta_1 \geq \gamma_1$			(-5.21)			(-4.25)			(-8.29)			(-5.80)

Table 2.5: Cash ETFs and VIX

Predictive regression of average order imbalance for U.S. cash ETFs ($cashOI$) on percentage change in volatility index (ΔVIX), orthogonalised percentage change in local volatility index ($\Delta LVIX^o$), change in local interest rates (IR), lagged order imbalance (OI_{t-1}), as well as lags of ΔVIX and $\Delta LVIX^o$. Specifically, I estimate the model below:

$$CashOI_{i,t} = \alpha + \sum_{k=1}^{5} \beta_k \Delta VIX_{t-k} + \sum_{k=1}^{3} \gamma_k \Delta LVIX^o_{i,t-k} + \gamma_5 \Delta LVIX^o_{i,t-5} + \\ + \delta_1 \Delta IR_{i,t-1} + \delta_2 CashOI_{i,t-1} + \varepsilon_{i,t}$$

Cash ETFs are defined as funds that invest in U.S. treasuries with less than 1 year of maturity. The number of lags in the model is determined using Akaike Information Criterion and Bayesian information criterion jointly. Coefficients and t-statistics (based on Newey and West (1987) robust standard errors) are presented. The lag for predictive regression is 1 day. Data is at a daily frequency and covers the period from 2006 or the first trading day of ETF (whichever is later) until the end of June 2018. ***,**,* denote significance at the 1%, 5% and 10% level.

Country	Const	ΔVIX_{t-1}	$\Delta LVIX^o_{t-1}$	ΔIR_{t-1}	$cashOI_{t-1}$	ΔVIX_{t-2}	$\Delta LVIX^o_{t-2}$	ΔVIX_{t-3}	$\Delta LVIX^o_{t-3}$	ΔVIX_{t-4}	ΔVIX_{t-5}	$\Delta LVIX^o_{t-5}$	Obs.	R^2
AUS	0.020**	0.289***	0.039	-0.065	0.301***	0.383***	0.002	0.233***	0.007	0.21***	0.257***	0.022	2,637	0.113
	(2.22)	(3.83)	(0.37)	(-0.27)	(13.28)	(3.84)	(0.02)	(2.74)	(0.07)	(2.58)	(3.37)	(0.22)		
BEL	0.096***	0.217	0.193	0.12	0.392***	0.451***	0.058	0.264	0.131	0.419**	0.045	0.184	966	0.176
	(5.58)	(1.38)	(0.77)	(0.72)	(8.83)	(2.85)	(0.28)	(1.62)	(0.68)	(2.47)	(0.22)	(1.06)		
CAN	0.007	0.290***	-0.037	0.034	0.276***	0.391***	-0.005	0.208***	0.114	0.179**	0.259***	0.016	2,196	0.097
	(0.74)	(3.59)	(-0.63)	(0.39)	(11.5)	(3.89)	(-0.06)	(2.76)	(1.56)	(2.20)	(3.24)	(0.25)		
CHN	0.035***	0.314***	-0.097	0.003	0.35***	0.478***	0.073	0.254***	-0.174	0.306***	0.197**	0.023	2,723	0.146
	(3.68)	(3.58)	(-0.61)	(0.08)	(13.05)	(5.65)	(0.46)	(3.12)	(-1.12)	(3.05)	(1.99)	(0.15)		
FRA	0.039***	0.307***	0.039	0.02*	0.341***	0.369***	0.063**	0.208***	0.051*	0.24***	0.169*	0.019	2,878	0.140
	(4.18)	(4.02)	(1.20)	(1.66)	(12.92)	(4.16)	(2.01)	(2.99)	(1.66)	(2.95)	(1.93)	(0.77)		
DEU	0.039***	0.304***	0.125	0.02*	0.343***	0.349***	0.143	0.189**	0.016	0.247***	0.168*	0.085	2,878	0.139
	(4.18)	(3.99)	(0.93)	(1.67)	(12.98)	(3.8)	(1.00)	(2.39)	(0.14)	(2.93)	(1.91)	(0.70)		
HKG	0.038***	0.303***	-0.016	0.202	0.342***	0.396***	-0.013	0.235***	-0.078	0.279***	0.181**	0.176*	2,886	0.140
	(4.11)	(4.00)	(-0.13)	(1.18)	(12.98)	(4.06)	(-0.11)	(3.13)	(-0.68)	(3.13)	(2.08)	(1.65)		
IND	0.022**	0.275***	-0.079	0.255***	0.304***	0.404***	0.132	0.217***	-0.06	0.222***	0.257***	0.021	2,678	0.116
	(2.45)	(3.67)	(-0.67)	(2.61)	(13.59)	(4.44)	(1.12)	(2.91)	(-0.54)	(2.91)	(3.43)	(0.2)		
ITA	0.039***	0.305***	0.041	0.019*	0.343***	0.366***	0.146	0.187***	0.075	0.235***	0.168*	0.072	2,878	0.139
	(4.17)	(4.01)	(0.34)	(1.64)	(12.97)	(4.10)	(1.10)	(3.32)	(0.68)	(2.80)	(1.91)	(0.66)		
JPN	0.04***	0.326***	0.022	0.010	0.348***	0.369***	0.015	0.229***	0.114	0.209**	0.158*	-0.011	2,829	0.142
	(4.29)	(4.2)	(0.21)	(1.17)	(13.06)	(3.63)	(0.15)	(2.76)	(1.06)	(2.32)	(1.78)	(-0.11)		
MEX	0.033***	0.286***	0.272**	1.416	0.355***	0.459***	0.112	0.247***	0.119	0.300***	0.166	0.061	2,616	0.152
	(3.44)	(3.22)	(2.01)	(1.19)	(12.93)	(5.68)	(0.66)	(3.09)	(0.73)	(3.06)	(1.61)	(0.48)		
NLD	0.039***	0.306***	-0.008	0.019	0.343***	0.376***	0.120	0.191**	0.064	0.236***	0.168*	-0.054	2,878	0.139
	(4.16)	(4.00)	(-0.08)	(1.64)	(13.01)	(4.18)	(0.84)	(2.35)	(0.55)	(2.74)	(1.88)	(-0.42)		
RUS	0.027***	0.291***	-0.148	-0.003	0.349***	0.534***	0.075	0.266***	-0.118	0.333***	0.182*	0.049	2,500	0.149
	(2.74)	(3.25)	(-1.01)	(-0.01)	(12.22)	(6.36)	(0.62)	(3.08)	(-1.18)	(3.31)	(1.72)	(0.55)		
ZAF	0.038***	0.305***	0.908***	-0.638**	0.322***	0.315***	0.648***	0.159**	0.394	0.211**	0.152*	0.423**	2,868	0.139
	(4.05)	(3.98)	(3.86)	(-2.02)	(12.73)	(3.7)	(3.13)	(2.27)	(1.59)	(2.52)	(1.69)	(2.14)		
KOR	0.039***	0.309***	-0.168	2.298***	0.342***	0.442***	0.024	0.235***	0.018	0.253***	0.176**	0.165	2,886	0.141
	(4.17)	(4.11)	(-1.31)	(2.97)	(13.08)	(4.34)	(0.19)	(3.11)	(0.15)	(2.94)	(1.99)	(1.53)		
SWE	0.039***	0.300***	0.064	-0.009	0.344***	0.366***	0.229**	0.167**	0.104	0.225***	0.173**	0.019	2,885	0.140
	(4.17)	(3.93)	(0.61)	(-0.62)	(13.01)	(4.12)	(2.16)	(2.17)	(1.09)	(2.7)	(1.98)	(0.20)		
CHE	0.039***	0.289***	0.090	-0.006**	0.344***	0.35***	0.128	0.185**	0.088	0.204**	0.189**	-0.013	2,801	0.139
	(4.03)	(3.74)	(0.66)	(-2.12)	(12.73)	(3.63)	(0.94)	(2.26)	(0.66)	(2.26)	(2.09)	(-0.10)		
GBR	0.039***	0.296***	0.134	0.018	0.343***	0.342***	0.125	0.184**	0.092	0.223***	0.169*	0.032	2,886	0.139
	(4.15)	(3.92)	(1.42)	(0.28)	(13.01)	(3.78)	(1.09)	(2.41)	(0.99)	(2.65)	(1.91)	(0.34)		

Table 2.6: Asymmetric Reaction to Changes in VIX and LVIX

Panel regressions with random effects for predictive regression of order imbalance for 41 country-level MSCI based iShares ETFs on percentage change in volatility index (ΔVIX), percentage change in orthogonalised local volatility index ($\Delta LVIX^o$), ΔVIX^{Qk} and $\Delta LVIX^{o,Qk}$ (where $k \in [2,5]$) to capture asymmetric response to magnitude and direction of news, change in local interest rates (IR), lagged order imbalance (OI_{t-1}), dummy variable for a common language (L) and a distance to the U.S. (G), as well as lags of ΔVIX and $\Delta LVIX^o$. ΔVIX is split into quintiles by size. ΔVIX^{Q1} (ΔVIX^{Q5}) equals ΔVIX if the change is the smallest (the largest) and 0 otherwise. The same applies for $\Delta LVIX^{Q1}$ ($\Delta LVIX^{Q5}$). Specifically, I estimate the model below:

$$OI_{i,t} = \alpha + \sum_{k=2}^{5} \beta_k \Delta VIX_{t-1}^{Qk} + \sum_{k=2}^{5} \gamma_k \Delta LVIX_{t-1}^{o,Qk} + \sum_{I=1}^{5} \eta_I \Delta VIX_{t-I} + \sum_{I=1}^{2} \delta_I \Delta LVIX_{t-I}^{o} + \boldsymbol{\mu} Controls + \varepsilon_{i,t}$$

where the set of $Controls$ includes changes in the interests (ΔIR), order imbalances (OI) at time $t-1$, as well as as dummy variable for a common language (L) and a distance measure to the U.S. (G). The number of lags in the model is determined using Akaike Information Criterion and Bayesian information criterion jointly. Coefficients and z-statistics (based on robust standard errors) are presented. The lag for predictive regression is 1 day. Data is at a daily frequency and covers the period from 2006 or the first trading day of ETF (whichever is later) until the end of June 2018. ***,**,* denote significance at the 1%, 5% and 10% level.

	(1)	(2)	(3)	(4)
VARIABLES	*TOI*	*smallOI*	*retailOI*	*largeOI*
ΔVIX_{t-1}	-0.055*	0.011	-0.230***	-0.095**
	(-1.93)	(0.41)	(-3.55)	(-2.18)
ΔVIX_{t-1}^{Q2}	-0.101	-0.336***	-0.325**	0.012
	(-1.40)	(-6.31)	(-2.43)	(0.13)
ΔVIX_{t-1}^{Q3}	0.002	0.114	-0.455	-0.098
	(0.01)	(0.50)	(-1.17)	(-0.30)
ΔVIX_{t-1}^{Q4}	-0.388***	-0.227**	-0.048	-0.453**
	(-2.58)	(-2.01)	(-0.25)	(-2.10)
ΔVIX_{t-1}^{Q5}	-0.174***	-0.163***	-0.018	-0.207***
	(-4.18)	(-3.62)	(-0.23)	(-3.47)
$\Delta LVIX_{t-1}^{o}$	0.025	0.045	0.080	0.043
	(0.60)	(1.26)	(1.12)	(0.78)
$\Delta LVIX_{t-1}^{o,Q2}$	0.006	-0.024	0.012	-0.034
	(0.04)	(-0.26)	(0.06)	(-0.18)
$\Delta LVIX_{t-1}^{o,Q3}$	0.429	0.088	0.854	0.536
	(0.91)	(0.25)	(1.09)	(0.95)
$\Delta LVIX_{t-1}^{o,Q4}$	0.075	0.035	0.230	0.029
	(0.40)	(0.28)	(0.75)	(0.12)
$\Delta LVIX_{t-1}^{o,Q5}$	-0.114**	-0.126***	-0.141*	-0.137*
	(-2.06)	(-2.88)	(-1.67)	(-1.83)
ΔIR_{t-1}	-0.004**	-0.002	-0.006*	-0.002
	(-2.22)	(-1.54)	(-1.96)	(-0.79)

Continue on the next page

Table continued from previous page

	(1)	(2)	(3)	(4)
VARIABLES	TOI	$smallOI$	$retailOI$	$largeOI$
G	-0.000	-0.000	-0.000**	0.000
	(-0.22)	(-0.69)	(-2.56)	(0.13)
L	0.003	-0.005**	0.013	0.001
	(0.38)	(-2.23)	(0.87)	(0.14)
OI_{t-1}	0.169***	0.193***	0.129***	0.113***
	(12.31)	(8.23)	(8.87)	(10.57)
ΔVIX_{t-2}	-0.020	-0.015	-0.100***	-0.031
	(-1.48)	(-1.16)	(-4.06)	(-1.29)
$\Delta LVIX^{o}_{t-2}$	-0.055**	-0.026	-0.050*	-0.071***
	(-2.32)	(-1.20)	(-1.86)	(-2.71)
ΔVIX_{t-3}	-0.042**	-0.014	-0.057**	-0.046**
	(-2.53)	(-1.35)	(-2.43)	(-2.13)
ΔVIX_{t-4}	-0.062***	-0.048***	-0.079***	-0.074***
	(-4.15)	(-3.23)	(-4.44)	(-3.66)
ΔVIX_{t-5}	-0.036**	-0.021	-0.022	-0.046*
	(-2.45)	(-1.58)	(-1.16)	(-1.95)
Constant	0.022***	0.008**	0.040***	0.018***
	(2.85)	(2.09)	(3.31)	(2.74)
Observations	57,683	52,578	54,455	55,346
Adjusted R^2	0.032	0.041	0.020	0.016

Table 2.7: Panel Results- Political Uncertainty

Panel regressions with random effects for predictive regression of order imbalance for 41 country-level MSCI based iShares ETFs region on percentage change in volatility index (ΔVIX), percentage change in orthogonalised local volatility index ($\Delta LVIX^o$), U.S. political uncertainty ($USPU$), orthogonalised local political uncertainty (LPU^o), change in local interest rates (IR), lagged order imbalance (OI_{t-1}), dummy variable for a common language (L) and a distance to the US (G), as well as lags of ΔVIX, $\Delta LVIX^o$, $USPR$ and LPR^o. Specifically, I run the model below:

$$OI_{i,t} = \alpha + \sum_{k=1}^{5} \beta_k \Delta VIX_{t-k} + \sum_{k=1}^{2} \gamma_k \Delta LVIX^o_{i,t-k} + \sum_{k=1}^{2} \delta_k USPU_{t-1} + \sum_{k=1}^{1,5} \phi_k LPU^o_{t-1} + \boldsymbol{\mu} Controls + \varepsilon_{i,t}$$

where the set of $Controls$ includes changes in the interests (ΔIR), order imbalances (OI) at time $t-1$, as well as as dummy variable for a common language (L) and a distance measure to the U.S. (G). Political uncertainty is based on SVI from Google trends (as described in section 2.3.4). The number of lags in the model is determined using Akaike Information Criterion and Bayesian information criterion jointly. Coefficients and z-statistics (based on robust standard errors) are presented. The lag for predictive regression is 1 day. Data is at a daily frequency and covers the period from 2006 or the first trading day of ETF (whichever is later) until the end of June 2018. ***,**,* denote significance at the 1%, 5% and 10% level.

	(1)	(2)	(3)	(4)
VARIABLES	*TOI*	*smallOI*	*retailOI*	*largeOI*
ΔVIX_{t-1}	-0.172***	-0.107***	-0.257***	-0.233***
	(-7.10)	(-4.49)	(-9.42)	(-7.94)
$\Delta LVIX^o_{t-1}$	-0.046**	-0.035***	-0.013	-0.050*
	(-2.39)	(-2.72)	(-0.50)	(-1.79)
$USPU_{t-1}$	0.010	0.017**	-0.014	0.014
	(1.07)	(2.08)	(-0.92)	(1.19)
LPU^o_{t-1}	-0.010	-0.004	0.005	-0.012
	(-0.98)	(-0.43)	(0.37)	(-0.86)
ΔIR_{t-1}	-0.004**	-0.002*	-0.006*	-0.002
	(-2.37)	(-1.68)	(-1.85)	(-0.86)
OI_{t-1}	0.166***	0.181***	0.131***	0.109***
	(14.03)	(9.98)	(9.09)	(10.20)
L	0.005	-0.003	0.013	0.004
	(0.66)	(-1.50)	(0.89)	(0.45)
G	-0.000	0.000	-0.000**	0.000
	(-0.00)	(0.15)	(-2.54)	(0.41)
$USPU_{t-2}$	0.044***	0.050***	0.027	0.040***
	(4.33)	(5.79)	(1.62)	(3.87)
LPU^o_{t-5}	-0.007	-0.004	-0.012	-0.002
	(-0.99)	(-0.63)	(-1.20)	(-0.14)
ΔVIX_{t-2}	-0.037**	-0.024*	-0.109***	-0.049*
	(-2.42)	(-1.67)	(-4.83)	(-1.91)
$\Delta LVIX^o_{t-2}$	-0.048**	-0.024	-0.049*	-0.067**
	(-1.97)	(-1.08)	(-1.74)	(-2.36)
ΔVIX_{t-3}	-0.061***	-0.030***	-0.064***	-0.066***
	(-3.20)	(-2.93)	(-2.96)	(-2.80)
ΔVIX_{t-4}	-0.072***	-0.053***	-0.076***	-0.082***
	(-4.42)	(-3.34)	(-4.62)	(-3.91)
ΔVIX_{t-5}	-0.036**	-0.022	-0.026	-0.047**
	(-2.54)	(-1.55)	(-1.52)	(-2.04)
Constant	0.010	-0.003	0.038***	0.004
	(1.45)	(-1.24)	(3.09)	(0.63)
Observations	55,567	50,543	53,863	53,493
Adjusted R^2	0.031	0.036	0.019	0.015

Table 2.8: Correlation with the U.S. Market: Staggered Introduction

Panel regression with random effects of innovations in rolling 36-day correlation (measured as a residual of AR(1) process) between S&P 500 returns and local MSCI market returns ($\Delta\rho$) on the dummy variable (for the introduction of ETFs) $Intro^{ETF}$, on the dummy variable (for periods of high U.S. volatility conditional on low volatility in local markets) $D_{US|L}$, the interaction of two dummy variables and the lagged innovation in correlation. Specifically, I estimate the model below:

$$\Delta\rho_t = \alpha + \beta_1 Intro_t^{ETF} + \beta_2 Intro_t^{ETF} D_{US|L,t} + \beta_3 D_{US|L,t} + \beta_4 \Delta\rho_{t-1} + \varepsilon_{i,t}$$

Dummy variable $D_{US|L}$ takes the value of 1 when volatility in U.S. market is high (greater than mean plus one standard deviation) and volatility in the local market is not high (less than mean plus one standard deviation). Dummy variable $Intro^{ETF}$ takes the value of 1 throughout the life of an ETF (from the inception date until the fund end date) and is 0 otherwise. Data is at a monthly frequency and covers the period from January 1988 or the first trading day on which MSCI return is available (whichever is later) until the end of June 2018. ***,**,* denote significance at the 1%, 5% and 10% level.

	(1)
VARIABLES	$\Delta\rho_t$
$Intro^{ETF}$	0.001
	(1.10)
$Intro^{ETF} \times D_{US\|L,t}$	0.008**
	(2.59)
$D_{US\|L,t}$	0.002
	(0.78)
$\Delta\rho_{t-1}$	0.015
	(1.28)
Constant	-0.001
	(-1.60)
Observations	12,268
Adjusted R^2	0.004

Table 2.9: Correlation with the U.S. Market: Arbitrage and High Volatility

Panel regression with random effects of innovations in rolling 36-month correlation (measured as a residual of AR(1) process) between S&P 500 returns and local MSCI market returns ($\Delta\rho$) on the dummy variable (for periods of high U.S. volatilities conditional on low volatility in local markets) $D_{US|L}$, measure of arbitrage activity ($|\Delta SO_t|$), interaction of the dummy variable with the arbitrage proxy, lagged innovation in correlation and Amihud (2002) illiquity measure of MSCI index ($ILLIQ_{MSCI}$). Specifically, I estimate the model below:

$$\Delta\rho_t = \alpha + \beta_1|\Delta SO_t| + \beta_2|\Delta SO_t|D_{US|L,t} + \beta_3 D_{US|L,t} + \beta_4\Delta\rho_{t-1} + \beta_5 ILLIQ_{MSCI,t} + \varepsilon_{i,t}$$

Dummy variable takes the value of 1 when volatility in U.S. market is high (greater than mean plus one standard deviation) and volatility in the local market is not high (less than mean plus one standard deviation). Arbitrage activity $|\Delta SO_t|$ is proxied by an absolute value of percentage change of shares outstanding of ETF fund from time $t-1$ to t. Data is at a monthly frequency and covers the period from 2006 or the first trading day of ETF (whichever is later) until the end of June 2018. Illiquidity measure is in 10^{-4} ***,**,* denote significance at the 1%, 5% and 10% level.

	(1)	(2)
VARIABLES	$\Delta\rho_t$	$\Delta\rho_t$
$\lvert\Delta SO_t\rvert$	0.23*	-0.048
	(1.80)	(-0.33)
$\lvert\Delta SO_t\rvert \times D_{US\lvert L,t}$		0.649**
		(2.05)
$D_{US\lvert L,t}$		-0.003
		(-1.36)
$\Delta\rho_{t-1}$	0.075***	0.136***
	(3.99)	(10.34)
$ILLIQ_{MSCI,t}$	0.56*	1.35***
	(1.67)	(2.62)
Constant	-0.001**	-0.000
	(-2.46)	(-0.68)
Observations	3,243	1,939
Adjusted R^2	0.007	0.019

Table 2.10: Price Discovery, Market Illiquidity and Correlation

This table presents average 36-month rolling correlations of MSCI index returns on S&P 500 returns and average adjusted R-squares of MSCI index returns on the first principal component (PC1) of all available countries. I form 3 (or 4) monthly portfolios based on price discovery and limits to arbitrage proxies. Each portfolio is presorted based on information available during previous month. Panel A shows the sort into terciles by β_i from equation $R_{i,t}^{NAV} = \alpha + \beta_i \left(\frac{P_{i,t-1} - NAV_{i,t-1}}{NAV_{i,t-1}} \right) + \varepsilon$. I use a 36-month period to estimate β_i. Sort in panel B is based on Amihud's illiquidity ratio (Amihud (2002)) of the underlying markets ($ILLIQ$). Panel C shows the result of a double sort by median based on proxies from Panel A and Panel B. Panel D shows the result of a double sort by median based on price discovery as in panel A and illiquidity mismatch as defined in equation 2.12. t-statistics is based on Newey and West (1987) robust standard errors.

	Panel A: Price Discovery				
Sorting Variable	low	medium	high	*HML*	*t-stat*
	Rolling Correlations				
NAV sensitivity to Premium	0.59	0.69	0.76	0.17	(32.02)
	Adjusted R-squared based on PC1				
NAV sensitivity to Premium	0.50	0.61	0.67	0.17	(10.67)
	Panel B: Limits to Arbitrage				
Sorting Variable	low	medium	high	*HML*	*t-stat*
	Rolling Correlations				
Amihud's Illiquidity Ratio	0.75	0.69	0.65	-0.11	(-10.11)
	Adjusted R-squared based on PC1				
Amihud's Illiquidity Ratio	0.66	0.61	0.54	-0.12	(-10.26)
	Panel C: Double Sort 1				
		Amihud's Illiquidity Ratio			
Sorting Variables		low	high	*HML*	*t-stat*
	Rolling Correlations				
NAV sensitivity to Premium	low	0.67	0.60	-0.07	(-5.81)
	high	0.76	0.72	-0.04	(-3.31)
	HML	0.09	0.12		
	t-stat	(8.71)	(14.08)		
	Panel D: Double Sort 2				
		Liquidity Mismatch			
Sorting Variables		low	high	*HML*	*t-stat*
	Rolling Correlations				
NAV sensitivity to Premium	low	0.61	0.65	0.06	(3.25)
	high	0.71	0.77	0.06	(5.04)
	HML	0.10	0.11		
	t-stat	(9.48)	(6.54)		

Table 2.11: Price Discovery, Noise and Correlation

This table presents average 36-month rolling correlations of MSCI index returns on S&P 500 returns for all available countries. I form 3 (or 4) monthly portfolios based on one of the three proxies that affect price discovery due to noise. Each portfolio is presorted based on information available during previous month. Panel A shows the sort into terciles by the $|VR-1|$, where VR (variance ratio) is the ratio of 15-sec returns over three times of 5-sec return variances. Sort in panel B is based on the ratio of dollar volume of $smallOI$ and of TOI. Panel C shows the result of a sort by ratio of dollar volume of $smallOI$ and $retailOI$. For each panel, the double sort by median based on one of 3 sorting criteria and Amihud's illiquidity ratio (Amihud (2002)) of the underlying markets ($ILLIQ$) is also provided. t-statistics is based on Newey and West (1987) robust standard errors.

	Panel A: Variance Ratio				
Sorting Variable	low	medium	high	*HML*	*t-stat*
Variance Ratio	0.72	0.67	0.68	-0.03	(-2.64)
		Variance Ratio			
Sorting Variables		low	high	*HML*	*t-stat*
Amihud's Illiquidity Ratio	low	0.73	0.73	-0.01	(-0.42)
	high	0.65	0.65	-0.01	(-0.45)
	HML	-0.08	-0.08		
	t-stat	(-6.55)	(-4.27)		
	Panel B: Small Dollar Volume to Total Dollar Volume				
Sorting Variable	low	medium	high	*HML*	*t-stat*
Small/Totl Dollar Volume	0.71	0.73	0.66	-0.05	(-2.93)
		Small/Total Dollar Volume			
Sorting Variables		low	high	*HML*	*t-stat*
Amihud's Illiquidity Ratio	low	0.75	0.72	-0.02	(-2.79)
	high	0.66	0.64	-0.02	(-1.56)
	HML	-0.08	-0.08		
	t-stat	(-7.91)	(-11.41)		
	Panel C: Small Dollar Volume to Retail Dollar Volume				
Sorting Variable	low	medium	high	*HML*	*t-stat*
Small/Retail Dollar Volume	0.71	0.71	0.67	-0.04	(-4.15)
		Small/Retail Dollar Volume			
Sorting Variables		low	high	*HML*	*t-stat*
Amihud's Illiquidity Ratio	low	0.76	0.72	-0.04	(-7.56)
	high	0.67	0.63	-0.04	(-4.36)
	HML	0.09	0.09		
	t-stat	(-11.10)	(-10.52)		

Table 2.12: Panel Results- Foreign Exchange Risk

Panel regressions with random effects for predictive regression of order imbalance for 41 country-level MSCI based iShares ETFs region on percentage change in volatility index (ΔVIX), percentage change in orthogonalised local volatility index ($\Delta LVIX^o$), exchange rate return (Δer_{t-1}), absolute value of exchange rate return ($|\Delta er_{t-1}|$), change in local interest rates (IR), lagged order imbalance (OI_{t-1}), dummy variable for a common language (L) and a distance to the US (G), as well as lags of ΔVIX and $\Delta LVIX^o$. Specifically, I estimate the model below:

$$OI_{i,t} = \alpha + \sum_{k=1}^{5} \beta_k \Delta VIX_{t-k} + \sum_{k=1}^{2} \gamma_k \Delta LVIX^o_{i,t-k} + \delta_1 \Delta er_{t-1} + \delta_2 |\Delta er_{t-1}| + \\ + \delta_3 \Delta IR_{i,t-1} + \delta_4 OI_{i,t-1} + \delta_5 L_i + \delta_6 G_i + \varepsilon_{i,t}$$

The number of lags in the model is determined using Akaike Information Criterion and Bayesian information criterion jointly. Coefficients and z-statistics (based on robust standard errors) are presented. The lag for predictive regression is 1 day. Data is at a daily frequency and covers the period from 2006 or the first trading day of ETF (whichever is later) until the end of June 2018. ***,**,* denote significance at the 1%, 5% and 10% level.

	(1)	(2)	(3)	(4)
VARIABLES	*TOI*	*smallOI*	*retailOI*	*largeOI*
ΔVIX_{t-1}	-0.176***	-0.113***	-0.249***	-0.235***
	(-7.08)	(-4.73)	(-9.71)	(-8.05)
$\Delta LVIX^o_{t-1}$	-0.039*	-0.032**	0.002	-0.037
	(-1.95)	(-2.01)	(0.09)	(-1.39)
Δer_{t-1}	-0.680***	-0.387**	-1.296***	-0.856**
	(-4.23)	(-2.27)	(-3.59)	(-2.50)
$\|\Delta er_{t-1}\|$	-1.020***	-1.559***	-2.017***	-0.663
	(-2.78)	(-4.76)	(-3.14)	(-1.63)
ΔIR_{t-1}	-0.004**	-0.002	-0.006**	-0.002
	(-2.28)	(-1.63)	(-2.02)	(-0.86)
OI_{t-1}	0.169***	0.193***	0.128***	0.113***
	(12.39)	(8.30)	(8.83)	(10.50)
L	0.002	-0.006***	0.012	0.001
	(0.33)	(-2.64)	(0.81)	(0.11)
G	-0.000	-0.000	-0.000**	0.000
	(-0.23)	(-0.93)	(-2.55)	(0.14)
ΔVIX_{t-2}	-0.027*	-0.023*	-0.095***	-0.040
	(-1.92)	(-1.80)	(-4.19)	(-1.55)
$\Delta LVIX^o_{t-2}$	-0.055**	-0.027	-0.054**	-0.072***
	(-2.32)	(-1.25)	(-2.03)	(-2.68)
ΔVIX_{t-3}	-0.053***	-0.024**	-0.057***	-0.060***
	(-3.13)	(-2.42)	(-2.70)	(-2.64)
ΔVIX_{t-4}	-0.071***	-0.056***	-0.081***	-0.085***
	(-4.65)	(-3.58)	(-4.89)	(-4.17)
ΔVIX_{t-5}	-0.042***	-0.028**	-0.023	-0.055**
	(-2.78)	(-2.07)	(-1.25)	(-2.22)
Constant	0.020***	0.011***	0.049***	0.013*
	(2.99)	(3.02)	(4.02)	(1.85)
Observations	57,683	52,578	54,455	55,346
Adjusted R^2	0.032	0.041	0.020	0.016

Table 2.13: Volatility Regimes

Panel regressions with random effects for predictive regression of order imbalance for 41 country-level MSCI based iShares ETFs on percentage change in volatility index (ΔVIX), percentage change in orthogonalised local volatility index ($\Delta LVIX$), change in local interest rates (IR), lagged order imbalance (OI_{t-1}), dummy variable for a common language (L) and a distance to the US (G), as well as lags of ΔVIX and $\Delta LVIX^o$. Specifically, I estimate the model below:

$$OI_{i,t} = \alpha + \sum_{k=1}^{5} \beta_k \Delta VIX_{t-k} + \sum_{k=1}^{2} \gamma_k \Delta LVIX^o_{i,t-k} + \delta_1 \Delta IR_{i,t-1} + \delta_2 OI_{i,t-1} + \delta_3 L_i + \delta_4 G_i + \varepsilon_{i,t}$$

The number of lags in the model is determined using Akaike Information Criterion and Bayesian information criterion jointly. Coefficients and z-statistics (based on robust standard errors) are presented. The lag for predictive regression is 1 day. Data is at a daily frequency and covers the period from 2006 or the first trading day of ETF (whichever is later) until the end of June 2018. The data sample is split into 3 sub-samples: period of low VIX ,period of medium VIX, and period of high VIX. The split is based on terciles of of historic VIX level distribution from Jan 1990 - June 2018. ***,**,* denote significance at the 1%, 5% and 10% level.

	Low				Medium				High			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
VARIABLES	*TOI*	*smallOI*	*retailOI*	*largeOI*	*TOI*	*smallOI*	*retailOI*	*largeOI*	*TOI*	*smallOI*	*retailOI*	*largeOI*
ΔVIX_{t-1}	-0.155***	-0.016	-0.230***	-0.262***	-0.197***	-0.121***	-0.215***	-0.273***	-0.117***	-0.095***	-0.207***	-0.134***
	(-5.55)	(-0.57)	(-3.71)	(-5.25)	(-5.34)	(-3.04)	(-4.80)	(-5.90)	(-5.23)	(-4.14)	(-4.67)	(-4.71)
$\Delta LVIX^o_{t-1}$	-0.022	-0.043**	-0.016	-0.035	-0.105**	-0.028	-0.045	-0.098*	-0.030	-0.049	0.009	0.004
	(-0.94)	(-2.17)	(-0.57)	(-0.92)	(-2.02)	(-0.61)	(-0.72)	(-1.67)	(-1.01)	(-1.63)	(0.15)	(0.09)
ΔIR_{t-1}	-0.002	-0.002	-0.006	0.001	-0.005	-0.006	0.008	0.007	-0.005*	-0.001	-0.004	-0.007*
	(-0.65)	(-0.38)	(-0.97)	(0.29)	(-1.12)	(-1.50)	(0.86)	(0.86)	(-1.66)	(-0.27)	(-0.86)	(-1.75)
OI	0.195***	0.228***	0.099***	0.122***	0.125***	0.138***	0.131***	0.092***	0.179***	0.209***	0.150***	0.112***
	(10.01)	(6.98)	(8.03)	(6.66)	(7.14)	(6.94)	(6.12)	(5.39)	(6.83)	(6.15)	(6.22)	(11.43)
L	-0.012	-0.026***	-0.007	-0.011	-0.002	-0.004	0.010	-0.007	0.027***	0.017**	0.038**	0.028***
	(-0.88)	(-2.72)	(-0.32)	(-0.75)	(-0.20)	(-0.69)	(0.59)	(-0.68)	(2.69)	(2.12)	(2.11)	(2.93)

Continue on the next page

Table continued from previous page

	Low				Medium				High			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
VARIABLES	*TOI*	*smallOI*	*retailOI*	*largeOI*	*TOI*	*smallOI*	*retailOI*	*largeOI*	*TOI*	*smallOI*	*retailOI*	*largeOI*
G	-0.000	-0.000	-0.000	-0.000	0.000	-0.000	-0.000*	0.000	0.000	0.000	-0.000*	0.000
	(-0.99)	(-1.38)	(-1.46)	(-0.95)	(0.02)	(-0.15)	(-1.79)	(0.53)	(0.96)	(1.36)	(-1.86)	(1.34)
ΔVIX_{t-2}	-0.016	0.024	-0.066*	-0.057	0.021	-0.019	-0.095*	0.031	-0.001	0.016	-0.028	-0.015
	(-0.49)	(0.81)	(-1.70)	(-1.21)	(0.58)	(-0.65)	(-1.83)	(0.70)	(-0.03)	(0.76)	(-0.79)	(-0.41)
$\Delta LVIX^{o}_{t-2}$	-0.043	0.016	-0.054	-0.074**	-0.098**	-0.109*	-0.077	-0.116**	-0.054*	-0.059**	-0.066	-0.048
	(-1.47)	(0.81)	(-1.05)	(-2.03)	(-1.99)	(-1.94)	(-1.04)	(-2.05)	(-1.66)	(-2.18)	(-1.26)	(-1.26)
ΔVIX_{t-3}	-0.006	0.026	0.100*	-0.006	-0.072*	-0.026	-0.028	-0.080*	-0.010	0.030	-0.052*	-0.030
	(-0.17)	(0.94)	(1.79)	(-0.15)	(-1.90)	(-0.86)	(-0.62)	(-1.71)	(-0.56)	(1.32)	(-1.68)	(-0.92)
ΔVIX_{t-4}	-0.003	0.019	0.101**	-0.031	-0.083***	-0.072**	-0.073**	-0.090***	-0.060**	-0.043**	-0.110***	-0.086***
	(-0.10)	(0.64)	(2.09)	(-0.84)	(-3.50)	(-2.25)	(-2.39)	(-2.89)	(-2.56)	(-2.11)	(-3.43)	(-2.64)
ΔVIX_{t-5}	-0.026	0.011	0.050	-0.006	-0.007	-0.021	-0.005	-0.027	-0.033	0.008	0.014	-0.055
	(-0.84)	(0.40)	(0.85)	(-0.16)	(-0.23)	(-0.73)	(-0.11)	(-0.68)	(-1.18)	(0.38)	(0.41)	(-1.56)
Constant	0.053***	0.042***	0.081***	0.052***	0.017*	0.005	0.033***	0.008	-0.031***	-0.049***	-0.006	-0.035***
	(3.29)	(3.37)	(4.09)	(3.15)	(1.92)	(0.65)	(2.68)	(0.96)	(-2.82)	(-4.29)	(-0.41)	(-3.30)
Observations	21,407	20,187	19,576	20,680	15,664	13,840	14,838	14,750	14,652	13,109	14,324	14,193
Adjusted R^2	0.032	0.042	0.010	0.015	0.019	0.021	0.019	0.013	0.032	0.043	0.025	0.014

Chapter 3

ETF Short Interest and Distressed Stocks

3.1 Introduction

With the rise of assets under management of exchange traded funds (ETFs) significant attention is devoted to studying the effect of ETFs on the underlying stocks. Yet, the majority of previous research focuses on the common effects of ETFs on every stock in the holding basket, regardless of the individual characteristics of such stocks. In particular, many studies investigate whether ETFs bring noise to underlying securities (e.g. Ben-David et al., 2018). At the same time, the growth of coverage of ETF market led to inclusion of securities of heterogeneous types in the portfolios. Therefore, it is important to consider differences in stocks in ETF basket when assessing potential impact of such investment vehicles on equity markets. In this paper, I consider the credit quality as a characteristic that varies among ETF holdings and measures the impact of ETF trading on stocks.

I investigate the effect of industry ETF membership on stocks that are in financial distress in contrast to non-distressed stocks. Industry ETFs (also known as sector ETFs) is a growing product with $385 billion dollars of assets under management in the U.S. alone as of June 2019[1]. For this study I consider industry

[1] According to ICI Global: https://www.ici.org/research/stats/etf/etfs_06_19

ETF as a fund that tracks the performance of one sector of U.S. economy (e.g. iShares U.S. Real Estate ETF) via physical (non-synthetic) replication. Such type of ETFs has a large market coverage and is widely available. It is also an ETF product with one of the highest levels of short-selling (Huang, O'Hara, and Zhong, 2018). I investigate the role of industry ETF short-selling on the underlying stocks and, in contrast to previous studies, show that distressed securities are positively affected by industry ETF membership in a form of improved price efficiency.

I also study stocks that are part of industry ETF baskets and find that they do not show signs of financial distress anomaly in contrast to non-ETF-member stocks. The financial distress anomaly is one of the widely investigated puzzles in the asset-pricing literature. Stocks that experience financial distress are more risky. Apart from bankruptcy risk (and numerous costs associated with it) such stocks tend to have larger market betas, have smaller market capitalization, have larger volatility, leverage and lower liquidity. Despite having a higher risk than other less distressed firms they under-perform relative to the common benchmarks (e.g Campbell, Hilscher, and Szilagyi, 2008; Dichev, 1998). Such high risk and low return relationship is not consistent with rational investment behaviour of market participants. Although there is no clear consensus within the current research on the potential explanations of distress anomaly, strong evidence suggest that distressed stocks tend to be overpriced due to the existence of higher barriers for short-selling (e.g. higher fees, low institutional ownership and therefore availability etc.). Campbell et al. (2008) find that distressed stocks have higher loadings on market, SMB and HML factors. They show that stocks with higher limits to arbitrage tend to exhibit a stronger anomaly. The profit and loss of the long-short strategy on the non-distressed and distressed stocks is correlated with VIX, suggesting that investors sell stocks with high probability of failure when facing an increase in market uncertainty. Gao, Parsons, and Shen (2017) support the behavioural explanation based on limits to arbitrage suggesting that when it is hard for institutional investors to correct the mispricing negative news might not be reflected in the price immediately. The slow incorporation of information results in negative returns consistent with under-performance of distressed stocks. Specifically, they show that high overconfidence among investors results in significantly negative returns of distressed stocks. Following the period of high market gains the returns of portfolio of distressed stocks is low. Finally, they show that within the distressed group returns are particularly low for firms that have experienced recent bad news. Avramov, Chordia, Jostova, and Philipov (2013) show that profitability of many anomalies such as, for example, earnings and price momentum is concentrated in distressed stocks. Once stocks with low credit ratings are removed from the sample the anomalies disappear. Moreover, it is the short side of the strategy that generates the positive return. Such finding supports the distressed stock mispricing explanation. Stambaugh, Yu, and Yuan (2012) link investor sentiments to profitability of the short-leg of the long-short distressed strategy. They show that distressed anomaly is stronger following the period of high sentiments and the short-leg experiences a significantly lower return. Interestingly, the effect of sentiments on the long-leg is much weaker. Such asymmetric effect of sentiments is consistent with anomaly reflecting the mispricing.

Motivated by the mispricing explanation of the anomaly and the existence of high barriers for the direct short-selling of distressed stocks I investigate whether investors can reduce the overpricing of such stocks by obtaining a short position indirectly via ETFs. Karmaziene and Sokolovski (2015) as well as Li and Zhu (2018) describe the create-to-lend mechanism that allows the authorised participants (APs)[2] to easily expand the supply of ETF shares available to borrow. The transmission mechanism of ETF shocks to the underlying stocks is widely described in the ETF literature (e.g Ben-David et al., 2018; Da and Shive, 2018, etc.). The sufficient short-selling of industry ETFs in the secondary market leads to the ETF price declining

[2]Authorised Participant is a dealer in the ETF market who has an agreement with ETF sponsor for creation/redemption of ETF shares. To create new shares he can deliver the basket of underlying securities to ETF sponsor and exchange them for newly created shares. For redemption he can deliver the ETF shares and exchange them for underlying stocks.

below the net asset value of the fund. Observing the discount (or experiencing a positive inventory shock if the AP is also a market maker), the AP has the incentive to close the arbitrage opportunity by purchasing the ETF shares (or using ETF stocks from his inventory) and delivering them to the ETF sponsor in exchange for the underlying basket of securities. Simultaneously, underlying stocks are sold-short by the AP creating a negative price pressure. The short-selling is much easier for the AP as he has a certainty of receiving the underlying securities locked in the ETF basket. In addition, the short-sale of the ETF should be a strong signal to the market participants about the value of the underlying stocks. Over time such signal is reflected in negative returns in the underlying market (Li and Zhu, 2018). I find that 9.11% of the companies in the baskets of industry ETFs are in distress. Such percentage is high because many passive funds are forced to keep distressed stocks as they are a part of broad sector benchmarks that industry funds replicate. I first test the effect of U.S. industry ETF short-selling on returns of distressed and non-distressed stocks next month. I find that while ETF short-selling has a very limited effect on non-distressed stocks (consistent with absence of overpricing in these stocks) the selling pressure results in negative returns of distressed stocks during the following month. Such response is consistent with the existence of risk-averse APs that propagate ETF shocks to the underlying markets. I also investigate the effect of differences in the short-selling motives of industry ETFs on underlying securities. Huang et al. (2018) argue that industry ETFs provide a new hedging tool to active investors (e.g. hedge funds) to hedge their industry risk when making risky bets on underlying stocks. An informed investor would utilise his positive private information about a stock by taking a long position in it and simultaneously hedging his industry exposure by going short on a corresponding industry ETF. Such motive is dominant during non-crisis period. The alternative reason to short-sell industry ETF is to make a negative bet on the future returns of the industry (speculative motive). Huang et al. (2018) show that such reason for short-selling is the main one during the crisis period. Since short-selling for hedging purposes does not contain fundamental information about the industry such activity should result in return reversal of underlying stocks next month. I test whether such explanation is consistent in the case of distressed stocks by splitting my sample into crisis and non-crisis periods. I find that during the crisis period fundamental short-selling creates an equal downwards pressure on both types of stocks. During the period of hedging-motivated short-selling I observe a reversal for non-distressed stocks. In contrast, even when the initial motive for short-selling is non-fundamental the distressed stocks react negatively in line with an overvaluation hypothesis proposed by a number of studies. Overall, such results show that irrespective of short-selling motives industry ETFs partially reduce overpricing of distressed securities.

In contrast to Huang et al. (2018) I demonstrate that high short-selling of industry ETFs is not uniform across all sectors of the economy. The negative returns of the underlying stocks after indirect short-selling is concentrated in the cyclical industries. Such evidence is consistent with high proportion of distressed stocks concentrating in high beta (more volatile) sectors. Following the findings of Stambaugh et al. (2012) I also test whether ETF short-selling is driven by the existence of overpricing in the market. I show that on average the largest amount of short-selling is concentrated in the month following the period of high investor sentiments proxied by Baker and Wurgler (2006) measure and by the Michigan sentiment index. Finally, I test whether the alleviation of the short-selling constraint embedded in the distressed stocks translates into a higher price efficiency of distressed stocks. Consistent with previous studies I find that ETF short-selling increases the price delay for non-distressed stocks (negative effect on efficiency). However, the effect is completely opposite for distressed stocks, which highlights the positive effect of industry ETFs on the price efficiency of their distressed member stocks. I also investigate whether the ETF effect on efficiency remains in the presence of alternative short-selling mechanisms - stock options. I find that although the net effect is lower the result is robust when controlling for availability of call and put options.

I investigate the financial distress anomaly among stocks with low credit ratings conditional on ETF membership. I first sort stocks based on S&P long-term credit rating and then based on ETF membership. I find that since ETFs reduce the overpricing of distressed stocks, securities that are members of industry ETFs do not under-perform common benchmarks in contrast to non-member stocks. This effect is robust to different proxies of financial distress. Such finding provides a novel evidence of ETFs improving general market efficiency in contrast to numerous previous studies that show the opposite. Garlappi, Shu, and Yan (2008) and Garlappi and Yan (2011) suggest that one of the potential explanations of distress anomaly is that when default is almost certain the distress risk decreases as equity investors can expropriate value from debt holders. Gao et al. (2017) do not find the support for such hypothesis internationally. I show that my findings are not driven by such explanation either.

This paper is related to the literature that links pricing efficiency to the short-selling constraints at the individual stock level. Drechsler and Drechsler (2016) demonstrate the connection between high short-selling fees and profitability of anomaly returns. They show that many anomalies (including the distress anomaly) only exist in the high short-selling fee group. Previous research shows that short-selling constraints (Nagel, 2005, etc.), as well as short-selling risk in the form of uncertainty about future short-selling fees (Engelberg, Reed, and Ringgenberg, 2018) leads to a reduction in price efficiency of stocks. Saffi and Sigurdsson (2010) show that stocks with higher constraints have slower reaction to market developments. My work is also related to the literature discussing the value relevant information contained in the short-selling. Boehmer, Jones, and Zhang (2008) show that short-sellers are well-informed and able to predict negative returns at the stock level. Engelberg, Reed, and Ringgenberg (2012) provide evidence that such informativeness comes from the ability to better process public information. Boehmer, Huszár, Wang, and Zhang (2018) find that short-selling predictability varies internationally and is concentrated in countries with less severe short-selling regulations and better market quality. Rapach, Ringgenberg, and Zhou (2016) show that the short-interest ratio at the aggregate level is a better predictor of U.S. returns than many other known variables. Huszár, Tan, and Zhang (2017) demonstrate that short-sellers target specific industries with higher information complexity. Huszár, Tan, and Zhang (2019) show that industry level short-selling contains a better (more economically important) information about future returns than at the stock level due to the presence of binding short-selling constraints. Consistent with these findings I use industry ETFs to assess the alternative short-selling venues. Beneish, Lee, and Nichols (2015) show that the predictive ability of short-interest ratio is concentrated in stocks with binding constraints. Similarly, Guo and Wu (2019) show that predictive ability of short-interest ratio is concentrated in the low-rated stocks. This highlights the reason why in this study I use credit quality as a defining characteristic when evaluating the impact of ETF short-selling on stocks. This paper is also related to the literature on ETF short-selling. Karmaziene and Sokolovski (2015) demonstrate that during short-sale ban of financial stocks in 2008 investors moved to financial sector ETFs to alleviate the constraint. Li and Zhu (2018) show that stocks with higher short-selling constraints at the stock level and higher short-interest ratio at the ETF level experience lower returns. Overall, evidence suggest that investors use ETFs as an alternative short-selling route when the direct short-selling is complicated. My results support such findings. However, previous studies do not consider the effect of such ETF short-selling on the distressed stocks. Sorting stocks by credit rating allows me to highlight the role of ETFs, as an indirect short-selling mechanism, in improving price efficiency of distressed stocks.

Numerous studies show the negative effects of ETFs on underling stocks from different dimensions. Israeli et al. (2017) argue that ETFs increase trading costs of underlying stocks and reduce their information efficiency. Da and Shive (2018) shows that ETFs increase the co-movement of underlying stocks. Ben-David et al. (2018) show that ETFs propagate liquidity shocks to underlying securities leading to a rise of non-fundamental volatility. In contrast to such studies, by recognising that ETF baskets include stocks of heterogeneous types (rather than considering the joint effect on all stocks) I show positive impacts of ETFs in restoring market efficiency. Opposite to the literature that focuses on the noise propagation the research of Lettau and Madhavan (2018) as well as Madhavan and Sobczyk (2016) highlights the role of ETFs in improving the price discovery at the stock level. This paper contributes to this stream of literature by demonstrating the improvement in price discovery of distressed stocks. Evans, Karakaş, Moussawi, and Young (2019) show that the short-selling of the ETFs creates the so-called "phantom ETF" and "phantom underlying securities". ETF stocks carry the voting rights (executed by the ETF sponsor) as they are backed by the portfolio of stocks in the ETF basket. In contrast, ETF stocks that are borrowed and then short-sold are backed by the collateral held by the broker (may be a hedging instrument, rather than the original securities) and due to broker limitation on voting do not carry voting rights. Authors find that firms with high amount of "phantom ETF shares" tend to underperform due to a deterioration of firm's governance. In addition due to many underlying shares being locked up in the broker account as collateral the value of votes increases. Kalay, Karakaş, and Pant (2014) discuss the increase in lending fees for stocks around special shareholder meetings and consequently a positive correlation with voting premium. In addition, the availability of voting rights is likely to be reduced by the ETF phantom effect. While this paper does not directly study the voting premium it is likely to be reflected in the higher daily cost of borrowing score (DCBS) for ETF-member security. Interestingly, in my sample I do not find the statistically significant difference between DCBS of ETF member and non-member distressed securities, which suggests that other factors such as liquidity and lending supply might out-weight the voting premium effect. In addition, I find that in contrast to Evans et al. (2019) ETF-member distressed stocks perform better than non-ETF stocks relative to a common risk benchmark (zero vs negative alpha). Such finding suggests that for distressed securities, on a risk adjusted bases, the reduction in mispricing has a stronger effect than the deterioration of company governance.

Section 3.2 describes the construction of stock and industry ETF sample, as well as procedure for sorting stocks by credit quality. Section 3.3 provides the empirical results. Section 3.4 shows the results of robustness tests. Section 3.5 concludes.

3.2 Data and Methodology

In this section I describe my sample construction including the procedures to identify industry ETFs. I then provide the methodology to clean the universe of common stocks used in this study and to construct the short-interest ratio for each of such stocks, as well as for each industry ETF. Finally, I explain different methodologies for constructing the financial distress measure and for obtaining the ETF holdings.

3.2.1 Industry ETFs

In order to construct the list of industry ETFs I first obtain all plain vanilla equity ETFs that are traded on one of the U.S. exchanges. I start with the universe of stocks from Center for Research in Security Prices (CRSP) database. I filter for ETFs by only keeping stocks with share code of 73. I combine this list with ETFs from Compustat database that, as in Israeli et al. (2017), are identified using issue type of '%'. Using CRSP Survivor-biased-free Mutual Fund database I identify equity ETFs from my list by keeping only those funds that have Lipper Asset Code 'EQ'. I follow Ben-David et al. (2018) and filter funds by Lipper Objective Codes to remove ETFs with synthetic replication, as well as leveraged or active products. Finally, to ensure that I obtain the funds that invest in equities I check that they have at least 80% of AUM invested in common stocks.

I follow Huang et al. (2018) to identify industry ETFs. I obtain holdings for each equity ETF from Thomson-Reuters Mutual Fund holding database (S12). For every equity that funds have in their holdings I obtain Standard Industry Classification codes (SIC) from Compustat. The assigned SIC codes are as of the fiscal year ending in a calendar year $t-1$. I use 12 industry classifications from

Kenneth French's website to allocate ETFs to industry groups (see table 3.1 for the description of each industry). Industry ETF is defined as a fund that has at least 30% of assets invested in the dominating industry and has at least 30 stocks in holdings in general. I check the obtained list and manually filter the ETFs that do not intend to invest in any particular industry. For example, I remove ETFs that invest in broad indices such as Russell 3000 and S&P 500. Finally, I ensure that 80% of AUM are invested in U.S common domestic stocks. For example, this results in iShares Global Tech ETF (IXN) being excluded from the list and iShares North America Tech ETF (IGM) remaining on the list. Overall, the final list consists of 127 ETFs, which is similar to 121 funds identified in Huang et al. (2018). The common feature of almost all funds in the final sample is that their names contain clearly identifiable industries (e.g. Vanguard Consumer Staples ETF or iShares U.S. Basic Materials ETF). The sample starts in January 1999 and ends in February 2017 due to ETF and S&P rating data availability.

[insert Table 3.1 here]

An obvious concern is that if distress risk is concentrated among small stocks then they are likely to be excluded from ETF benchmarks. However, very often this is not the case. Many of the funds in my sample track benchmarks constructed by MSCI or S&P. MSCI constructs sector indices based on "Investable Market Index". This index achieves 99% of the coverage of the market. Figure 3.1 shows that Investable Market Index consists of Standard Index (Large Cap and Mid Cap indices) and a Small Cap index[3]. For example, Vanguard Energy ETF (VDE) tracks MSCI U.S. Investable Market Energy 25/50 Index that includes large, medium and small companies in energy sector. Blackrock's iShares U.S. Consumer Goods ETF (IYK) tracks Dow Jones U.S. Consumer Goods Index. Dow Jones index covers 95% of U.S. market capitalization. Overall, while some funds track a narrowly defined indices that include only firms with large or medium capitalisation, it is relatively

[3]Based on MSCI Global Investable Market Indexes Methodology available at MSCI.com

common for index funds to include small stocks. I find that on average 9.11% of the number of companies that are in ETF benchmark are in financial distress. Moreover, concentration rules such as "25/50" limit the maximum weight per stock and the total weight of large stocks in the fund. Furthermore, although most of financially distressed firms are concentrated in smaller size companies the distress is not exclusive to this category.

[insert Figure 3.1 here]

In this study I use industry ETFs because such funds are the most shorted by investors among equity ETFs. Figure 3.2 shows the list of top 15 most shorted ETFs as of 24 July, 2018. Out of 15 funds 7 can be classified as industry ETFs. In addition, this type of funds is more widely available than vanilla funds that exclusively follow small stocks (9 funds as of February 2019[4]).

[insert Figure 3.2 here]

3.2.2 Short-Interest

I am interested in the effect of indirect short-selling via ETFs on the underlying stocks. I obtain monthly short interest for each ETF in my sample from Compustat. The short-interest ratio (SIR) is defined as the amount of shares sold short scaled by the number of shares outstanding:

$$SIR_t = \frac{SharesSoldShort_t}{SharesOustanding_t} \tag{3.1}$$

I use the number of shares outstanding from CRSP database, as it is more complete and has less severe errors for this variable. I compute short interest ratio for ETFs ($SIR_{j,t}$, where j is one of 127 industry ETFs), as well as at the stock level ($SIR_t^{stock_i}$) to control for the direct short-selling channel. In this study I use short interest ratio at the monthly frequency. Sometimes, the SIR can be above 100%, which means

[4]According to ETF.com

that ETF shares were re-borrowed multiple times. Similarly to other literature on the short-interest ratio (e.g. Huang et al., 2018), I replace SIR greater than 100% with 100%.

3.2.3 Universe of Common Stocks and ETF holdings

I obtain daily prices for all U.S. stocks traded on major exchanges from CRSP daily stock file. I only include common stocks with share class '11' and '12' that are listed on NYSE, AMEX or NASDAQ (exchange codes '1', '2' or '3'). Stocks with prices of less than $5 at the end of the month are removed. When working with distressed stocks it is important to account for delistings. I adjust the returns by including delisting returns based on methodology of Shumway and Warther (1999). Overall, the sample of common stocks contains 10,595 firms from 1999-2017. I also calculate numerous stock related control variables from Compustat, CRSP and IBES databases. See table C.1 in the appendix for more details.

I obtain quarterly ETF holdings for each industry ETF in my sample from S12 Thomson Reuters Mutual Fund Holdings database. The reporting quarters are not always aligned across funds. I assume that the constituents of ETF basket are constant for the reporting quarter (from previous holdings report date or fund inception date until the current report date). Such assumption is realistic, as major index providers usually rebalance their indices once a quarter.

3.2.4 Measure of Financial Distress

In this study I evaluate the difference between the effect of ETF short-selling on the distressed and non-distressed stocks. Previous research introduced many different proxies for financial distress: Z-scores (Altman, 1968), O-scores (Ohlson, 1980), Moody-KMV's expected default frequency (Gao et al., 2017; Garlappi et al., 2008) etc.. I follow Avramov, Chordia, Jostova, and Philipov (2009) and Avramov et al. (2013) and use domestic long-term issuer credit ratings provided by S&P as a proxy for financial distress. This measure shows the capacity and willingness of a company to meet its long-term financial obligations. The main advantage of this dataset is that it is readily available from Compustat database at a monthly frequency. In contrast, alternative measures are often constructed from annual or quarterly accounting data that limits the frequency of the analysis. S&P assigns ratings from AAA (extremely strong capacity to meet obligations) to SD (selective default). Similarly to Avramov et al. (2009, 2013), I exclude SD category from the analysis and assign numeric scores to each rating (AAA = 1 ... BBB-=10 ... D=22). Table C.2 in the appendix shows the distribution of stocks by credit scores. Every rating category below 10 is considered as non-investment grade. Based on terciles of monthly credit score distributions I split stocks into 3 groups: Distressed Stocks (the highest scores), Medium Rating and High Rating stocks (the lowest scores). The allocation is performed monthly. After removing stocks without credit rating my sample contains 2721 unique stocks.

In the robustness tests, I use Altman's Z-scores (Altman, 1968) and the proxy for Moody's KMV distance to default measure that is based on the model of Merton (1974). The Altman's Z-score combines 5 accounting ratios:

$$Z = 1.2X_1 + 1.4X_2 + 3.3X_3 + 0.6X_4 + 0.99X_5 \tag{3.2}$$

where X_1 is the ratio of Working Capital to Total assets, X_2 is the ratio of Retained Earnings to Total assets, X_3 is the ratio of EBIT to Total assets, X_4 is the ratio of Market Value of Equity to Book Value of Total Debt and X_5 is the ratio of Sales to Total Assets. Typically, companies with z-scores above 2.99 are considered as safe. Group of companies with z-scores between 1.81 and 2.99 are in a "grey zone" and have the highest proportion of error classification. Firms with z-scores below 1.81 are considered as distressed. In order to obtain a higher frequency data, I compute Altman Z-scores quarterly.

I follow Bharath and Shumway (2008) to construct the distance to default measure. While not identical to Moody's KMV measure authors estimate that the correlation is high reaching around 79%. This measure is based on the representation of the value of equity as a call option on company's assets with a strike price that is equal to the face value of debt. The value of the distance to default measure can be estimated as follows:

$$DD = \frac{\ln \frac{V}{F} + (\mu - 0.5\sigma_V^2)T}{\sigma_V \sqrt{T}} \tag{3.3}$$

where V is the value of firm's assets, F is the face value of firm's debt issued with maturity T, μ is a continuously compounded rate of return and σ_v is the volatility of the total value of the firm. I use the expected default frequency (EDF) as a sorting variable to proxy the probability of default:

$$EDF = \mathcal{N}(-DD) \tag{3.4}$$

In order to estimate EDF for each firm in my sample I proxy the risk-free rate with a 1-year Treasury Constant Maturity Rate obtained from Federal Reserve Bank Reports on WRDS. The value of debt is estimated as a sum of the values of short-term liabilities and one-half of long-term liabilities available from Compustat. Appendix C.1 provides the detailed description of how the EDF measure is computed numerically.

3.3 Empirical Results

In this section I evaluate the effect of ETF short-selling on the returns of distressed and non-distressed stocks. I highlight the differences in effects conditional on industry sensitivity to business cycles and on short-selling motives. I further investigate the implications of industry ETF membership on the existence of the distress anomaly and demonstrate the asymmetric impact of ETF short-selling on the price efficiency of distressed stocks.

Panel A of table 3.2 shows the summary statistics of ETFs across 11 industries. None of the ETFs in my sample track the performance of Consumer Durables industry[5]. The highest number of ETFs are in the Financial Sector and Business Equipment (that includes the IT sector) industries. Most of the funds have a high SIR on average. Retail and Financial sectors (Shop and Money) have the highest short interest (18.25% and 17.13%), as these are the typical industries that investors bet against during recession periods. Overall, the SIR tends to be higher in cyclical industries in line with a speculation motive. While Huang et al. (2018) argue that the value of SIR for industry ETFs is very high at the 95th percentile I show that the level of short-selling varies significantly across industries. The values are consistent with the average SIR ranking and are relatively low (in comparison to other industries) for Non-Durables and Telecommunication sectors (20.19% and 22.35%). This suggests that these industries are less likely to experience extreme levels of short-selling often observed in the Financial Services industry. Panel B of table 3.2 shows the summary statistics of stocks in my sample aggregated based on credit rating and ETF membership. On average, all distressed stocks have a credit rating of 14.04 (approximately B+), medium rating stocks have a score of 10.77 (BB+) and high rating stocks' score is 6.98 (A-). Distressed stocks that are members of ETF basket have almost identical credit scores[6] to non-ETF members, suggesting that it is not only the least distressed stocks that are members of ETFs. This is important as it shows that the results presented in this paper are not driven by relatively less distressed conditions among distressed stocks in ETF baskets. I further test this in the Robustness section and show that this is indeed not the case. Consistent with previous studies I find that distressed stocks have the lowest price and have the smallest market capitalisation. When comparing within the distressed category ETF stocks are larger on average than non-ETF ones. This is consistent with the top-down approach based on capitalisation that is often used for index construction. In addition, ETF stocks are more liquid. Finally, I present the average Daily Cost of Borrowing Score (DCBS) obtained from Markit that measures how difficult it is

[5]This is consistent with 11 industries identified in Huang et al. (2018)

[6] The difference in average ratings is 0.093 and t-statistics of the mean comparison test is t=0.91

to short-sell a stock. Similarly to previous literature the results suggest that it is harder to short-sell stocks in the distressed category. In contrast to medium and high rating stocks both ETF and non-ETF distressed groups have DCBS above 2 at the 95th percentile suggesting that these groups contain a subset of stocks with binding short-sale constraints that are not present for stocks with higher ratings.

[*insert Table 3.2 here*]

3.3.1 Stock Returns and ETF Short-Selling

I investigate how the ETF short-selling, as an alternative route to direct stock short-selling, affects the returns of basket constituents. I perform a predictive panel regression of monthly stock returns on SIR of ETF as of previous month:

$$\begin{aligned} r_{i,t+1} = \alpha + \beta_1 sir_{j,t} + \beta_2(sir_{j,t} \times Distress_{i,t}) + \beta_3 Distress_{i,t} + \beta_4 sir_t^{stock_i} + \\ + \beta_5(sir^{stock_i} \times Distress_{i,t}) + \mu Controls_t + FE_i + FE_t + \varepsilon_{i,t} \end{aligned} \tag{3.5}$$

where r_{t+1} is the log return of stock i at month $t+1$ and sir is the log of SIR of ETF j that holds stock i at time t. In order to investigate if going short on an industry ETF has a different effect on distressed stocks than on financially stable stocks I include the interaction variable of sir with a dummy variable $Distress_{i,t}$ that takes the value of 1 if stock i is in the distressed category (based on the definition of financial distress in 3.2.4). I include SIR at the stock level and its interaction with the distressed dummy to control for the direct short-selling channel. For other controls I include the standard set of variables used in the literature for return predictive regressions based on Huang et al. (2018) and Green et al. (2017): 12-month cumulative return, market capitalisation at the stock level, asset growth, book-to-market ratio, growth in the long-term net operating assets, gross profitability, investment growth, operating profitability, accruals and Amihud (2002) illiquidity measure. Every variable is lagged by one month or one quarter (depending on availability). Finally, I control for time and stock fixed effects, bootstrap and cluster standard errors by industry[7]. Table 3.3 shows the result of such regression. On its own, the ETF short interest does not affect stock returns next period. In contrast, once the full specification of regression is considered, the interaction term is highly significant and negative. This suggests that on average non-distressed stocks absorb ETF shocks and are immune to ETF short selling, as they are less likely to be overpriced. In contrast, as argued by previous studies distressed stocks have a higher short-selling constraint at the stock-level. As a result such stocks are more likely to be overpriced. Switching to ETFs helps to alleviate such constraint. The negative loading on the interaction term suggests that ETF short interest is followed by negative returns of distressed stocks next month, as distressed stock prices are pushed closer to a fundamental level. This effect is not driven by relative illiquidity of distressed stocks or by stock reversals, as I control for these channels in the regression.

[*insert Table 3.3 here*]

In the results above I use the adjusted version of SIR capped at 100%. In my sample SIR for ETFs above 100% occurs in only 3.04% of observations. In the unreported results I confirm that using the raw version of SIR is quantitatively very similar.

3.3.2 Short-Selling Motives: Hedging vs Speculation

The results above are agnostic about the reason for the short-selling. Since the net asset value (NAV) of the ETF co-moves with its price due to the existence of arbitrage mechanism the significant short-selling of the ETFs should create a negative price pressure in both markets. According to Ben-David et al. (2018) if such pressure is non-fundamental the price-impact is temporary and should be reversed next period. However, if the short-selling represents the bearish view of market participants we should not see such reversal.

[7]Clustering by industry is important as one firm may be included into multiple ETFs in the same industry. In order to ensure that the results are not affected by the small number of cluster groups, in unreported results, I cluster standard errors by ETF funds. I do not find any material differences in results

Huang et al. (2018) show that during the non-crisis period hedging motive is a dominant reason for ETF short-selling. In other words, investors short-sell industry ETFs mainly to hedge the industry risk of their long position in underlying stocks. In contrast, they found that speculative motive dominates during the financial crisis. During such period, investors make directional bets on the future negative performance of an industry. I evaluate the impact of such different motives for short-selling on distressed stocks. I run regression 3.5 conditional on the crisis period: excluding the period from quarter 4 of 2006 until the end of 2008 and only during the crisis period. Table 3.4 shows the results of such regressions.

[*insert Table 3.4 here*]

During the non-crisis period ETF short-selling is not based on fundamental information about the industry. In contrast to previous results, the loading on ETF short interest ratio $sir_{j,t}$ is positive and significant (although the significance is low). This is consistent with Huang et al. (2018) who show similar reaction using ETF returns as a dependent variable. The non-fundamental shock to stock prices is reversed next period. The reversal of the non-distressed stocks is in line with initial motive of sophisticated investors with positive private information about such stocks. The investors are less likely to bet on distressed stocks that have a lower probability of positive developments. The interaction term is negative and significant. This suggests that there is no reversal for distressed stocks (net effect is negative), which is in line with overpricing of this type of stocks. In other words, even when the short-selling motive is non-fundamental, the effect on distressed stocks is fundamental (although smaller in magnitude), as it allows to overcome the short-selling barriers at the stock level.

Huang et al. (2018) conducts the analysis excluding the crisis period. In contrast, I compare the findings by testing the effect during 2007-2008 financial crisis. During the crisis-period when ETF short-selling is more likely to contain fundamental information about a particular industry the loading on sir_j is negative and significant, while the interaction term is not significant. This is in line with speculative hypothesis, as both types of stocks are equally affected. The effect of short-selling is not reversed and is predictive. As argued by Ben-David et al. (2018) and Malamud (2016) the propagation of the negative shock from the ETF market is performed via APs. The negative shock in the stock market might come either from AP hedging his positive inventory (that arose after counterparty's short-sale) by going short in the stock market (if AP is also a market maker[8]) or due to AP closing the arbitrage position (that arose due to ETF shares being traded at discount after short-selling).

3.3.3 Cyclical vs Defensive Industries

In this section, I further test the effect of ETF short-selling conditional on the sensitivities of industries to business cycles. I use MSCI classification to allocate 11 industries considered in this study into Defensive and Cyclical types (see Panel A of table 3.2 for details of allocation). Companies in cyclical industries tend to be more sensitive to fluctuations in business cycle. Their earnings and market performance are more likely to be worse during recession. In contrast, firms in defensive industries are less correlated with the market. Distressed stocks tend to have higher market betas and therefore, I expect the majority of such stocks to be operating in cyclical industries.

[*insert Table 3.5 here*]

Table 3.5 shows the results of separating ETFs into 2 sub-samples based on such market sensitivity classification. I exclude *Other* industry from the analysis as this category combines cyclical and defensive industries. In addition, I exclude *Energy* industry, as although officially it is classified as defensive one, the performance of such industry is unstable. The results are robust to including *Energy* in the analysis. Similarly to previous results, the interaction term for cyclical industries is negative

[8] Although authorised participant is often a market maker in the ETF market it is not always the case

and significant, highlighting the concentration of overpriced distressed stocks in such industries. In contrast, the interaction term for defensive industries is insignificant as the loading on $sir_{j,t}$ is positive and significant reflecting the reversal of non-fundamental short-selling and a relatively lower number of distressed stocks in such industries. I show that the results are robust to the crisis effect. Overall, sorting by industry types highlights that the short-selling motives differ not only during crisis and non-crisis periods, but also depend on the sensitivity of stocks to business cycles, with major effect concentrated in the cyclical industries.

3.3.4 Overpricing, Investor Sentiments and ETF Short-Selling

The results above confirm that distressed stocks are more likely to be overpriced than non-distressed ones, that investors use industry ETFs to indirectly short-sell ETF constituents and that the motive for such short-selling depends on the time period (e.g. crisis) and on the sensitivity of stocks to business cycles. In this section, I provide additional pieces of evidence that ETFs are used to bet against the overpriced stocks.

Stambaugh et al. (2012) show that profitability of many anomalies depends on the degree of overpricing of the short-legs of strategies that exploit such anomalies and that the overpricing is the highest following the period of high investor sentiments. If ETF short-selling is used to bet against the overpriced stocks then we should see the increase in the SIR following months with high sentiments. I test this in table 3.6 at the aggregate level. I regress the log of cross-sectional average SIR of all ETFs ($avsir_{t+1}$) on Baker and Wurgler (2006) measure of investor sentiments (S_t)[9] as of previous month:

$$avsir_{t+1} = \alpha + \beta_1 S_t + avsir_t + \varepsilon_t \tag{3.6}$$

As can be seen from table 3.6 a period of high investor sentiments is followed by a larger average short-selling of ETFs. Such evidence links the ETF short-selling

[9] I thank the authors for providing the investor sentiments data on their websites

to overpricing hypothesis, suggesting that investors sell ETFs to bet against such mispricing. Following Stambaugh et al. (2012) I also use the Michigan consumer sentiments index[10], a measure based on the household survey, which, as in Baker and Wurgler (2006), I orthogonalise to variables related to macro-information: the growth in industrial production, durable, non-durable and service consumption, as well as to NBER recession indicator and growth in employment. As shown in table 3.6 the result is numerically similar to the one presented above.

[*insert Table 3.6 here*]

3.3.5 Distressed Anomaly and Price Efficiency

As discussed previously, due to high short-selling constraint for distressed stocks at the stock level investors cannot reflect their negative views about such stocks and they are likely to be overpriced. Over time, slow adjustment of prices downwards is reflected in the under-performance of such stocks relative to commonly used benchmarks (the distressed anomaly). I test whether the ETF short-selling, as an alternative route to obtain the negative exposure to distressed stocks, helps to improve price efficiency of such stocks by reducing the likelihood of over-pricing. If this is the case then those distressed stock that are members of ETF baskets are less likely to exhibit under-performance. I first compare the level of mispricing from asset-pricing models among stocks of different credit quality and then conduct a direct test of the impact of short-selling on the price efficiency proxies.

Every month I sort stocks into Distressed, Medium and High Rating categories as described in section 3.2.4. I further split stocks in Distressed category into those held by ETFs and others. Similarly to Avramov et al. (2013), I construct equally weighted portfolios based on credit sorts as of previous month. I then measure the performance of each of the 5 portfolios relative to different benchmarks. Table 3.7 shows the results of such exercise.

[10]Calculated by University of Michigan and is available from FRED database

I first check the result using Fama-French 3 factor model. Consistent with distressed anomaly literature Distressed-All group shows small under-performance of 3.6% per annum (although the result is not statistically strong). In contrast, Medium and High groups out-perform the benchmark. When I split distressed stocks based on ETF membership, Distressed-non-ETF group shows a much stronger (both statistically and in magnitude) under-performance of 8.4% per annum. In contrast, Distressed-ETF portfolio does not show any under-performance. This shows that by providing an alternative short-selling route industry ETFs reduce overpricing and improve price efficiency of distressed stocks. Across all groups, distressed stocks have a higher market beta that is gradually declining with lower credit scores (higher ratings). Interestingly, Distressed-ETF portfolio has a higher market beta than the non-ETF one. ETF membership amplifies the sensitivity of constituents to market movements. Based on SMB loadings stocks with lower credit ratings are smaller (higher loading). Unsurprisingly, Distressed-ETF stocks tend to be slightly larger than non-ETF ones. It is possible, that distressed stocks are past losers. Da and Gao (2010) show that outperformance of distressed stocks (instead of under-performance) demonstrated by Vassalou and Xing (2004) is due to short-term reversals. I extend Fama-French 3-factor model with short-term reversal factor obtained from Kenneth French website. The results are almost identical. Distressed stocks are less liquid in comparison to non-distressed stocks. Therefore, I also extend 3 factor model with Pástor and Stambaugh (2003) liquidity factor. Interestingly, alpha for Distressed-All group becomes insignificant. However, non-ETF alpha remains negative and significant. The Distressed-ETF result is robust to including the liquidity factor. Finally, I benchmark portfolio returns against Carhart (1997) 4-factor model that includes momentum. I do not find materially different results.

[*insert Table 3.7 here*]

Having established that industry ETFs reduce the likelihood of existence of the distressed anomaly the important question is whether stocks that are members of

ETFs incorporate negative fundamental information faster than non-ETF members and, therefore are more price efficient. I test if the ETF short-selling is a channel via which such negative information is reflected in stock prices.

I employ Price Delay measure (D) of Hou and Moskowitz (2005) to directly test the speed of information incorporation in stock prices. I regress the stock return ($R_{j,t}$) on the contemporaneous market return ($R_{m,t}$) and its four lags:

$$R_{j,t} = \alpha_j + \beta_j R_{m,t} + \sum_{n=1}^{4} \delta_j^{-n} R_{m,t-n} + \varepsilon_{j,t} \tag{3.7}$$

The 2 versions of the delay measure are then defined as follows:

$$D1 = 1 - \frac{R^2_{\delta_j^{-n}=0,\forall n \in [1,4]}}{R^2} \qquad D2 = \frac{|\sum_{n=1}^{4} \delta_j^{-n}|}{|\beta_j| + |\sum_{n=1}^{4} \delta_j^{-n}|} \tag{3.8}$$

$D1$ compares the ability of the reduced specification model (no lags) and full model in explaining stock returns. $D2$ captures the relative importance of longer over shorter lags in the regression. In both cases, a higher measure represents a longer delay. Following Boehmer and Wu (2012) I perform the regression 3.7 at a monthly level (using daily returns) and require at least 15 observations per month. I then perform the predictive panel regression of different versions of the price delay measure on the SIR at the ETF and stock level during previous month conditional on the credit quality of stocks:

$$\begin{aligned} d_{i,t+1} = \alpha + \beta_1 sir_{j,t} + \beta_2(sir_{j,t} \times Distress_{i,t}) + \beta_3 Distress_{i,t} + \beta_4 sir_t^{stock_i} + \\ + \beta_5(sir^{stock_i} \times Distress_{i,t}) + \mu Controls + FE_i + FE_t + \varepsilon_{i,t} \end{aligned} \tag{3.9}$$

where $d_{i,t+1}$ is a log of distressed measures $D1$ and $D2$. Following Boehmer and Wu (2012) the controls include (all lagged by one month) volume-weighted average price (VWAP), log of market value of equity, log of volume orthogonalised to log of market value of equity and the number of analysts covering the stock. Huang et al. (2018) show that industry ETF membership reduces the delay of member stocks, as using ETFs hedge funds are able to hedge industry risk exposure and therefore can more easily incorporate private information into the stock price. In contrast, since the intention is to measure the impact of ETF short-selling, I conduct the panel regression exclusively on the ETF member stocks.

[*insert Table 3.8 here*]

Table 3.8 shows the results of regression 3.9. The results confirm the finding of Ben-David et al. (2018) that ETFs bring noise to underlying stocks - unconditionally, as ETF short-selling increases the delay measure of stocks. However, once the interaction with distress dummy is considered the result is the opposite. Consistent with overpricing hypothesis and barriers for direct short-selling, higher industry ETF SIR reduces the delay of member stocks. Such results are in contrast to previous research that mainly focuses on the negative impact of the ETF market on the efficiency of underlying stocks. The split of ETF-member stocks into distressed and non-distressed categories allows to cast a new light on the benefits of ETFs for the stock market.

3.4 Robustness

In this section I perform the robustness tests to ensure that the previous findings are not driven by the choice of methodology and cannot be explained by alternative theories. I begin by providing the evidence that the results do not depend on the choice of the distress classification. I proceed by analysing the alternative explanation of the distressed anomaly within the ETF context and finally I investigate if the existence of other products, that can also provide the ability to obtain a short-position, dominates the ETF channel.

3.4.1 Alternative Distress Classification

In the previous results sorting to distressed category is based on the S&P long-term credit rating. In order to test if the results are robust to the choice of distress classification I repeat the analysis using Altman Z-score as a proxy for financial distress. The major drawback of such measure is that it relies on the accounting variables and therefore is only available quarterly. The allocation to distressed group is based on the lowest tercile of the distribution of quarterly scores. Consequently, I hold the portfolios for 3 months. I consider stock to be a member of ETF if it was a part of a fund basket in at least 1 month during the quarter. Table 3.9 shows the performance of such portfolios relative to different performance benchmarks.

[*insert Table 3.9 here*]

The results are consistent with previous monthly findings based on S&P credit ratings. As before, the under-performance of the distressed group is concentrated in the non-ETF-member stocks. In contrast, the distressed ETF-member stocks show a small over-performance although at a marginal significance level.

Moody's KMV model is also a popular choice in the literature to estimate the likelihood of default. As mentioned previously, I use a close proxy of this measure based on methodology of Bharath and Shumway (2008). In a similar fashion to the analysis above I allocate stocks into terciles based on EDF measure from distance-to-default model described in equation 3.4. Higher EDF values imply deeper financial distress. Table 3.10 shows the results of allocation based on EDF measure. In contrast to the non-ETF group, distressed members of ETF do not show underpricing and therefore do not display a distressed anomaly.

[*insert Table 3.10 here*]

Overall, using Altaman's Z-score and Moody's KMV model I show that my finding that distressed anomaly is not present in the ETF-member stocks is robust to the choice of the distress measure.

3.4.2 Shareholder Advantage Theory

The distressed anomaly is based on the idea that equity risk growths together with default risk. From risk trade-off perspective shareholders of a firm with a high probability of default need to be compensated with a higher stock return. Garlappi et al. (2008) and Garlappi and Yan (2011) argue that this is true only when there is an upward sloping relationship between default probability and equity risk. However, once shareholders' bargaining power at near default is considered such relationship is inversed. A firm with a high bargaining power may be able to renegotiate debt agreements and avoid liquidation costs. Therefore, for such firms a relatively higher probability of default risk is associated with lower equity risk and subsequently lower stock returns.

The difference between average credit ratings of distressed non-ETF and distressed ETF stocks is low (14.42 and 13.94 respectively). However, it may be possible that my results in the previous section are driven by a small sub-sample of non-ETF stocks that have a relatively higher degree of distress. In order to ensure that my findings do not capture the non-linear effect of shareholder recovery I first exclude stocks that have the lowest credit rating (highly distressed stocks - top decile of credit scores). I then split the remaining stocks into ETF and Non-ETF members and repeat the portfolio tests.

[*insert Table 3.11 here*]

Panel A shows alphas from different benchmark models. As can be seen from the table, distressed stocks that are members of ETFs do not show any signs of distress anomaly, while the group of non-ETF stocks has a negative alpha even after excluding the most risky securities. The under-performance of the non-ETF group is smaller relative to the results in table 3.7, which suggests that the hump-shaped relationship between distress risk and equity risk is present in the sample, but cannot explain my main findings.

In Panel B of table 3.11 I test this further and focus only on the highly distressed group and sort those stocks based on ETF and non-ETF membership. My previous results can be partially explained by the effect described in Garlappi et al. (2008) and Garlappi and Yan (2011) under two conditions: if the proportion of highly distressed stocks is larger in the non-ETF group and if sorting within highly distressed groups produces negative alphas independently on the ETF membership. Surprisingly, on average, there are more ETF-member unique stocks (38.12) than non-ETF-member stocks (24.49) per month in such a highly distressed group. The average rating of both sub-groups are similar with latter group having a marginally larger score (15.78 and 16.06 respectively). The average score of all stocks in a highly distressed group is 15.84 (between B- and B) - 6 scores below the investment grade threshold. The results show that consistently with previous findings under-performance is concentrated in the non-ETF group. The ETF group's alphas remain insignificant suggesting that the result is not driven by the effects described by the shareholder advantage theory.

According to the shareholder advantage theory the degree of hump-shaped relationship between equity beta and the probability of default is determined by the parameter that proxies for the recoverable fraction of asset value. Higher the value that can be extracted during the distress - lower the equity risk. Such variable is proportional to the shareholder bargaining power (if the recovery through renegotiation is possible) and the amount of liquidation costs (higher costs provide higher incentives to renegotiate). Garlappi and Yan (2011) show that the former measure can be captured by a firm's value of assets (small firms have a higher concentration of monitoring debt holders). Liquidation costs can be proxied by the degree of asset specificity. Following Garlappi and Yan (2011) I capture asset specificity with a measure of industry concentration proxied by the Herfindahl index (HI) of sales in an industry. HI for industry j at time t is defined as:

$$HI_{j,t} = \sum_{i=1}^{N_{j,t}} sales_{i,j}^2 \tag{3.10}$$

where $N_{j,t}$ is the number of stocks in the industry j at time t.

In panel C of table 3.11 I sort stocks independently into terciles[11] based on

[11] I move back to tercile distress sorting (instead of deciles) to ensure that corner portfolios contain the sufficient number of stocks. The limitation of this is that the double-sort includes all distressed

Distress and Asset Size, as well as Distress and Industry Concentration. Similarly, in panel D of table 3.11 I sort stocks based on Distress and Asset Specificity. For brevity I only show the results for high-high group sorting. Both sorts confirm that the results found in this paper are not driven by a non-linear relationship between distress and equity risks. Even for those firms with high shareholder recovery the ETF-member stocks do not display significant under-performance. Overall, the evidence above show that the finding of this paper is not driven by the non-linear relationship between equity risk and distress.

3.4.3 Alternative Short-Selling Mechanisms

The short-selling via ETFs is not the only way to obtain the negative exposure to stocks that are hard to bet against directly. Abhyankar, Filippou, Garcia-Ares, and Haykir (2019) study the effect of the existence of stock options on the ability of investors to short stocks and subsequently on price efficiency. They show that momentum strategy that is profitable due to its short leg can be partially explained by the stock optionality. Loser stocks with available options tend to be more informationally efficient than stocks without options. Investors can write a call option or buy a put option to obtain a negative exposure to a stock. Such feature of optionality provides a competing mechanism for indirect selling of securities.

In order to test if the ETF short-selling effect on distressed stocks survives once controlling for alternative short-selling avenues I perform regression 3.9 conditional on a call or a put option existing for stock i during month t.[12] Table 3.12 shows the results of such regression. As can be seen from the table, despite controlling for option availability, the interaction term of ETF short-interest ratio and the *Distress* dummy is still significant, although it is slightly smaller in comparison to previous findings. This suggest that ETF is an independent channel via which investors correct the mispricing in distressed stocks and improve price efficiency.

stocks and not just highly-distressed group

[12] I thank the authors for kindly providing the data on the existence of call and put options (based on OptionMetrics IvyDB US database)

[*insert Table 3.12 here*]

3.5 Conclusion

I investigate the effect of industry ETF short-selling on the underlying stocks in the fund basket conditional on the credit quality of such stocks. I find that 9.11% of companies in the benchmarks of this type of funds are distressed. I first show that the ETF short-selling helps to reduce the overpricing present in distressed stocks. I demonstrate that short-selling via ETFs alleviates the direct short-selling constraint and can predict negative returns of stocks in the lowest credit rating group. My findings show that in contrast to non-distressed stocks, for which the ETF short-selling is absorbed or reversed (apart from the financial crisis period), the effect on distressed stocks is always negative and fundamental (no reversal) irrespectively of short-selling motives. I find that the effect is concentrated mostly in the cyclical industries. Such finding is consistent with higher concentration of overpriced distressed stocks in such industries as well as with ETF investors betting against such mispricing.

Motivated by findings that distressed anomaly being caused by overpricing of distressed stocks I show that once sorted by ETF membership distressed firms that are a part of ETF portfolio do not show the under-performance relative to common benchmarks. Such result is robust to the choice of measure of credit quality and alternative explanations of distress anomaly. I provide new evidence that the reduction in overpricing via ETF short-selling leads to an improved price efficiency of such stocks, which is in contrast to the negative unconditional effect of ETFs on underlying stocks.

Overall, in contrast to literature showing the negative impact of ETFs (e.g Filippou, Gozluklu, and Rozental, 2019) this paper provides a novel evidence of the positive effect of industry ETFs on the distressed stocks. In addition, it highlights the increasing heterogeneity of types of stocks within the ETF basket and the importance of considering differences in underlying securities when assessing the impact of this investment vehicle on financial markets. This work can be important to policymakers when evaluating different channels via which the growing ETF market can influence the general economy.

Figures and Tables

Figure 3.1: Structure of MSCI Global Investable Market Index

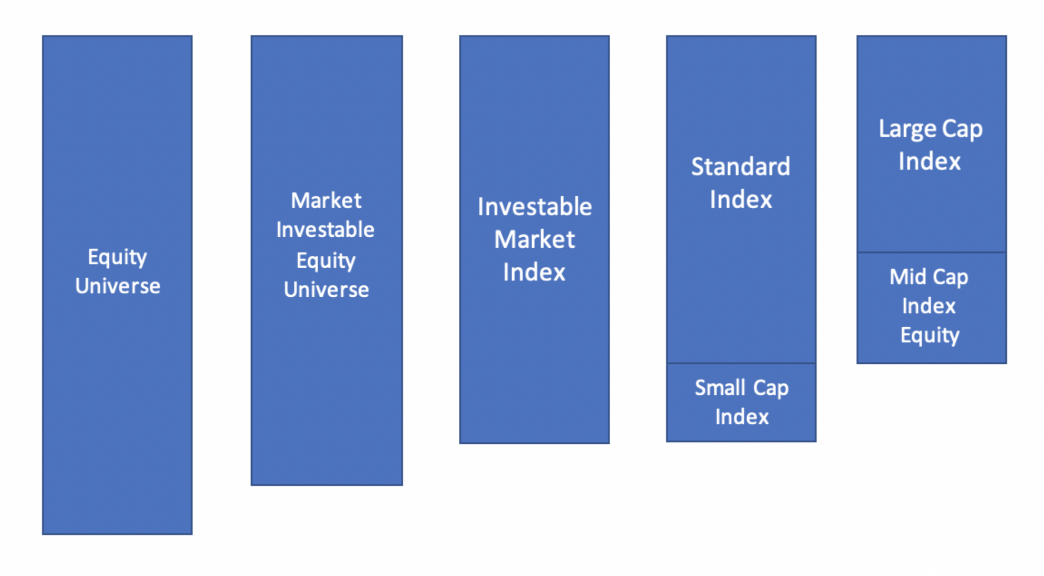


Figure 3.2: Top 15 most Shorted ETFs

This figure shows the 15 most shorted ETFs as of July 24, 2018 reported by ETF.com (https://www.etf.com/sections/features-and-news/most-shorted-etfs). Short interest is calculated as a percentage of shares outstanding that was shorted on that day.

Ticker	Fund	Short Interest %
SMH	VanEck Vectors Semiconductor ETF	211.73
XRT	SPDR S&P Retail ETF	193.06
DSLV	VelocityShares 3X Inverse Silver ETN	140.41
VXX	iPath S&P 500 VIX Short-Term Futures ETN	128.42
VXZ	iPath S&P 500 VIX Mid-Term Futures ETN	97.36
XOP	SPDR S&P Oil & Gas Exploration & Production ETF	93.65
UNG	United States Natural Gas Fund LP	84.51
ERY	Direxion Daily Energy Bear 3X Shares	81.77
DUST	Direxion Daily Gold Miners Index Bear 3x Shares	76.26
JDST	Direxion Daily Junior Gold Miners Index Bear 3X Shares	73.15
XBI	SPDR S&P BIOTECH ETF	65.88
DGAZ	VelocityShares 3X Inverse Natural Gas ETN	59.34
FXE	Invesco CurrencyShares Euro Trust	56.04
IYR	iShares U.S. Real Estate ETF	54.90
XHB	SPDR S&P Homebuilders ETF	52.61

Table 3.1: Fama-French 12 industry Classification

Fama-French 12 industry classification based on SIC codes aggregation. Table includes short group name, industry and the description of sub-industry constituents for each industry group.

Group Name	Industry	Industry Definition
NoDur	Consumer NonDurables	Food, Tobacco, Textiles, Apparel, Leather, Toys
Durbl	Consumer Durables	Cars, TV's, Furniture, Household Appliances
Manuf	Manufacturing	Machinery, Trucks, Planes, Off Furn, Paper, Com Printing
Enrgy	Energy	Oil, Gas, and Coal Extraction and Products
Chems	Chemicals	Chemicals and Allied Products
BusEq	Business Equipment	Computers, Software, and Electronic Equipment
Telcm	Telecommunication	Telephone and Television Transmission
Utils	Utilities	Utilities
Shops	Shops	Wholesale, Retail, and Some Services (Laundries, Repair Shops)
Hlth	Healthcare	Healthcare, Medical Equipment, and Drugs
Money	Finance	Finance
Other	Other	Mines, Constr, BldMt, Trans, Hotels, Bus Serv, Entertainment

Table 3.2: Summary Statistics

Panel A reports the summary statistics of all 127 ETFs used in this study. Mean SIR is computed as $\frac{1}{T}\Sigma_{t=1}^{T}(\frac{1}{N_{k,t}}\Sigma_{j=1}^{N_{k,t}}SIR_{k,j,t})$, where $N_{k,t}$ is the number of ETFs in industry k at time t and $SIR_{k,j,t}$ is the short-interest ratio of fund j in industry k at time t. Industry Sensitivity is based on GICS classification. Minimum SIR is reported in 10^1. Panel B reports the summary statistics for stocks allocated into Distressed (All, members (ETF) and non-members (NoNETF) of ETF basket), Medium Rating and High Rating groups based on terciles of the distribution of S&P long-term credit ratings. Rating is an average value that represents the converted alphabetic credit rating into numeric score. Higher score represents lower rating (AAA=1 ... D=22). Size is an average market capitalization of firms within the category and is expressed in billions of USD. Illiquidity is based on Amihud (2002) illiquidity measure and is displayed in 10^9. Daily Cost of Borrowing Score (DCBS) is from Markit and reflects the cost and relative difficulty to borrow a stock.

Panel A: Industry ETFs						
Group Name	Number of ETFs	Mean SIR	Min SIR	Max SIR	95th SIR percentile	Industry Sensitivity
NoDur	8	5.06%	0.00%	53.14%	20.19%	Defensive
Manuf	12	13.88%	0.00%	100.0%	43.17%	Cyclical
Enrgy	15	15.43%	0.00%	100.0%	88.42%	Cyclical/Defensive
Chems	5	12.63%	0.00%	100.0%	46.23%	Cyclical
BusEq	21	12.38%	0.00%	100.0%	42.49%	Cyclical
Telcm	3	6.81%	0.00%	49.64%	22.35%	Defensive
Utils	7	7.54%	0.04%	69.22%	32.72%	Defensive
Shops	10	18.25%	0.00%	100.0%	100.00%	Cyclical
Hlth	16	11.98%	0.00%	100.0%	49.11%	Defensive
Money	25	17.13%	0.00%	100.0%	100.00%	Cyclical
Other	5	9.13%	0.00%	100.0%	40.38%	Mixed
Panel B: Stocks						
Type of stocks:	Rating	Price	Size	Illiquidity	DCBS	95th DCBS percentile
Distressed-All	14.04	21.81	1.79	5.64	1.27	2.62
Distressed-ETF	13.94	25.06	2.73	4.74	1.24	2.44
Distressed-NoNETF	14.42	17.65	0.96	6.12	1.33	3.18
Medium Rating	10.77	37.31	4.79	4.05	1.11	1.38
High Rating	6.98	172.41	24.30	3.98	1.06	1.31

Table 3.3: Stock Return and ETF Short-Selling

This table reports the results of a panel regression examining the effect of ETF short-selling on distressed stocks:

$$r_{i,t+1} = \alpha + \beta_1 sir_{j,t} + \beta_2(sir_{j,t} \times Distress_{i,t}) + \beta_3 Distress_{i,t} + \\ + \beta_4 sir_t^{stock_i} + \beta_5(sir_t^{stock_i} \times Distress_{i,t}) + \mu Controls + FE_i + FE_t + \varepsilon_{i,t}$$

where $r_{i,t+1}$ is a log return of stock i at month $t+1$. $sir_{j,t}$ is a previous month's log of short interest ratio of the industry ETF j that holds stock i in the basket during month t and $sir_t^{stock_i}$ is the short interest ratio at the stock level. Short interest ratio is measured as a quantity of shares sold short scaled by the number of shares outstanding. $Distress_{i,t}$ is a dummy variable that takes the value of 1 if stock i is classified as distressed during month t. Distress classification is based on the bottom 33rd percentile of the distribution of S&P long-term credit ratings. Controls include one period lagged monthly or quarterly values (depending on availability) of $illiq_{i,t}$ (monthly average of daily Amihud (2002) illiquidity ratios for stock i), $r_{i,t-1,t-12}$, log of market capitalisation,asset growth, book-to-market ratio, growth in the long-term net operating assets, gross profitability, investment growth, operating profitability and accruals. The regression frequency is monthly and the time period is from January 1999 - February 2017. Loadings on $sir_{j,t}$ are expressed in 10^1. Where stated regressions include firm and year fixed effects. All standard errors are bootstrapped and clustered by ETF industries. *,**,*** denote statistical significance at 10%, 5% and 1% levels.

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	$r_{i,t+1}$	$r_{i,t+1}$	$r_{i,t+1}$	$r_{i,t+1}$	$r_{i,t+1}$	$r_{i,t+1}$
$sir_{j,t}$	-0.002	-0.004	-0.003	-0.002	-0.003	-0.004
	(-0.26)	(-0.49)	(-0.34)	(-0.27)	(0.27)	(-0.12)
$sir_{j,t} \times Distress_{i,t}$			-0.003***	-0.002***	-0.003***	-0.002***
			(-5.31)	(-3.74)	(-4.61)	(-3.26)
$Distress_{i,t}$			0.022***	0.010***	0.016***	0.015**
			(3.32)	(2.88)	(2.97)	(2.18)
$sir_t^{stock_i}$					-0.001***	-0.001**
					(-2.71)	(-2.05)
$sir_t^{stock_i} \times Distress_{i,t}$					-0.001	-0.001
					(-0.62)	(-0.48)
$Constant$	0.003	-0.027	-0.001	-0.001	-0.001	-0.024**
	(1.10)	(-1.53)	(-0.63)	(-1.41)	(-0.11)	(-1.97)
Controls	No	Yes	No	Yes	No	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Quarter-Year FE	No	Yes	No	Yes	No	Yes
Observations	742,327	474,727	742,327	474,727	733,361	472,422
$Adjusted \quad R^2(\%)$	0.00	11.59	0.15	11.64	0.16	11.63

Table 3.4: Speculation and Hedging Hypotheses

This table reports the results of a panel regression examining the effect of ETF short-selling on distressed stocks and non-distressed stocks during the financial crisis period and outside of the such period:

$$r_{i,t+1} = \alpha + \beta_1 sir_{j,t} + \beta_2(sir_{j,t} \times Distress_{i,t}) + \beta_3 Distress_{i,t} + \\ + \beta_4 sir_t^{stock_i} + \beta_5(sir_t^{stock_i} \times Distress_{i,t}) + \mu Controls + FE_i + FE_t + \varepsilon_{i,t}$$

where $r_{i,t+1}$ is a log return of stock i at month $t+1$. $sir_{j,t}$ is a previous month's log of short interest ratio of the industry ETF j that holds stock i in the basket during month t and $sir_t^{stock_i}$ is the short interest ratio at the stock level. Short interest ratio is measured as a quantity of shares sold short scaled by the number of shares outstanding. $Distress_{i,t}$ is a dummy variable that takes the value of 1 if stock i is classified as distressed during month t. Distress classification is based on the bottom 33rd percentile of the distribution of S&P long-term credit ratings. Controls include one period lagged monthly or quarterly values (depending on availability) of $illiq_{i,t}$ (monthly average of daily Amihud (2002) illiquidity ratios for stock i), $r_{i,t-1,t-12}$, log of market capitalisation, asset growth, book-to-market ratio, growth in the long-term net operating assets, gross profitability, investment growth, operating profitability and accruals. Crisis results are estimated during 2006Q4 - 2008Q4 and non-crisis results are estimated on the whole sample excluding the crisis period. The regression frequency is monthly and the time period is from January 1999 - February 2017. Panel regressions include firm fixed effects. All standard errors are bootstrapped and clustered by ETF industries. *,**,*** denote statistical significance at 10%, 5% and 1% levels.

	(1)	(2)
	Non-Crisis	Crisis
VARIABLES	$r_{i,t+1}$	$r_{i,t+1}$
$sir_{j,t}$	0.002**	-0.005**
	(2.21)	(-1.96)
$sir_{j,t} \times Distress_{i,t}$	-0.003***	-0.001
	(-3.87)	(-1.12)
$Distress_{i,t}$	0.019***	-0.028***
	(3.22)	(2.98)
$sir_t^{stock_i}$	0.002**	-0.014***
	(2.13)	(-2.84)
$sir_t^{stock_i} \times Distress_{i,t}$	0.001	-0.007*
	(0.42)	(1.68)
Constant	0.017***	-0.132***
	(3.67)	(-3.81)
Controls	Yes	Yes
Firm FE	Yes	Yes
Quarter-Year FE	No	No
Observations	396,045	76,377
Adjusted R^2(%)	1.29	0.64

Table 3.5: Cyclical and Defensive ETFs

This table reports the results of a panel regression examining the effect of ETF short-selling on distressed stocks conditional on the type of industry that ETF is benchmarked against (Cyclical vs Defensive):

$$r_{i,t+1} = \alpha + \beta_1 sir_{j,t} + \beta_2 (sir_{j,t} \times Distress_{i,t}) + \\ + \beta_3 Distress_{i,t} + \beta_4 sir_t^{stock_i} + \beta_5 (sir_{j,t}^{stock_i} \times Distress_{i,t}) + FE_i + FE_t + \varepsilon_{i,t}$$

where $r_{i,t+1}$ is a log return of stock i at month $t+1$. $sir_{j,t}$ is a previous month's log of short interest ratio of the industry ETF j that holds stock i in the basket during month t and $sir_t^{stock_i}$ is the short interest ratio at the stock level. Short interest ratio is measured as a quantity of shares sold short scaled by the number of shares outstanding. $Distress_{i,t}$ is a dummy variable that takes the value of 1 if stock i is classified as distressed during month t. Distress classification is based on the bottom 33rd percentile of the distribution of S&P long-term credit ratings. Controls include one period lagged monthly or quarterly values (depending on availability) of $illiq_{i,t}$ (monthly average of daily Amihud (2002) illiquidity ratios for stock i), $r_{i,t-1,t-12}$, log of market capitalisation,asset growth, book-to-market ratio, growth in the long-term net operating assets, gross profitability, investment growth, operating profitability and accruals. $Cyclical$ and $Defensive$ industries are based on Global Industry Classficiation Standards (*GICS*) official sector classification matched to corresponding SIC industries. *Cyclical* industries include *Manuf*, *Chems*, *BusEq*, *Shops*, *Money*. $Defensive$ industries include *NoDur*, *Telcm*, *Utils*, *Hlth*. *Other* and *Energy* industry is excluded from the analysis. The regression frequency is monthly and the time period is from January 1999 - February 2017. Loadings on $sir_{j,t}$ are expressed in 10^1. Where stated regressions include firm and year fixed effects. All standard errors are bootstrapped and clustered by ETF industries. *,**,*** denote statistical significance at 10%, 5% and 1% levels.

VARIABLES	Defensive (1) $r_{i,t+1}$	Defensive (2) $r_{i,t+1}$	Cyclical (3) $r_{i,t+1}$	Cyclical (4) $r_{i,t+1}$
$sir_{j,t}$	-0.009**	0.006**	0.001	0.004
	(-2.14)	(2.06)	(0.23)	(0.68)
$sir_{j,t} \times Distress_{i,t}$	-0.001	-0.001	-0.002***	-0.002***
	(-0.79)	(-0.76)	(-4.83)	(-6.11)
$Distress_{i,t}$	0.013	0.004	0.018**	0.017**
	(1.28)	(0.46)	(2.91)	(2.67)
$sir_t^{stock_i}$	0.001	-0.001*	-0.002**	-0.002**
	(0.64)	(-1.72)	(-2.20)	(-2.07)
$sir_t^{stock_i} \times Distress_{i,t}$	-0.007	-0.002	0.003	0.001
	(-1.23)	(-0.86)	(1.19)	(1.06)
$Constant$	-0.003	-0.046***	-0.013	-0.030
	(-1.01)	(-4.87)	(-1.41)	(-1.19)
Controls	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Quarter-Year FE	No	Yes	No	Yes
Observations	125,185	125,185	251,257	251,257
Adjusted R^2(%)	0.75	9.21	0.73	12.66

Table 3.6: ETF Short-Interest and Investor Sentiments

This table reports the results of the regression of the log of cross-sectional average of ETF SIR across all funds in the sample at month $t+1$ ($avsir_{t+1}$) on one of the sentiments proxies: the Baker and Wurgler (2006) investor sentiment measure as of month t (S_t) and Michigan consumer sentiments index (MS_t) orthogonalised to 12-month growth in industrial production, durable, nondurable and service consumption, as well as to NBER recession indicator and growth in employment:

$$avsir_{t+1} = \alpha + \beta_1 Sentiments_t + avsir_t + \varepsilon_{i,t}$$

The regression frequency is monthly and the time period is from January 1999 - February 2017. The table reports regression coefficients and Newey and West (1987) corrected t-statistics with six-month lags. *,**,*** denote statistical significance at 10%, 5% and 1% levels.

	(1)	(2)
VARIABLES	$avsir_{t+1}$	$avsir_{t+1}$
S_t	0.038**	
	(2.29)	
MS_t		0.003**
		(2.48)
$avsir_t$	0.801***	0.808***
	(13.78)	(15.15)
$Constant$	-0.412***	-0.395***
	(-3.33)	(-3.62)
Observations	230	230
$Adjusted \quad R^2(\%)$	79.27	79.43

Table 3.7: Distressed Anomaly and ETF Ownership

This table presents the results of OLS regression of portfolio excess returns of Distressed, Medium and High rated stocks on Fama and French (1993) three-factor model (FF3). In addition, it reports alphas obtained by running the same regression on three-factor model with short-term reversals (FF3+STR α), three-factor model with Pástor and Stambaugh (2003) liquidity factor (FF3 + Liq α) and Carhart (1997) four-factor model (4-factor α). A numeric score is assigned for every S&P long-term credit rating (from AAA=1...D=22). Every month stocks are ranked based on terciles of the distribution of such ranks. *Distressed* stocks are defined as firms with credit rating in the lowest tercile as of month $t-1$. Equally-weighted portfolios are formed and held for 1 month. ETF stocks are stocks that are part of any industry ETF basket during month $t-1$. Regression frequency is monthly and the time period is from January 1999 - February 2017. The table reports Newey and West (1987) corrected t-statistics with six-month lags. *,**,*** denote statistical significance at 10%, 5% and 1% levels.

	Distressed		Distressed	Medium	High
	NonETF	ETF	All	All	All
FF3 α	-0.007***	0.002	-0.003*	0.002**	0.003***
	(-3.09)	(0.87)	(-1.76)	(1.98)	(3.84)
MRP	1.104***	1.489***	1.296***	1.045***	0.881***
	(23.43)	(25.72)	(32.90)	(25.61)	(43.77)
SMB	0.684***	0.526***	0.660***	0.254*	-0.075
	(5.50)	(3.34)	(4.64)	(1.96)	(-1.09)
HML	0.650***	0.159	0.450***	0.560***	0.460***
	(8.94)	(1.42)	(4.77)	(5.77)	(7.76)
FF3+STR α	-0.006***	0.002	-0.003*	0.002*	0.003***
	(-2.89)	(0.83)	(-1.72)	(1.97)	(3.78)
FF3+Liq α	-0.006***	0.001	-0.002	0.002*	0.003***
	(-2.74)	(0.67)	(-1.49)	(1.68)	(3.67)
4-factor α	-0.007***	0.001	-0.003*	0.002*	0.003***
	(-2.98)	(0.71)	(-1.70)	(1.93)	(3.27)

Table 3.8: Price Efficiency and ETF Short-Selling

This table reports the results of a panel regression examining the effect of ETF short-selling on stock efficiency:

$$d_{i,t+1} = \alpha + \beta_1 sir_{j,t} + \beta_2(sir_{j,t} \times Distress_{i,t}) + \beta_3 Distress_{i,t} + \beta_4 sir_t^{stock_i} + \\ + (\beta_5 sir^{stock_i} \times Distress_{i,t}) + \beta_6 d_{i,t} + \mu Controls + FE_i + FE_t + \varepsilon_{i,t}$$

where $d_{i,t+1}$ is a log of one of the 2 versions of price delay measure of stock i during month $t+1$. 2 versions of delay D1 and D2 are defined in equation 3.8. $sir_{j,t}$ is a previous month's log of short interest ratio of the industry ETF j that holds stock i in the basket during month t and $sir_t^{stock_i}$ is the short interest ratio at the stock level. Short interest ratio is measured as a quantity of shares sold short scaled by the number of shares outstanding. $Distress_{i,t}$ is a dummy variable that takes the value of 1 if stock i is classified as distressed during month t. Distress classification is based on the bottom 33rd percentile of the distribution of S&P long-term credit ratings. Controls include one period lagged monthly or quarterly values (depending on availability) of $VWAP$ (log of daily volume-weighted average price), log of market capitalization, log of trading volume orthogonalised to market capitalization and log of number of analysts covering the stock. Regression frequency is monthly and the time period is from January 1999 - February 2017. Where stated regressions include firm and year fixed effects. All standard errors are bootstrapped and clustered by ETF industries. *,**,*** denote statistical significance at 10%, 5% and 1% levels.

VARIABLES	(1) $d1_{i,t+1}$	(2) $d2_{i,t+1}$
$sir_{j,t}$	0.003**	0.001*
	(1.99)	(1.78)
$sir_{j,t} \times Distress_{i,t}$	-0.012**	-0.019**
	(-2.47)	(-2.23)
$Distress_{i,t}$	0.001	-0.003
	(0.14)	(-0.42)
$sir_t^{stock_i}$	-0.018*	-0.006**
	(-1.86)	(-2.10)
$sir_t^{stock_i} \times Distress_{i,t}$	-0.006***	-0.005***
	(-2.81)	(-3.24)
$d_{i,t}$	0.035***	0.019***
	(4.21)	(3.85)
$Constant$	-0.588**	-0.690***
	(-2.38)	(-4.57)
Controls	Yes	Yes
Firm FE	Yes	Yes
Quarter-Year FE	Yes	Yes
Observations	709,719	709,719
$Adjusted\ \ R^2(\%)$	15.48	8.47

Table 3.9: Robustness: Altman's Z-score, Distressed Anomaly and ETF Ownership

This table presents the results of OLS regression of portfolio excess returns of Distressed, Medium and High rated stocks on Fama and French (1993) three-factor model (FF3). In addition, it reports alphas obtained by running the same regression on three-factor model with short-term reversals (FF3+STR α), three-factor model with Pástor and Stambaugh (2003) liquidity factor (FF3 + Liq α) and Carhart (1997) four-factor model (4-factor α). Altman (1968) Z-scores are computed quarterly for each stock. Every quarter stocks are ranked based on terciles of the distribution of such scores. *Distressed* stocks are defined as firms with credit rating in the lowest tercile as of quarter $t-1$. Equally-weighted portfolios are formed and held for 1 month. ETF stocks are stocks that are part of any industry ETF basket during quarter $t-1$. Regression frequency is quarterly and the time period is from January 1999 - February 2017. The table reports Newey and West (1987) corrected t-statistics with four-month lags. *,**,*** denote statistical significance at 10%, 5% and 1% levels.

	Distressed		Distressed	Medium	High
	NonETF	ETF	All	All	All
FF3 α	-0.021***	0.007*	-0.015***	0.018***	0.012***
	(-3.03)	(1.79)	(-3.30)	(6.65)	(3.12)
MRP	1.290***	1.715***	1.669***	0.978***	1.009***
	(6.17)	(15.13)	(14.75)	(19.29)	(23.38)
SMB	0.502***	0.323**	0.500***	0.234***	0.028*
	(4.58)	(2.41)	(7.74)	(9.72)	(1.76)
HML	0.490***	0.681***	0.593***	0.380***	-0.321***
	(5.59)	(6.83)	(9.63)	(3.76)	(-5.21)
FF3+STR α	-0.019***	0.007*	-0.014***	0.018***	0.011***
	(-2.89)	(1.68)	(-2.41)	(6.28)	(2.72)
FF3+Liq α	-0.017***	0.006*	-0.015**	0.017***	0.011***
	(-3.54)	(1.76)	(-1.98)	(5.86)	(2.90)
4-factor α	-0.021***	0.005	-0.014**	0.018***	0.011***
	(-2.94)	(1.57)	(-2.02)	(5.41)	(2.62)

Table 3.10: Robustness: Moody's KMV, Distressed Anomaly and ETF Ownership

This table presents the results of OLS regression of portfolio excess returns of Distressed, Medium and High rated stocks on Fama and French (1993) three-factor model (FF3). In addition, it reports alphas obtained by running the same regression on three-factor model with short-term reversals (FF3+STR α), three-factor model with Pástor and Stambaugh (2003) liquidity factor (FF3 + Liq α) and Carhart (1997) four-factor model (4-factor α). Expected default frequencies (based on Bharath and Shumway (2008) proxy of Moody's KMV model) are computed quarterly for each stock. Every quarter stocks are ranked based on terciles of the distribution of EDFs. *Distressed* stocks are defined as firms with credit rating in the lowest tercile as of quarter $t-1$. Equally-weighted portfolios are formed and held for 1 month. ETF stocks are stocks that are part of any industry ETF basket during quarter $t-1$. Regression frequency is quarterly and the time period is from January 1999 - February 2017. The table reports Newey and West (1987) corrected t-statistics with four-month lags. *,**,*** denote statistical significance at 10%, 5% and 1% levels.

	Distressed		Distressed	Medium	High
	NonETF	ETF	All	All	All
FF3 α	-0.008***	0.002	-0.005**	0.005***	0.004***
	(-5.12)	(0.82)	(-2.32)	(5.42)	(5.20)
MRP	0.749***	1.402***	1.071***	0.877***	0.750***
	(25.80)	(18.87)	(26.43)	(31.18)	(19.06)
SMB	0.656***	0.262	0.577***	0.536***	0.484***
	(5.48)	(1.22)	(5.14)	(9.33)	(19.99)
HML	0.444***	0.585***	0.549***	0.344***	0.137**
	(5.75)	(3.29)	(5.61)	(8.77)	(2.17)
FF3+STR α	-0.008***	0.002	-0.005**	0.004***	0.004***
	(-4.38)	(0.76)	(-2.29)	(5.60)	(5.08)
FF3+Liq α	-0.008***	0.002	-0.005*	0.004***	0.003***
	(-3.42)	(0.73)	(-1.88)	(3.48)	(4.74)
4-factor α	-0.007***	0.001	-0.005**	0.004***	0.004***
	(-3.92)	(0.59)	(-1.99)	(5.26)	(5.17)

Table 3.11: Robustness: Shareholder Advantage, Distressed Anomaly and ETF Ownership

This table presents the alpha esimates of OLS regression of portfolio excess returns of Distressed, Medium and High rated stocks on Fama and French (1993) three-factor model (FF3 α). In addition, it reports alphas obtained by running the same regression on three-factor model with short-term reversals (FF3+STR α), three-factor model with Pástor and Stambaugh (2003) liquidity factor (FF3 + Liq α) and Carhart (1997) four-factor model (4-factor α). A numeric score is assigned for every S&P long-term credit rating (from AAA=1...D=22). Every month stocks are ranked based on deciles of the distribution of such ranks. Distressed (Highly Distressed) stocks are defined as firms with credit rating in the lowest tercile (decile) as of month $t-1$ (highest tercile (decile) of a credit score). Stocks are also independently sorted into terciles by asset size to obtain High Asset Size group and by asset specificity to obtain High Asset Specificity group. Asset Specificity is proxied by the Herfindahl index of sales in an industry. Equally-weighted portfolios are formed and held for 1 month. ETF members are stocks that are part of any industry ETF basket during month $t-1$. Panel A shows the results for all stocks excluding a Highly Distressed group. Panel B uses only Highly Distressed Stocks. Panel C shows the results for a group of stocks that are both in the Distressed group and in the High Asset Size group. Panel D shows the results for a group of stocks that are both in the Distressed group and in the High Asset Specificity group. Regression frequency is monthly and the time period is from January 1999 - February 2017. The table reports Newey and West (1987) corrected t-statistics with six-month lags. *,**,*** denote statistical significance at 10%, 5% and 1% levels.

	FF3-α	t-stat	FF3+STR-α	t-stat	FF3+Liq-α	t-stat	4-factor-α	t-stat
Panel A: Distressed Group Excluding Highly Distressed Stocks								
All	-0.002*	(-1.80)	-0.002*	(-1.74)	-0.001	(-1.61)	-0.002*	(-1.74)
NonETF	-0.004**	(-2.24)	-0.004**	(-2.02)	-0.004**	(-2.11)	-0.004**	(-2.15)
ETF	0.002	(0.70)	0.002	(0.68)	0.001	(0.50)	0.002	(0.42)
Panel B: Highly Distressed Stocks								
All	-0.005**	(-2.12)	-0.005*	(-1.86)	-0.004*	(-1.76)	-0.004**	(-2.03)
NonETF	-0.011***	(-3.25)	-0.011***	(-3.14)	-0.011***	(-3.21)	-0.010***	(-3.18)
ETF	0.004	(0.94)	0.004	(0.88)	0.003	(0.77)	0.006	(0.85)
Panel C: Distressed Stocks & High Asset Size								
All	-0.004**	(-2.07)	-0.003**	(-2.01)	-0.002*	(-1.87)	-0.003**	(-2.04)
NonETF	-0.010***	(-2.93)	-0.010***	(-2.86)	-0.009***	(-2.77)	-0.009**	(-2.15)
ETF	-0.001	(-0.69)	-0.001	(-0.49)	-0.000	(-0.58)	-0.001	(-0.18)
Panel D: Distressed Stocks & High Asset Specificity								
All	-0.003**	(-2.43)	-0.003**	(-2.42)	-0.002**	(-1.82)	-0.003**	(-2.30)
NonETF	-0.004***	(-2.67)	-0.004***	(-2.65)	-0.003**	(-2.47)	-0.004***	(-2.45)
ETF	-0.002	(-0.90)	-0.002	(-0.82)	-0.001	(-0.71)	-0.001	(-0.86)

Table 3.12: Price Efficiency and Alternative Short-Selling

This table reports the results of a panel regression examining the effect of ETF short-selling on stock efficiency conditional on the existence of call or put options:

$$d_{i,t+1} = \alpha + \beta_1 sir_{j,t} + \beta_2(sir_{j,t} \times Distress_{i,t}) + \beta_3 Distress_{i,t} + \beta_4 sir_t^{stock_i} + \\ + (\beta_5 sir^{stock_i} \times Distess_{i,t}) + \beta_6 d_{i,t} + \mu Controls + FE_i + FE_t + \varepsilon_{i,t}$$

where $d_{i,t+1}$ is a log of one of the 2 versions of price delay measure of stock i during month $t+1$. 2 versions of delay D1 and D2 are defined in equation 3.8. $sir_{j,t}$ is a previous month's log of short interest ratio of the industry ETF j that holds stock i in the basket during month t and $sir_t^{stock_i}$ is the short interest ratio at the stock level. The regression is performed conditionally if a put or a call option exists for stock i during month t. Short interest ratio is measured as a quantity of shares sold short scaled by the number of shares outstanding. $Distress_{i,t}$ is a dummy variable that takes the value of 1 if stock i is classified as distressed during month t. Distress classification is based on the bottom 33rd percentile of the distribution of S&P long-term credit ratings. Controls include one period lagged monthly or quarterly values (depending on availability) of $VWAP$ (log of daily volume-weighted average price), log of market capitalization, log of trading volume orthogonalised to market capitalization and log of number of analysts covering the stock. Regression frequency is monthly and the time period is from January 1999 - February 2017. Where stated regressions include firm and year fixed effects. All standard errors are bootstrapped and clustered by ETF industries. *,**,*** denote statistical significance at 10%, 5% and 1% levels.

	(1)	(2)
VARIABLES	$d1_{i,t+1}$	$d2_{i,t+1}$
$sir_{j,t}$	0.001	0.001
	(1.58)	(0.38)
$sir_{j,t} \times Distress_{i,t}$	-0.004**	-0.017**
	(-1.98)	(-2.03)
$Distress$	0.004	-0.001
	(0.42)	(-0.22)
$sir_t^{stock_i}$	-0.018**	-0.010**
	(-2.39)	(-2.11)
$sir_t^{stock_i} \times Distress_{i,t}$	-0.005**	-0.004**
	(-1.99)	(-2.24)
$d_{i,t}$	0.031***	0.015***
	(5.14)	(3.46)
$Constant$	-0.596***	-0.689***
	(-4.31)	(-2.96)
Controls	Yes	Yes
Firm FE	Yes	Yes
Quarter-Year FE	Yes	Yes
Observations	646,649	646,649
$Adjusted \quad R^2(\%)$	13.97	6.28

Concluding Remarks

In this thesis, I investigate the impact of strong ETF industry growth observed in recent years on the underlying equity markets. In contrast to the majority of ETF literature that mostly focuses on the negative effects of the funds my research finds that the consequences of industry and country-level ETF trading on the underlying stocks is multidimensional. This thesis provides a better understanding of this type of product and highlights the systematic importance of the ETF as an investment tool.

In chapter 1, I study the ETF arbitrage mechanism and present evidence that it is not risk-free and therefore, is not always functional (which is in contrast to what is often advertised by ETF sponsors). I argue that APs and other arbitragers would only attempt to eliminate such deviation of law of one price when the reward is substantial to compensate for arbitrage risks. I relate significant differences of ETF prices from NAVs to the level of aggregate financial frictions. While liquidity is often cited as a main driver of deviations I show that currency and equity market illiquidity risks are only able to explain up to 24% of variation in premiums, suggesting that other risk factors are also important. In particular, I show that the absolute value of ETF premium is a robust proxy for multiple dimensions of country-level frictions risk including credit risk, funding illiquidity and information uncertainty. Such measure is better at a country-level than many other known proxies due the availability of cross-sectional data that, in contrast to, for example, ADRs, does not rely on individual stock mispricing. I investigate what stock characteristics drive the sensitivity of securities to aggregate frictions and find that small cyclical stocks with strong price uncertainty (low analyst coverage and high dispersion in forecasts), as well as high leverage are the most exposed. Finally, I show that friction risk based on this ETF measure is priced internationally as investors demand a compensation for being exposed to it.

In chapter 2, I show that the impact of country-level ETFs on the diversification ability of U.S. investors is negative. I develop a shock propagation mechanism and show that it is responsible for an increased correlation between foreign countries and the U.S. market. I investigate how investors make their decisions regarding country-level ETF holdings when facing U.S. economic uncertainty in contrast to foreign uncertainty. Using U.S. and local VIX I show that they mainly react to U.S. developments by selling country-ETFs and switching to cash products. In contrast, they only react to significant negative news in the foreign economies. I also study the reaction of investors when facing an increase in political uncertainty and show that, differently to economic uncertainty, they leave the U.S. market and move their capital abroad. The time-series analysis shows that U.S. shocks are propagated to foreign stock markets via ETF arbitrage mechanism, which, as a result, leads to an increase in correlation of stock returns across countries. Controlling for business cycles and trading channel, the cross-sectional sorts demonstrate that such propagation is stronger when limits to arbitrage are lower and when the transmitted information is fundamental. I find that countries with a more liquid stock market (which makes the arbitrage easier) and that have a high price discovery in the ETF market are more correlated with the U.S..

In chapter 3, I demonstrate the positive role of industry ETFs in improving the price efficiency of the underlying stocks. I emphasize the heterogeneity of the type of stocks in the ETF basket and show the contrasting effect of ETFs when sorting stocks by credit quality. I show that since investors are limited in the direct short-selling of overpriced distressed securities, they can achieve the negative exposure indirectly via industry ETFs. I demonstrate that ETF short-selling negatively predicts the returns of distressed stocks and that the level of short-selling is higher in the period when overpricing is more likely. I further study the under-performance of stocks with high credit risk (the distressed anomaly) that is often related to stock overpricing. I show that when stocks are sorted by industry ETF membership, distressed securities that are part of the basket are less likely to show the sings of anomaly. In addition, in the direct test of price efficiency I show that ETF short-selling reduces the price delay measure for distressed stocks. The results presented in the paper are robust to different measures of credit risk, different distressed anomaly explanations and when accounting for the presents of alternative shot-selling mechanisms.

Overall, my work highlights the complexity of assessing the consequences of ETF development and promotes further studies in this direction. In addition, this thesis provides novel evidence on the importance of the ETF industry to the global economy and may be useful for policy makers when attempting to regulate the markets.

Appendix A

Supporting Documentation: Chapter 1

A.1 Tables

Table A.1: Correlation of Absolute Premiums

This table shows the correlation between different versions of NAV adjustment for each country in the sample. NA means no adjustment, GIR is the adjustment based on Goetzmann et al. (2001) and ES is the adjustment based on Engle and Sarkar (2006). Frequency is daily and the sample period is from June 2002- June 2018.

Countries	$\rho(GIR, NA)$	$\rho(ES, NA)$	$\rho(GIR, ES)$
AUS	0.61	0.51	0.95
AUT	0.78	0.76	0.95
BEL	0.81	0.78	0.93
CAN	0.90	0.92	0.92
DNK	0.79	0.78	0.93
FIN	0.83	0.81	0.91
FRA	0.75	0.63	0.78
DEU	0.83	0.65	0.78
HKG	0.87	0.59	0.84
IRL	0.95	0.93	0.96
ISR	0.94	0.79	0.92
ITA	0.84	0.66	0.83
JPN	0.79	0.56	0.84
NLD	0.77	0.64	0.83
NZL	0.81	0.63	0.92
NOR	0.80	0.77	0.90
SGP	0.89	0.69	0.91
ESP	0.84	0.67	0.86
SWE	0.83	0.68	0.85
CHE	0.81	0.74	0.92
GBR	0.80	0.67	0.86
USA	-	-	-

Table A.2: Correlation of Absolute Premiums across Countries: no adjustment

	AUS	AUT	BEL	CAN	DNK	FIN	FRA	DEU	HKG	IRL	ISR	ITA	JPN	NLD	NZL	NOR	SGP	ESP	SWE	CHE	GBR	USA
AUS	1.00																					
AUT	0.55	1.00																				
BEL	0.57	0.71	1.00																			
CAN	0.35	0.27	0.33	1.00																		
DNK	0.36	0.49	0.52	0.22	1.00																	
FIN	0.25	0.55	0.56	0.21	0.46	1.00																
FRA	0.60	0.72	0.74	0.31	0.52	0.56	1.00															
DEU	0.58	0.70	0.72	0.30	0.52	0.56	0.83	1.00														
HKG	0.71	0.48	0.46	0.29	0.27	0.19	0.47	0.47	1.00													
IRL	0.16	0.22	0.27	0.08	0.27	0.30	0.29	0.27	0.16	1.00												
ISR	0.49	0.50	0.49	0.30	0.28	0.20	0.45	0.44	0.48	0.18	1.00											
ITA	0.58	0.71	0.73	0.28	0.52	0.57	0.84	0.78	0.47	0.32	0.42	1.00										
JPN	0.63	0.39	0.36	0.24	0.22	0.14	0.38	0.38	0.61	0.12	0.45	0.38	1.00									
NLD	0.61	0.72	0.74	0.32	0.53	0.58	0.83	0.79	0.51	0.27	0.47	0.80	0.40	1.00								
NZL	0.66	0.46	0.43	0.32	0.29	0.25	0.43	0.46	0.45	0.16	0.32	0.42	0.34	0.45	1.00							
NOR	0.40	0.51	0.48	0.25	0.45	0.41	0.48	0.50	0.27	0.18	0.18	0.51	0.18	0.50	0.32	1.00						
SGP	0.65	0.50	0.48	0.35	0.31	0.23	0.47	0.48	0.72	0.13	0.53	0.44	0.51	0.52	0.52	0.32	1.00					
ESP	0.60	0.72	0.75	0.30	0.52	0.58	0.85	0.81	0.48	0.30	0.45	0.85	0.40	0.82	0.42	0.50	0.46	1.00				
SWE	0.59	0.67	0.71	0.35	0.49	0.66	0.76	0.73	0.49	0.24	0.45	0.72	0.39	0.78	0.40	0.48	0.49	0.76	1.00			
CHE	0.53	0.64	0.66	0.37	0.46	0.54	0.69	0.67	0.43	0.25	0.46	0.65	0.36	0.71	0.43	0.43	0.49	0.68	0.69	1.00		
GBR	0.54	0.58	0.62	0.32	0.47	0.49	0.66	0.65	0.44	0.36	0.43	0.63	0.39	0.66	0.40	0.40	0.44	0.65	0.65	0.62	1.00	
USA	0.25	0.31	0.31	0.27	0.10	0.11	0.20	0.25	0.29	0.09	0.50	0.18	0.22	0.24	0.25	0.07	0.33	0.18	0.25	0.28	0.27	1.00

Table A.3: Correlation of Absolute Premiums across Countries: Goetzmann et al. (2001) adjustment

	AUS	AUT	BEL	CAN	DNK	FIN	FRA	DEU	HKG	IRL	ISR	ITA	JPN	NLD	NZL	NOR	SGP	ESP	SWE	CHE	GBR	USA
AUS	1.00																					
AUT	0.37	1.00																				
BEL	0.39	0.63	1.00																			
CAN	0.30	0.30	0.35	1.00																		
DNK	0.15	0.39	0.43	0.16	1.00																	
FIN	0.24	0.46	0.48	0.14	0.38	1.00																
FRA	0.37	0.60	0.62	0.21	0.43	0.49	1.00															
DEU	0.38	0.59	0.59	0.23	0.42	0.49	0.73	1.00														
HKG	0.39	0.30	0.27	0.23	0.10	0.11	0.26	0.28	1.00													
IRL	0.11	0.20	0.26	0.09	0.25	0.28	0.28	0.28	0.14	1.00												
ISR	0.27	0.35	0.36	0.24	0.16	0.12	0.27	0.28	0.29	0.16	1.00											
ITA	0.35	0.57	0.60	0.18	0.41	0.52	0.73	0.68	0.26	0.31	0.24	1.00										
JPN	0.24	0.17	0.17	0.16	0.07	0.14	0.15	0.14	0.29	0.08	0.24	0.14	1.00									
NLD	0.38	0.61	0.64	0.27	0.43	0.48	0.73	0.68	0.30	0.28	0.29	0.69	0.13	1.00								
NZL	0.41	0.27	0.24	0.20	0.13	0.20	0.25	0.30	0.16	0.15	0.15	0.27	0.11	0.27	1.00							
NOR	0.22	0.42	0.37	0.22	0.37	0.31	0.39	0.40	0.12	0.16	0.08	0.39	0.07	0.39	0.18	1.00						
SGP	0.43	0.35	0.35	0.30	0.19	0.19	0.25	0.32	0.54	0.10	0.39	0.25	0.21	0.35	0.29	0.18	1.00					
ESP	0.34	0.59	0.62	0.21	0.43	0.50	0.74	0.70	0.26	0.29	0.30	0.78	0.14	0.72	0.25	0.39	0.25	1.00				
SWE	0.40	0.56	0.59	0.26	0.36	0.60	0.65	0.60	0.27	0.23	0.28	0.61	0.15	0.65	0.24	0.36	0.31	0.64	1.00			
CHE	0.38	0.52	0.57	0.31	0.39	0.45	0.56	0.55	0.23	0.23	0.29	0.49	0.11	0.59	0.24	0.33	0.35	0.54	0.55	1.00		
GBR	0.36	0.43	0.51	0.26	0.38	0.42	0.52	0.50	0.27	0.36	0.26	0.48	0.19	0.52	0.22	0.29	0.30	0.49	0.52	0.47	1.00	
USA	0.27	0.31	0.32	0.29	0.09	0.13	0.19	0.25	0.27	0.11	0.48	0.15	0.14	0.23	0.19	0.08	0.35	0.16	0.23	0.26	0.25	1.00

Table A.4: Correlation of Absolute Premiums across Countries: Engle and Sarkar (2006) adjustment

	AUS	AUT	BEL	CAN	DNK	FIN	FRA	DEU	HKG	IRL	ISR	ITA	JPN	NLD	NZL	NOR	SGP	ESP	SWE	CHE	GBR	USA
AUS	1.00																					
AUT	0.36	1.00																				
BEL	0.37	0.60	1.00																			
CAN	0.32	0.29	0.35	1.00																		
DNK	0.16	0.33	0.39	0.13	1.00																	
FIN	0.22	0.38	0.41	0.12	0.33	1.00																
FRA	0.35	0.51	0.53	0.23	0.39	0.42	1.00															
DEU	0.33	0.50	0.50	0.22	0.36	0.37	0.63	1.00														
HKG	0.37	0.27	0.25	0.19	0.04	0.08	0.24	0.23	1.00													
IRL	0.11	0.17	0.23	0.06	0.23	0.28	0.24	0.22	0.11	1.00												
ISR	0.21	0.35	0.36	0.19	0.16	0.08	0.20	0.22	0.23	0.12	1.00											
ITA	0.31	0.50	0.51	0.21	0.39	0.45	0.64	0.54	0.19	0.26	0.20	1.00										
JPN	0.23	0.16	0.17	0.16	0.06	0.13	0.15	0.13	0.25	0.07	0.23	0.13	1.00									
NLD	0.35	0.55	0.56	0.27	0.37	0.40	0.63	0.58	0.26	0.23	0.29	0.57	0.16	1.00								
NZL	0.37	0.25	0.23	0.24	0.13	0.18	0.24	0.24	0.13	0.14	0.11	0.22	0.14	0.26	1.00							
NOR	0.22	0.37	0.33	0.21	0.33	0.25	0.35	0.32	0.08	0.14	0.06	0.34	0.06	0.33	0.16	1.00						
SGP	0.41	0.33	0.33	0.27	0.14	0.18	0.25	0.27	0.49	0.07	0.29	0.20	0.18	0.32	0.26	0.15	1.00					
ESP	0.33	0.52	0.55	0.23	0.39	0.47	0.64	0.58	0.21	0.25	0.21	0.67	0.14	0.60	0.23	0.35	0.21	1.00				
SWE	0.38	0.50	0.51	0.26	0.30	0.49	0.55	0.50	0.26	0.20	0.20	0.48	0.19	0.55	0.22	0.30	0.30	0.53	1.00			
CHE	0.37	0.48	0.51	0.32	0.35	0.38	0.49	0.47	0.22	0.19	0.25	0.40	0.15	0.54	0.23	0.29	0.33	0.47	0.51	1.00		
GBR	0.33	0.35	0.42	0.26	0.33	0.29	0.42	0.39	0.23	0.31	0.16	0.36	0.19	0.42	0.22	0.23	0.26	0.37	0.42	0.42	1.00	
USA	0.27	0.35	0.37	0.27	0.07	0.11	0.18	0.24	0.24	0.11	0.45	0.17	0.14	0.28	0.15	0.07	0.32	0.16	0.22	0.27	0.23	1.00

Table A.5: ETF Premium and Illiquidity Risks

This table shows the results of a panel regression of ETF premiums on the currency illiquidity (IML) and global market illiquidity ($MILLIQ^G$), as well as monthly fixed effects and country fixed effects:

$$p_{i,t} - NAV_{i,t} = \alpha + \beta_1 IML_t + \beta_2 MILLIQ_t^G + FE_i + FE_t + \varepsilon_{i,t}$$

The results are presented for 2 versions of ETF premiums: the one based on Goetzmann et al. (2001) adjustment (FFP^G) and the one based on Engle and Sarkar (2006) adjustment (FFP^E). The IML ($MILLIQ^G$) is measured as a long-short portfolio of currency pairs (stocks) sorted by bid-ask spread (Amihud (2002) illiquidity ratio) and rebalanced daily. The regression is performed at a daily level. The sample is from June 2002- June 2018. ***,**,* show the significance at 1%, 5% and 10%.

Variables	FFP^G	FFP^E
	(1)	(2)
IML	0.094***	0.050***
	(4.73)	(6.41)
$MILLIQ^G$	0.078**	0.033***
	(2.16)	(3.35)
$Constant$	0.001	0.001
	1.05	0.84
Country FE	Yes	Yes
Time FE	Yes	Yes
Observations	75,915	75,915
Countries	22	22
Adjusted R^2	4.63	4.95

Appendix B

Supporting Documentation: Chapter 2

B.1 Madhavan and Sobczyk Model

In this appendix I briefly introduce the Madhavan and Sobczyk (2016) model of ETF price and NAV.

Unobservable expected value of the underlying assets is modelled as a random walk:

$$v_t = v_{t-1} + r_t, \qquad where\ r_t \sim (\mu_r, \sigma_r^2) \tag{B.1}$$

Price is the fundamental value plus a "true premium":

$$p_t = v_t + u_t \tag{B.2}$$

The true premium is represented as an autoregressive model with a coefficient ψ that represents the speed of error correction and a liquidity shock $\varepsilon_t \sim (\mu_\varepsilon, \sigma_\varepsilon^2)$.

$$u_t = \psi u_{t-1} + \varepsilon_t \tag{B.3}$$

Defining the official NAV of the fund as n_t I can show the premium at any point in time as:

$$\pi_t = p_t - n_t = (p_t - v_t) + (v_t - n_t) = u_t + (v_t - n_t) \tag{B.4}$$

The deviation of price from NAV can be due to staleness in NAV or due to the impact of secondary market on ETF price through shock ε and slow arbitrage $\psi > 0$. When $u_t = 0$ the entire premium represents the staleness in NAV and the deviation represents a price discovery in ETF market. The portion of variance not due to transitory component u_t is:

$$D = 1 - \left(\frac{\sigma_u}{\sigma_\pi}\right)^2, \qquad where\ \sigma_u = \frac{\sigma_\varepsilon}{\sqrt{1-\psi^2}} \tag{B.5}$$

This is defined as a price discovery component and is negatively related to variance of liquidity shock and is positively related to the speed of arbitrage.

B.2 Methods for Order Imbalance Construction

Lee and Ready (1991) provide an algorithm for classifying trades into buys and sells. Trade price is compared to prevailing quote. Prevailing quote is a current quote if it is older than 5 seconds. It is a quote 5 seconds ago, otherwise.

1. If price=bid - trade is classified as a sell trade
2. If price=ask - trade is classified as a buy trade
3. If price is at mid-point of bid-ask spread tick test is used:
   (a) If price is larger than of a previous trade price it is a buy trade
   (b) If price is smaller than of a previous trade price it is a sell trade
4. If price is inside bid-ask spread, but not at mid-quote classification is based on proximity to either bid or ask. Trades closer to the bid (ask) are sell (buy) trades.

Holden and Jacobsen (2014) provide an Interpolated Time technique to match trades and quotes happening within a millisecond. There are N trades and K orders happening in 1 millisecond and I know the order for trades and for quotes. The method assumes a uniform distribution of trades and quotes. Trade n in second s is assigned to time:

$$s + \frac{2n-1}{2N}, \qquad n = 1, 2 \dots N$$

Similarly, trade k in second s is assigned to time:

$$s + \frac{2k-1}{2K}, \qquad k = 1, 2 \dots K$$

B.3 Figures and Tables

Figure B.1: GNI per capita and Liquidity Mismatch

Scatter plot of average Gross National Income (GNI) per capita and liquidity mismatch (as defined in equation 2.12) for 41 countries over the sample period of 2006-2018. Trend line is shown in red. Regression and adjusted R^2 is provided. GNI per capita is from World Bank database and is expressed in 10^4.

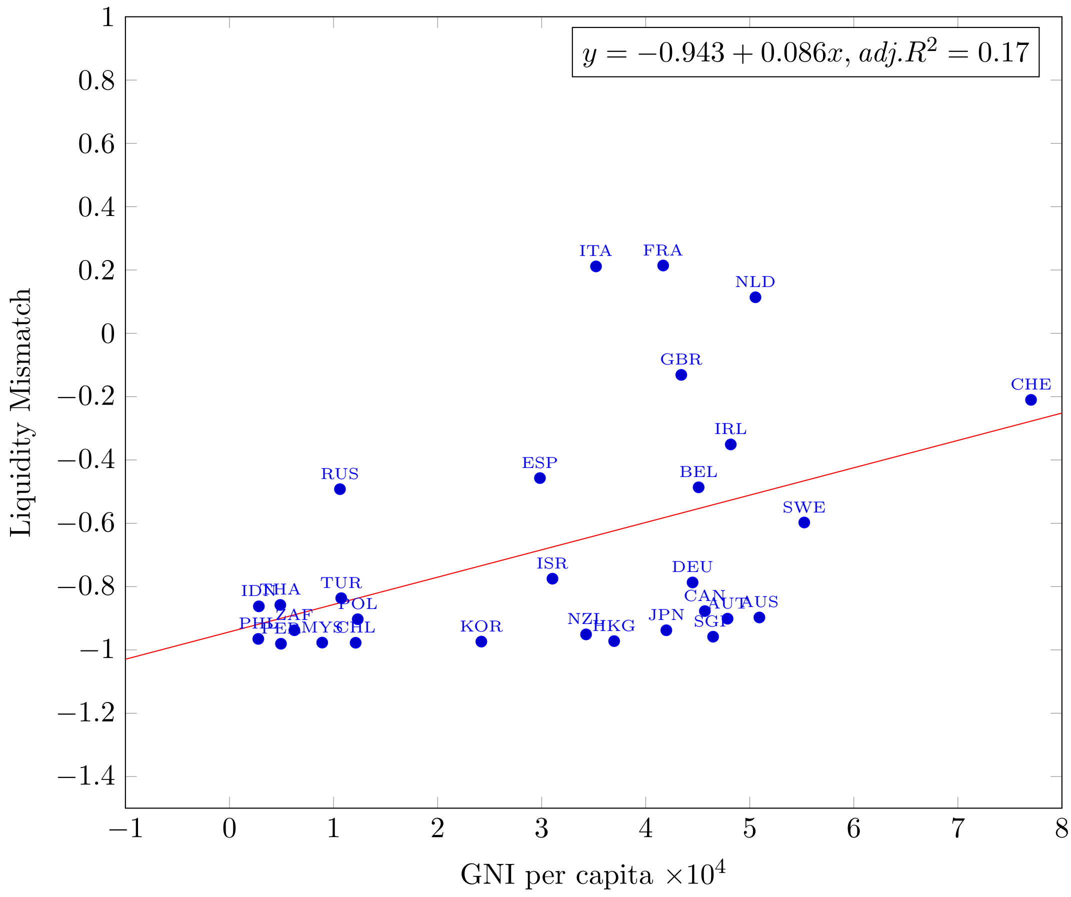
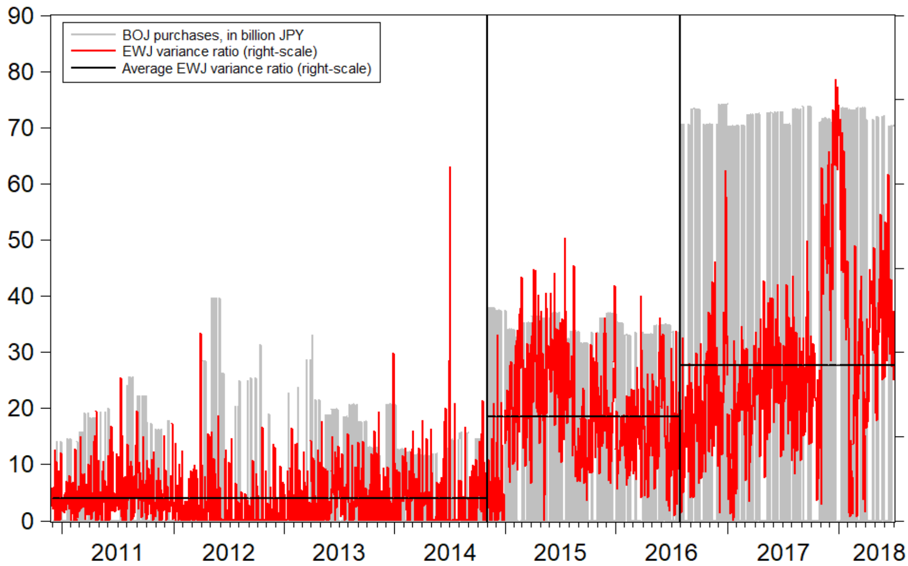


Figure B.2: BOJ ETF Purchases

The first panel shows $|VR - 1|$, where VR (variance ratio) is the ratio of 15-sec EWJ returns over three times of 5-sec EWJ return variances. The horizontal lines show the mean variance ratio ($|VR - 1|$) over three episodes, from 30/11/2010 to 31/10/2014, from 1/11/2014 to 29/07/2016 and 30/07/2016 to 29/06/2018. The vertical lines indicate the BOJ announcements of annual target changes. The second panel shows $|VR - 1|$, where VR (variance ratio) is the ratio of 15-sec SPY returns over three times of 5-sec SPY return variances. The grey bars exhibit the actual purchases in billion yen. EWJ (SPY) is the ETF that tracts MSCI Japan index (S&P 500). Both ETFs trade on NYSE Arca platform.



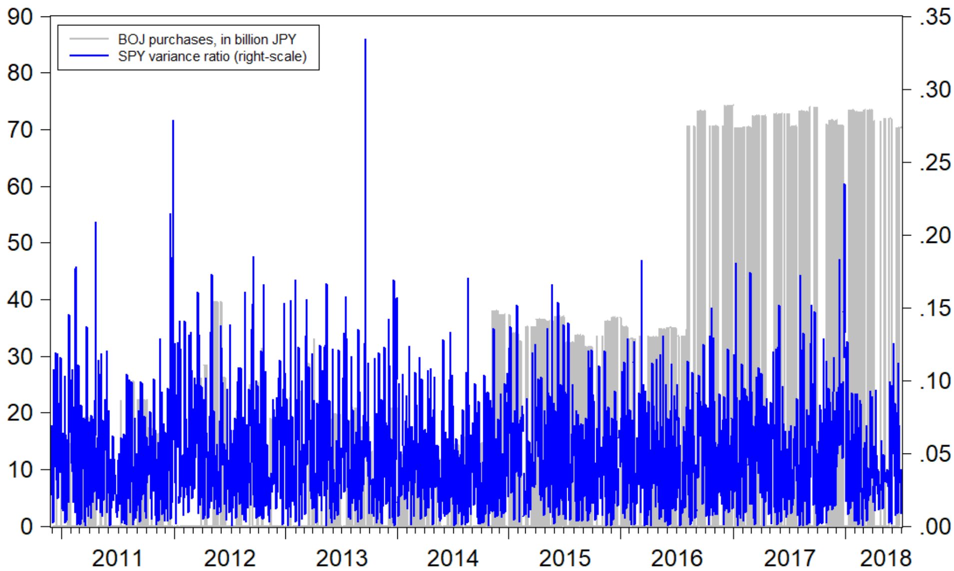


Figure B.3: BOJ ETF Purchase Days, VIX and Price Distortions

The first (second) pair of bar charts show $|VR - 1|$ (y-axis on the left), where VR (variance ratio) is the ratio of 15-sec EWJ (SPY) returns over three times of 5-sec EWJ (SPY) return variances, on non-intervention and BOJ invention days. The last pair of bar charts shows the daily change in VIX (y-axis on the right), on non-intervention and BOJ intervention days. The reported p-values indicate the significance of two-sample t-test comparing non-intervention and BOJ intervention days. Both ETFs trade on NYSE Arca platform.

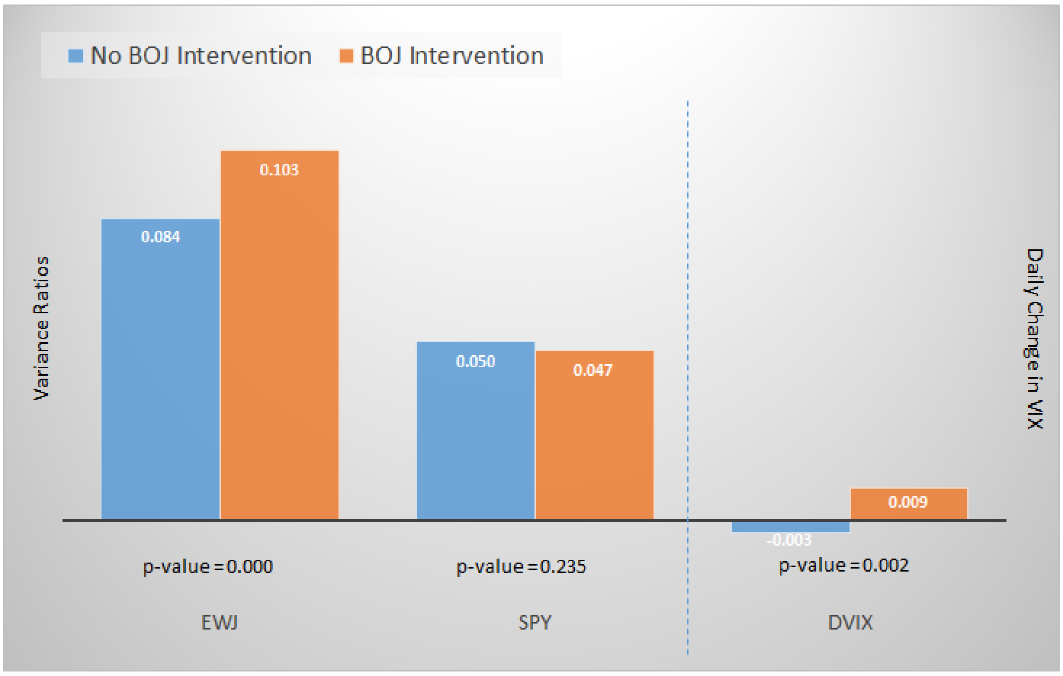


Figure B.4: BOJ ETF Purchase Days and ETF Correlations

The first (second) pair of bar charts show (annualized) daily returns of EWJ (SPY) ETF (y-axis on the left), on non-intervention and BOJ invention days. The last pair of bar charts shows the intraday correlation of EWJ and SPY ETF obtained by using 5-minute midquotes (y-axis on the right), on non-intervention and BOJ invention days. The reported p-values indicate the significance of two-sample t-test comparing non-intervention and BOJ intervention days. Both ETFs trade on NYSE Arca platform.

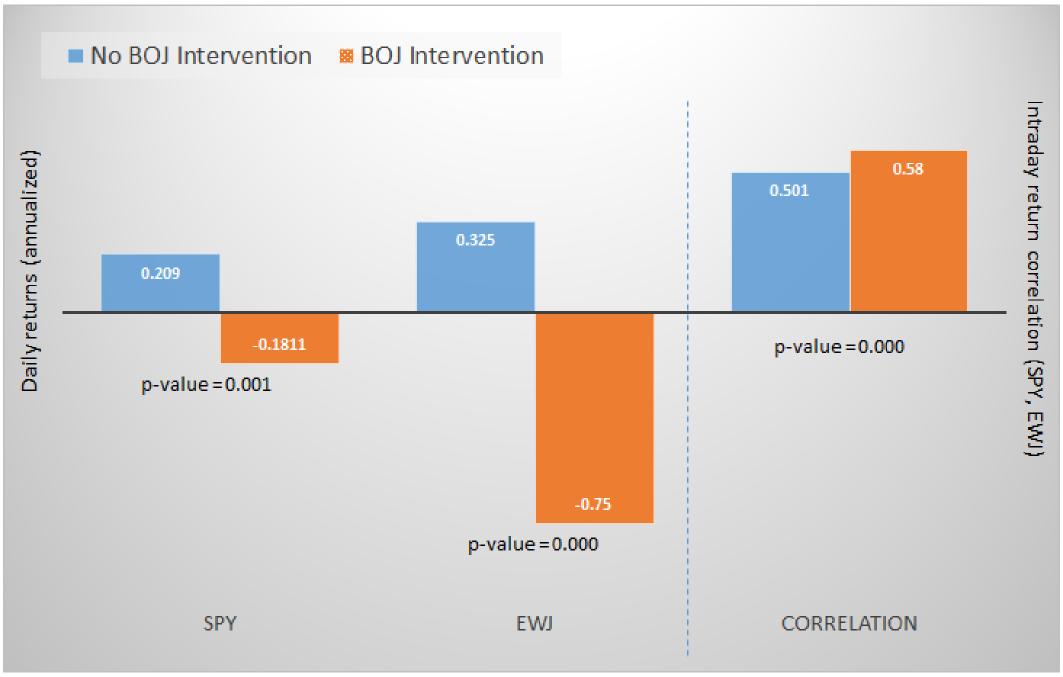


Table B.1: ETF Details

Countries in the sample, corresponding iShares country-level ETFs, their tickers, local market indices that funds track and the version of local volatility indices. Not all local market volatility indices are available. For some countries, general European index VSTOXX is used as a substitute. LVIX data is from Bloomberg.

Country	Ticker	Name: iShares MSCI	Tracking Index: MSCI	Volatility Index	Inception Date
AUS	EWA	Australia	Australia Index	S&P/ASX 200	12/03/1996
AUT	EWO	Austria Capped	Austria IMI 25/50	VSTOXX	12/03/1996
BEL	EWK	Belgium Capped	Belgium IMI 25/50	BEL 20	12/03/1996
BRA	EWZ	Brazil Capped	Brazil 25/50	-	10/07/2000
CAN	EWC	Canada	Canada Index	S&P/TSX 60 VIX	12/03/1996
CHL	ECH	Chile Capped	Chile IMI 25/50	-	12/11/2007
CHN	MCHI	China	China Index	ALPHASHARES CHINA	29/03/2011
COL	ICOL	Colombia	All Colombia Capped Index	-	18/06/2013
DNK	EDEN	Denmark	Denmark IMI 25/50 Index	-	25/01/2012
FIN	EFNL	Finland	Finland IMI 25/50 Index	VSTOXX	25/01/2012
FRA	EWQ	France	France Index	CAC40	12/03/1996
DEU	EWG	Germany	Germany Index	VDAX-NEW	12/03/1996
HKG	EWH	Hong Kong	Hong Kong Index	HSI	12/03/1996
IND	INDA	India	India Index	India VIX	02/02/2012
IDN	EIDO	Indonesia	Indonesia IMI	-	05/05/2010
IRL	EIRL	Ireland Capped	All Ireland Capped Index	VSTOXX	05/05/2010
ISR	EIS	Israel Capped	Israel Capped IMI	-	26/03/2008
ITA	EWI	Italy Capped	Italy 25/50	VSTOXX	12/03/1996
JPN	EWJ	Japan	Japan Index	NIKKEI STOCK AVERAGE	12/03/1996
MYS	EWM	Malaysia	Malaysia Index	-	12/03/1996
MEX	EWW	Mexico Capped	Mexico IMI 25/50	MEXICO	12/03/1996
NLD	EWN	Netherlands	Netherlands IMI	AEX	12/03/1996
NZL	ENZL	New Zealand Capped	New Zealand IMI 25/50	-	01/09/2010
NOR	ENOR	Norway	Norway IMI 25/50 Index	-	23/01/2012
PER	EPU	All Peru Capped	All Peru Capped Index	-	19/06/2009
PHL	EPHE	Philippines	Philippines IMI	-	28/09/2010
POL	EPOL	Poland Capped	Poland IMI 25/50	-	25/05/2010
QAT	QAT	Qatar	All Qatar Capped Index	-	29/04/2014
RUS	ERUS	Russia Capped	Russia 25/50 Index	RTS	09/11/2010
SAU	KSA	Saudi Arabia	Saudi Arabia IMI 25/50 Index	-	16/09/2015
SGP	EWS	Singapore	Singapore Index	-	12/03/1996
ZAF	EZA	South Africa	South Africa Index	SOUTH AFRICA	03/02/2003
KOR	EWY	South Korea	Korea 25/50 Index	VKOSPI	09/05/2000
ESP	EWP	Spain Capped	Spain 25/50	VSTOXX	12/03/1996
SWE	EWD	Sweden	Sweden Index	SIXVX	13/03/1996
CHE	EWL	Switzerland Capped	Switzerland 25/50	VSMI	12/03/1996
TWN	EWT	Taiwan Capped	Taiwan 25/50 Index	-	20/06/2000
THA	THD	Thailand Capped	Thailand IMI 25/50	-	26/03/2008
TUR	TUR	Turkey	Turkey IMI	-	26/03/2008
ARE	UAE	UAE	All UAE Capped Index	-	29/04/2014
GBR	EWU	United Kingdom	United Kingdom Index	FTSE 100	12/03/1996

Table B.2: Summary Statistics of VIX and LVIX

Summary statistics of daily changes in CBOE volatility index (VIX) and local alternatives (LVIX) for the period of 2006-2018. Details for LVIX are available in table B.1.

Country	Corr(ΔVIX, $\Delta LVIX$)	p-val	Mean	Std
AUS	0.145	< 0.01	0.19%	6.69%
AUT	0.537	< 0.01	0.22%	6.65%
BEL	0.422	< 0.01	0.25%	6.39%
BRA	–	–	–	–
CAN	0.331	< 0.01	0.42%	10.27%
CHL	–	–	–	–
CHN	0.475	< 0.01	0.12%	5.06%
COL	–	–	–	–
DNK	–	–	–	–
FIN	0.537	< 0.01	0.22%	6.65%
FRA	0.252	< 0.01	0.50%	13.53%
DEU	0.514	< 0.01	0.18%	6.08%
HKG	0.182	< 0.01	0.17%	5.81%
IND	0.168	< 0.01	0.13%	5.93%
IDN	–	–	–	–
IRL	0.537	< 0.01	0.22%	6.65%
ISR	–	–	–	–
ITA	0.537	< 0.01	0.22%	6.65%
JPN	0.138	< 0.01	0.20%	6.71%
MYS	–	–	–	–
MEX	0.444	< 0.01	0.11%	4.98%
NLD	0.452	< 0.01	0.26%	7.57%
NZL	–	–	–	–
NOR	–	–	–	–
PER	–	–	–	–
PHL	–	–	–	–
POL	–	–	–	–
QAT	–	–	–	–
RUS	0.221	< 0.01	0.23%	7.06%
SAU	–	–	–	–
SGP	–	–	–	–
ZAF	0.219	< 0.01	0.04%	3.05%
KOR	0.159	< 0.01	0.15%	5.78%
ESP	0.537	< 0.01	0.22%	6.65%
SWE	0.405	< 0.01	0.29%	7.62%
CHE	0.454	< 0.01	0.16%	5.60%
TWN	–	–	–	–
THA	–	–	–	–
TUR	–	–	–	–
ARE	–	–	–	–
GBR	0.450	< 0.01	0.27%	7.41%

Table B.3: Summary Statistics of USPU and LPU

Summary statistics of a proxy for U.S. political uncertainty ($USPU$) and local alternatives (LPU) for the period of 2006-2018. Details for political uncertainty construction are available in 2.3.4.

Country	Corr($USPU, LPU$)	p-val	Corr($\Delta LVIX, LPU$)	p-val	Mean	Std
AUS	0.348	< 0.01	0.043	0.03	0.61%	15.71%
AUT	0.392	< 0.01	0.031	0.09	0.08%	15.91%
BEL	0.430	< 0.01	-0.005	0.86	0.61%	16.64%
BRA	0.469	< 0.01	–	–	-0.03%	14.83%
CAN	0.513	< 0.01	-0.012	0.58	0.17%	15.39%
CHL	0.473	< 0.01	–	–	-0.31%	15.27%
CHN	0.396	< 0.01	0.007	0.79	-1.64%	15.88%
COL	0.435	< 0.01	–	–	-2.09%	18.58%
DNK	0.500	< 0.01	0.060	0.02	-3.13%	16.62%
FIN	0.533	< 0.01	0.086	0.01	-3.23%	16.24%
FRA	0.454	< 0.01	0.007	0.72	0.23%	15.79%
DEU	0.464	< 0.01	0.025	0.17	0.41%	15.72%
HKG	0.337	< 0.01	0.022	0.22	0.39%	16.16%
IND	0.503	< 0.01	0.016	0.53	-3.03%	16.54%
IDN	0.372	< 0.01	–	–	-1.3%	16.06%
IRL	0.462	< 0.01	0.050	0.02	-1.77%	14.14%
ISR	0.387	< 0.01	–	–	-0.63%	16.32%
ITA	0.442	< 0.01	–	–	0.51%	15.89%
JPN	0.365	< 0.01	0.019	0.29	0.86%	16.09%
MYS	0.317	< 0.01	–	–	0.90%	14.88%
MEX	0.540	< 0.01	-0.001	0.94	-0.02%	13.98%
NLD	0.471	< 0.01	0.022	0.22	0.19%	15.76%
NZL	0.406	< 0.01	–	–	-0.99%	13.88%
NOR	0.500	< 0.01	0.085	0.01	-3.03%	16.72%
PER	0.408	< 0.01	–	–	-0.52%	15.30%
PHL	0.387	< 0.01	–	–	-1.19%	14.93%
POL	0.449	< 0.01	–	–	-1.87%	14.84%
QAT	0.595	< 0.01	–	–	-3.66%	18.07%
RUS	0.380	< 0.01	-0.018	0.49	-1.62%	14.67%
SAU	0.512	< 0.01	–	–	-4.82%	14.43%
SGP	0.404	< 0.01	–	–	0.55%	15.38%
ZAF	0.450	< 0.01	0.040	0.03	0.05%	14.60%
KOR	0.387	< 0.01	0.031	0.09	0.24%	13.50%
ESP	0.457	< 0.01	0.052	0.01	0.25%	14.90%
SWE	0.425	< 0.01	0.037	0.04	0.51%	16.78%
CHE	0.394	< 0.01	0.043	0.02	0.68%	14.85%
TWN	0.395	< 0.01	–	–	0.75%	16.07%
THA	0.426	< 0.01	–	–	-0.45%	15.38%
TUR	0.436	< 0.01	–	–	-0.96%	14.57%
ARE	0.561	< 0.01	–	–	-3.15%	13.28%
GBR	0.455	< 0.01	0.053	0.01	0.19%	15.71%

Table B.4: Panel Results: Contemporaneous regression

Panel regressions with random effects for contemporaneous regression of order imbalance for 41 country-level MSCI based iShares ETFs on percentage change in volatility index (ΔVIX), percentage change in orthogonalised local volatility index ($\Delta LVIX^o$), change in local interest rates (IR), lagged order imbalance (OI_{t-1}), dummy variable for a common language (L) and a distance to the U.S. (G), as well as lags of ΔVIX and $\Delta LVIX^o$. Specifically, I estimate the model below:

$$OI_{i,t} = \alpha + \sum_{k=0}^{5} \beta_k \Delta VIX_{t-k} + \sum_{k=0}^{2} \gamma_k \Delta LVIX^o_{i,t-k} + \delta_1 \Delta IR_{i,t-1} + \delta_2 L_i + \delta_3 G_i + \delta_4 TOI_{i,t-1} + \varepsilon$$

The number of lags in the model is determined using Akaike Information Criterion and Bayesian information criterion jointly. Coefficients and z-statistics (based on robust standard errors) are presented. The lag for predictive regression is 1 day. Data is at a daily frequency and covers the period from 2006 or the first trading day of ETF (whichever is later) until the end of June 2018. Z-statistics is reported for difference in coefficients test $H_o : \beta_1 \geq \gamma_1$.***,**,* denote significance at the 1%, 5% and 10% level.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
VARIABLES	*totalOI*	*totalOI*	*totalOI*	*smallOI*	*smallOI*	*smallOI*	*retailOI*	*retailOI*	*retailOI*	*largeOI*	*largeOI*	*largeOI*
ΔVIX_t	-0.221***	-0.221***	-0.243***	-0.178***	-0.179***	-0.192***	-0.282***	-0.280***	-0.309***	-0.220***	-0.220***	-0.247***
	(-10.37)	(-10.37)	(-10.76)	(-7.34)	(-7.32)	(-7.55)	(-8.90)	(-8.80)	(-9.46)	(-9.17)	(-9.14)	(-9.75)
$\Delta LVIX^o_t$	-0.077**	-0.077**	-0.003	-0.063***	-0.063***	-0.017	-0.187***	-0.187***	-0.085**	-0.101**	-0.101**	-0.006
	(-2.34)	(-2.35)	(-0.10)	(-2.70)	(-2.70)	(-0.87)	(-3.94)	(-3.94)	(-2.54)	(-2.37)	(-2.37)	(-0.20)
ΔIR_{t-1}		-0.004**	-0.004**		-0.002	-0.002		-0.006*	-0.006*		-0.002	-0.002
		(-2.15)	(-2.28)		(-1.46)	(-1.53)		(-1.87)	(-1.87)		(-0.68)	(-0.78)
OI_{t-1}	0.174***	0.174***	0.170***	0.197***	0.197***	0.194***	0.135***	0.134***	0.130***	0.117***	0.117***	0.114***
	(12.46)	(12.42)	(12.30)	(8.21)	(8.24)	(8.19)	(8.85)	(9.03)	(8.91)	(11.05)	(10.96)	(10.62)
L		0.003	0.003		-0.005**	-0.005**		0.013	0.013		0.001	0.001
		(0.35)	(0.35)		(-2.28)	(-2.27)		(0.86)	(0.86)		(0.13)	(0.12)
G		-0.000	-0.000		-0.000	-0.000		-0.000**	-0.000**		0.000	0.000
		(-0.20)	(-0.19)		(-0.62)	(-0.61)		(-2.53)	(-2.53)		(0.15)	(0.15)

Continue on the next page

Table continued from previous page

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
VARIABLES	*totalOI*	*totalOI*	*totalOI*	*smallOI*	*smallOI*	*smallOI*	*retailOI*	*retailOI*	*retailOI*	*largeOI*	*largeOI*	*largeOI*
ΔVIX_{t-1}			-0.202***			-0.127***			-0.265***			-0.261***
			(-7.56)			(-5.19)			(-9.34)			(-8.67)
$\Delta LVIX^o_{t-1}$	-0.077***	-0.077***	-0.043**	-0.063***	-0.062***	-0.038**	-0.085***	-0.086***	-0.022	-0.085**	-0.085**	-0.043
	(-3.08)	(-3.07)	(-2.12)	(-3.46)	(-3.45)	(-2.44)	(-3.04)	(-3.04)	(-0.93)	(-2.45)	(-2.45)	(-1.46)
ΔVIX_{t-2}			-0.045***			-0.036***			-0.114***			-0.058**
			(-3.15)			(-2.84)			(-5.09)			(-2.28)
$\Delta LVIX^o_{t-2}$	-0.076***	-0.076***	-0.051**	-0.037	-0.037	-0.024	-0.078**	-0.079**	-0.049	-0.094***	-0.094***	-0.068**
	(-2.60)	(-2.60)	(-2.18)	(-1.48)	(-1.48)	(-1.09)	(-2.12)	(-2.12)	(-1.63)	(-2.87)	(-2.89)	(-2.53)
ΔVIX_{t-3}			-0.064***			-0.034***			-0.070***			-0.070***
			(-3.72)			(-3.33)			(-3.18)			(-3.06)
ΔVIX_{t-4}			-0.082***			-0.066***			-0.095***			-0.096***
			(-5.22)			(-4.03)			(-5.49)			(-4.56)
ΔVIX_{t-5}			-0.054***			-0.038***			-0.039**			-0.066***
			(-3.42)			(-2.64)			(-2.10)			(-2.66)
Constant	0.014***	0.015**	0.016**	0.000	0.004	0.004	0.010	0.039***	0.041***	0.010***	0.009	0.010
	(4.08)	(2.03)	(2.21)	(0.35)	(1.15)	(1.37)	(1.37)	(3.11)	(3.23)	(3.06)	(1.24)	(1.50)
Observations	57,894	57,703	57,661	52,789	52,789	52,556	54,620	54,620	54,433	55,550	55,550	55,324
Adjusted R^2	0.033	0.033	0.036	0.042	0.042	0.043	0.020	0.020	0.023	0.016	0.016	0.019

Table B.5: Panel Results- Developed Markets

Panel regressions with random effects for predictive regression of order imbalance for 21 developed market (classified by MSCI) country-level MSCI based iShares ETFs region on percentage change in volatility index (ΔVIX), percentage change in orthogonalised local volatility index ($\Delta LVIX^o$), change in local interest rates (IR), lagged order imbalance (OI_{t-1}), dummy variable for a common language (L) and a distance to the US (G), as well as lags of ΔVIX and $\Delta LVIX^o$.

$$OI_{i,t} = \alpha + \sum_{k=1}^{5} \beta_k \Delta VIX_{t-k} + \sum_{k=1}^{2} \gamma_k \Delta LVIX^o_{i,t-k} + \delta_1 \Delta IR_{i,t-1} + \delta_2 OI_{i,t-1} + \delta_3 L_i + \delta_4 G_i + \varepsilon_{i,t}$$

The number of lags in the model is determined using Akaike Information Criterion and Bayesian information criterion jointly. Coefficients and z-statistics (based on robust standard errors) are presented. The lag for predictive regression is 1 day. Data is at a daily frequency and covers the period from 2006 or the first trading day of ETF (whichever is later) until the end of June 2018. ***,**,* denote significance at the 1%, 5% and 10% level.

	(1)	(2)	(3)	(4)
VARIABLES	*TOI*	*smallOI*	*retailOI*	*largeOI*
ΔVIX_{t-1}	-0.203***	-0.123***	-0.266***	-0.261***
	(-6.63)	(-4.12)	(-8.77)	(-6.97)
$\Delta LVIX^o_{t-1}$	-0.041*	-0.043**	0.005	-0.037
	(-1.67)	(-2.57)	(0.15)	(-1.13)
ΔIR_{t-1}	-0.004**	-0.002*	-0.006*	-0.002
	(-2.29)	(-1.70)	(-1.89)	(-0.90)
OI_{t-1}	0.171***	0.205***	0.138***	0.114***
	(10.52)	(7.25)	(7.42)	(9.04)
L	0.001	-0.005**	0.018	-0.002
	(0.10)	(-2.09)	(1.19)	(-0.20)
G	-0.000	-0.000	-0.000***	-0.000
	(-1.20)	(-0.74)	(-3.39)	(-0.35)
ΔVIX_{t-2}	-0.044***	-0.022	-0.085***	-0.062**
	(-2.90)	(-1.52)	(-3.30)	(-2.05)
$\Delta LVIX^o_{t-2}$	-0.054**	-0.034	-0.050*	-0.065**
	(-2.01)	(-1.40)	(-1.90)	(-2.15)
ΔVIX_{t-3}	-0.061***	-0.025**	-0.056**	-0.077***
	(-2.82)	(-2.08)	(-2.11)	(-2.75)
ΔVIX_{t-4}	-0.090***	-0.066***	-0.080***	-0.100***
	(-5.03)	(-3.50)	(-5.01)	(-4.14)
ΔVIX_{t-5}	-0.053***	-0.037**	-0.029	-0.076***
	(-2.89)	(-2.30)	(-1.24)	(-2.66)
Constant	0.021***	0.004	0.048***	0.014**
	(2.80)	(1.19)	(3.18)	(1.99)
Observations	44,007	43,192	41,646	41,997
Adjusted R^2	0.033	0.045	0.022	0.017

Table B.6: Panel Results- Emerging Markets

Panel regressions with random effects for predictive regression of order imbalance for 21 emerging markets (classified by MSCI) country-level MSCI based iShares ETFs from Asia Pacific region on percentage change in volatility index (ΔVIX), percentage change in orthogonalised local volatility index ($\Delta LVIX^o$), change in local interest rates (IR), lagged order imbalance (OI_{t-1}), dummy variable for a common language (L) and a distance to the US (G), as well as lags of ΔVIX and $\Delta LVIX^o$. Specifically, I run the following model:

$$OI_{i,t} = \alpha + \sum_{k=1}^{5} \beta_k \Delta VIX_{t-k} + \sum_{k=1}^{2} \gamma_k \Delta LVIX^o_{i,t-k} + \delta_1 \Delta IR_{i,t-1} + \delta_2 OI_{i,t-1} + \delta_3 L_i + \delta_4 G_i + \varepsilon_{i,t}$$

The number of lags in the model is determined using Akaike Information Criterion and Bayesian information criterion jointly. Coefficients and z-statistics (based on robust standard errors) are presented. The lag for predictive regression is 1 day. Data is at a daily frequency and covers the period from 2006 or the first trading day of ETF (whichever is later) until the end of June 2018. ***,**,* denote significance at the 1%, 5% and 10% level.

	(1)	(2)	(3)	(4)
VARIABLES	*TOI*	*smallOI*	*retailOI*	*largeOI*
ΔVIX_{t-1}	-0.120***	-0.088***	-0.248***	-0.184***
	(-4.04)	(-3.37)	(-3.76)	(-5.59)
$\Delta LVIX^o_{t-1}$	-0.053***	0.014	-0.095	-0.062
	(-4.04)	(0.45)	(-1.36)	(-1.40)
ΔIR_{t-1}	0.049	0.079***	-0.156***	0.086*
	(1.18)	(6.93)	(-3.02)	(1.84)
OI_{t-1}	0.165***	0.133***	0.096***	0.110***
	(7.67)	(6.28)	(6.47)	(5.86)
L	0.007	-0.001	-0.017	0.010
	(0.21)	(-0.47)	(-0.32)	(0.39)
G	0.000	-0.000	0.000	0.000
	(0.48)	(-1.05)	(0.17)	(0.32)
ΔVIX_{t-2}	0.011	-0.044	-0.157***	0.009
	(0.48)	(-1.38)	(-3.96)	(0.21)
$\Delta LVIX^o_{t-2}$	-0.052	0.034	-0.063	-0.093**
	(-0.93)	(0.54)	(-0.58)	(-2.13)
ΔVIX_{t-3}	-0.040*	-0.040**	-0.082***	-0.017
	(-1.85)	(-2.54)	(-2.96)	(-0.51)
ΔVIX_{t-4}	-0.015	-0.023	-0.096*	-0.040
	(-1.13)	(-1.34)	(-1.95)	(-1.29)
ΔVIX_{t-5}	-0.018	-0.006	-0.023	0.002
	(-0.65)	(-0.24)	(-1.08)	(0.03)
Constant	0.004	0.004***	0.016	0.001
	(0.43)	(3.56)	(0.84)	(0.15)
Observations	13,676	9,386	12,809	13,349
Adjusted R^2	0.027	0.019	0.013	0.014

Table B.8: Average Monthly Correlations

Per country average monthly correlation of changes in MSCI index and changes in S&P 500. RW-36 (100) shows the results for partial rolling window correlation measure using 36 (100)-month estimation period. RW-36$[m, x]$ ($[IP]$) shows the 36-month partial-correlation controlling for export and import ratios with the U.S. (for percentage changes in industrial production index of the U.S. and a local country from month t-12 to month t). DCC shows the results for Dynamic Conditional Correlation model of Engle (2002) using GARCH(1,1) and constant mean for returns. G shows the geographical distance between capitals of corresponding countries and the capital of the U.S.. L is a dummy variable, which takes the value of 1 if the country has a common language with the U.S. and 0 otherwise. G and L measures are from CEPII's Geodist database.

Country	RW-36	RW-100	RW-36 $[m, x]$	RW-36 $[IP]$	DCC	G	L
AUS	0.77	0.81	0.76	0.77	0.69	15962	1
AUT	0.76	0.81	0.76	0.77	0.72	7130	0
BEL	0.79	0.81	0.79	0.80	0.77	6223	0
BRA	0.65	0.65	0.66	0.66	0.56	6794	0
CAN	0.79	0.81	0.77	0.79	0.70	737	1
CHL	0.54	0.54	0.54	0.54	0.46	8081	0
CHN	0.55	0.64	0.53	0.48	0.57	11159	0
COL	0.31	0.32	0.32	0.31	0.33	3815	0
DNK	0.57	0.65	0.57	0.57	0.62	6519	0
FIN	0.71	0.71	0.70	0.70	0.66	6943	0
FRA	0.82	0.85	0.83	0.83	0.77	6169	0
DEU	0.85	0.86	0.85	0.84	0.79	6718	0
HKG	0.68	0.69	0.68	0.68	0.63	13131	1
IND	0.50	0.52	0.50	0.51	0.52	12060	1
IDN	0.42	0.42	0.42	0.40	0.36	16371	0
IRL	0.68	0.77	0.68	0.69	0.70	5449	1
ISR	0.66	0.71	0.66	0.67	0.66	9451	1
ITA	0.72	0.76	0.72	0.72	0.69	7225	0
JPN	0.67	0.68	0.66	0.67	0.62	10919	0
MYS	0.62	0.61	0.62	0.61	0.53	15357	0
MEX	0.75	0.80	0.74	0.75	0.66	3038	0
NLD	0.84	0.86	0.85	0.84	0.79	6197	0
NZL	0.63	0.60	0.61	0.63	0.62	14220	1
NOR	0.59	0.66	0.61	0.59	0.63	6240	0
PER	0.43	0.48	0.43	0.43	0.45	5673	0
PHL	0.48	0.50	0.48	0.49	0.44	13794	1
POL	0.56	0.71	0.57	0.55	0.62	7184	0
QAT	0.28	0.22	0.28	0.28	0.19	11113	0
RUS	0.55	0.66	0.59	0.57	0.59	7835	0
SAU	0.20	0.20	0.02	0.20	0.18	10859	0
SGP	0.73	0.74	0.73	0.72	0.69	15564	1
ZAF	0.65	0.70	0.66	0.65	0.61	13040	1
KOR	0.67	0.73	0.67	0.66	0.58	11186	0
ESP	0.70	0.70	0.69	0.70	0.64	6092	0
SWE	0.80	0.82	0.80	0.80	0.77	6644	0
CHE	0.80	0.80	0.80	0.80	0.76	6603	0
TWN	0.67	0.70	0.67	0.65	0.59	12659	0
THA	0.51	0.59	0.51	0.51	0.55	14174	0
TUR	0.47	0.56	0.45	0.48	0.49	8733	0
ARE	0.28	0.19	0.28	0.28	0.22	11359	0
GBR	0.85	0.87	0.85	0.85	0.78	5901	1

Table B.7: Recession Periods

Panel regressions with random effects for predictive regression of order imbalance for 41 country-level MSCI based iShares ETFs on percentage change in volatility index (ΔVIX), percentage change in orthogonalised local volatility index ($\Delta LVIX$), dummy for recession in local countries $D_{R,m}$, change in local interest rates (IR), lagged order imbalance (OI_{t-1}), dummy variable for a common language (L) and a distance to the US (G), as well as lags of ΔVIX and $\Delta LVIX^o$. D_R takes the value of 1 if there is a recession in a country during month m (time t is during month m). Specifically, I estimate the model below:

$$OI_{i,t} = \alpha + \sum_{k=1}^{5} \beta_k \Delta VIX_{t-k} + \sum_{k=1}^{2} \gamma_k \Delta LVIX^o_{i,t-k} + \delta_1 D_{R,m} + \delta_2 \Delta IR_{i,t-1} + \delta_3 OI_{i,t-1} + \delta_4 L_i + \delta_5 G_i + \varepsilon_{i,t}$$

The number of lags in the model is determined using Akaike Information Criterion and Bayesian information criterion jointly. Coefficients and z-statistics (based on robust standard errors) are presented. The lag for predictive regression is 1 day. Data is at a daily frequency and covers the period from 2006 or the first trading day of ETF (whichever is later) until the end of June 2018. The data sample is split into 3 sub-samples: period of low VIX ,period of medium VIX, and period of high VIX. The split is based on terciles of of historic VIX level distribution from Jan 1990 - June 2018. ***,**,* denote significance at the 1%, 5% and 10% level.

	(1)	(2)	(3)	(4)
VARIABLES	*TOI*	*smallOI*	*retailOI*	*largeOI*
ΔVIX_{t-1}	-0.214***	-0.144***	-0.296***	-0.281***
	(-7.45)	(-5.42)	(-8.60)	(-8.23)
$\Delta LVIX^o_{t-1}$	-0.031	-0.024	-0.003	-0.021
	(-1.32)	(-1.48)	(-0.10)	(-0.69)
$D_{R,m}$	-0.040***	-0.053***	-0.038***	-0.048***
	(-3.45)	(-5.41)	(-2.82)	(-3.96)
ΔIR_{t-1}	-0.005**	-0.005***	-0.008***	-0.002
	(-1.97)	(-2.73)	(-3.51)	(-1.05)
OI_{t-1}	0.178***	0.203***	0.126***	0.120***
	(11.27)	(8.60)	(8.21)	(10.35)
L	0.005	-0.005	0.018	0.005
	(0.44)	(-1.04)	(1.08)	(0.44)
G	0.000	0.000	-0.000	0.000
	(0.70)	(1.27)	(-0.92)	(0.97)
ΔVIX_{t-2}	-0.040**	-0.042***	-0.121***	-0.056*
	(-2.02)	(-2.61)	(-4.55)	(-1.79)
$\Delta LVIX^o_{t-2}$	-0.067**	-0.030	-0.047	-0.085**
	(-2.06)	(-1.13)	(-1.43)	(-2.49)
ΔVIX_{t-3}	-0.057***	-0.033**	-0.070**	-0.057**
	(-3.60)	(-2.23)	(-2.46)	(-2.31)
ΔVIX_{t-4}	-0.094***	-0.092***	-0.092***	-0.095***
	(-3.99)	(-3.95)	(-4.08)	(-3.08)
ΔVIX_{t-5}	-0.072***	-0.050***	-0.059***	-0.072**
	(-3.42)	(-2.79)	(-2.89)	(-2.30)
Constant	0.033***	0.031***	0.046***	0.030***
	(3.11)	(4.87)	(3.01)	(3.21)
Observations	44,572	39,509	41,664	42,654
Adjusted R^2	0.041	0.056	0.022	0.021

Table B.9: Price Discovery, Market Illiquidity and Trading Channel

Average 36-month partial rolling correlations of MSCI index returns on S&P 500 returns of 3 (or 4) monthly portfolios formed based on price discovery and limits to arbitrage proxies. Correlation is defined as $\rho_t^L = \frac{1}{N_t^L}\sum_{i=1}^{N_t^L} partialCorr(\Delta MSCI_i, \Delta S\&P500 \quad | \frac{m_{US,t}}{m_{Total,t}}, \frac{x_{US,t}}{x_{Total,t}})$, where m_{US} (x_{US}) is the import(export) to (from) U.S. and m_{Total} (x_{Total}) is the total import (export) of a country with all its trading partners. Each portfolio is presorted based on information available during previous month. Panel A shows the sort into terciles by β_i from equation $R_{i,t}^{NAV} = \alpha + \beta_i \left(\frac{P_{i,t-1} - NAV_{i,t-1}}{NAV_{i,t-1}}\right) + \varepsilon$. I use a 36-month period to estimate β_i. Sort in panel B is based on Amihud's illiquidity ratio (Amihud (2002)) of the underlying markets ($ILLIQ$). Panel C shows the result of a double sort by median based on proxies from Panel A and Panel B. Panel D shows the result of a double sort by median based on price discovery as in panel A and illiquidity mismatch as defined in equation 2.12. t-statistics is based on Newey and West (1987) robust standard errors

	Panel A: Price Discovery				
Sorting Variable	low	medium	high	*HML*	*t-stat*
NAV sensitivity to Premium	0.59	0.69	0.76	0.17	(29.49)
	Panel B: Limits to Arbitrage				
Sorting Variable	low	medium	high	*HML*	*t-stat*
Amihud's Illiquidity Ratio	0.75	0.69	0.65	-0.10	(-11.03)
	Panel C: Double Sort 1				
		Amihud's Illiquidity Ratio			
Sorting Variables		low	high	*HML*	*t-stat*
NAV sensitivity to Premium	low	0.67	0.60	-0.07	(-5.01)
	high	0.76	0.73	-0.03	(-3.23)
	HML	0.09	0.12		
	t-stat	(8.43)	(11.10)		
	Panel D: Double Sort 2				
		Liquidity Mismatch			
Sorting Variables		low	high	*HML*	*t-stat*
NAV sensitivity to Premium	low	0.61	0.65	0.06	(2.99)
	high	0.71	0.77	0.07	(6.15)
	HML	0.10	0.11		
	t-stat	(8.88)	(6.22)		

Table B.10: Price Discovery, Market Illiquidity and Business Cycles

Average 36-month partial rolling correlations of MSCI index returns on S&P 500 returns of 3 (or 4) monthly portfolios formed based on price discovery and limits to arbitrage proxies. Correlation is defined as $\rho_t^L = \frac{1}{N_t^L}\sum_{i=1}^{N_t^L} partialCorr(\Delta MSCI_i, \Delta S\&P500 \quad |\Delta IP_{i,t}, \Delta IP_{US,t})$, where $\Delta IP_{US,t}$ ($\Delta IP_{i,t}$) is the percentage change in industrial production index of the U.S. (a local country) from month $t-12$ to month t. Each portfolio is presorted based on information available during previous month. Panel A shows the sort into terciles by β_i from equation $R_{i,t}^{NAV} = \alpha + \beta_i \left(\frac{P_{i,t-1} - NAV_{i,t-1}}{NAV_{i,t-1}}\right) + \varepsilon$. I use a 36-month period to estimate β_i. Sort in panel B is based on Amihud's illiquidity ratio (Amihud (2002)) of the underlying markets (*ILLIQ*). Panel C shows the result of a double sort by median based on proxies from Panel A and Panel B. Panel D shows the result of a double sort by median based on price discovery as in panel A and illiquidity mismatch as defined in equation 2.12. t-statistics is based on Newey and West (1987) robust standard errors

	Panel A: Price Discovery				
Sorting Variable	low	medium	high	*HML*	*t-stat*
NAV sensitivity to Premium	0.58	0.69	0.76	0.18	(23.10)
	Panel B: Limits to Arbitrage				
Sorting Variable	low	medium	high	*HML*	*t-stat*
Amihud's Illiquidity Ratio	0.75	0.69	0.64	-0.11	(-11.19)
	Panel C: Double Sort 1				
		Amihud's Illiquidity Ratio			
Sorting Variables		low	high	*HML*	*t-stat*
NAV sensitivity to Premium	low	0.66	0.60	-0.06	(-5.17)
	high	0.76	0.73	-0.03	(-2.89)
	HML	0.10	0.13		
	t-stat	(8.24)	(11.13)		
	Panel D: Double Sort 2				
		Liquidity Mismatch			
Sorting Variables		low	high	*HML*	*t-stat*
NAV sensitivity to Premium	low	0.61	0.64	0.05	(2.87)
	high	0.71	0.78	0.07	(4.97)
	HML	0.10	0.12		
	t-stat	(8.58)	(6.34)		

Table B.11: Correlation, Developed Markets and Financial Development

Average 36-month rolling correlations of MSCI index returns on S&P 500 returns of 3 (or 4) monthly portfolios formed based on price discovery and limits to arbitrage proxies. Correlation is defined as $\rho_t^L = \frac{1}{N_t^L}\sum_{i=1}^{N_t^L} Corr(\Delta MSCI_i, \Delta S\&P500)$. Each portfolio is presorted based on information available during previous month. Using a sub-sample of G10 countries panel A shows the sort into terciles by β_i from equation $R_{i,t}^{NAV} = \alpha + \beta_i\left(\frac{P_{i,t-1} - NAV_{i,t-1}}{NAV_{i,t-1}}\right) + \varepsilon$. I use a 36-month period to estimate β_i. Sort in panel B is based on a sub-sample of MSCI developed countries. In panel C and D correlation is defined as $\rho_t^L = \frac{1}{N_t^L}\sum_{i=1}^{N_t^L} partialCorr(\Delta MSCI_i, \Delta S\&P500 \quad |FD_{i,t})$, where $FD_{i,t}$ is the financial development proxy measured as a ratio of stock market capitalisation over GDP. Using a full sample Panel C shows the result of a double sort by median based on price discovery from Panel A and Amihud's illiquidity ratio (Amihud (2002)) of the underlying markets (*ILLIQ*). Panel D shows the result of a double sort by median based on price discovery as in panel A and illiquidity mismatch as defined in equation 2.12. t-statistics is based on Newey and West (1987) robust standard errors

	Panel A: G10 countries				
Sorting Variable	low	medium	high	*HML*	*t-stat*
NAV sensitivity to Premium	0.72	0.80	0.81	0.09	(6.37)
	Panel B: MSCI Developed Countries				
Sorting Variable	low	medium	high	*HML*	*t-stat*
NAV sensitivity to Premium	0.71	0.74	0.78	0.08	(11.20)
	Panel C: Partial Correlation- Double Sort 1				
		Amihud's Illiquidity Ratio			
Sorting Variables		low	high	*HML*	*t-stat*
NAV sensitivity to Premium	low	0.67	0.61	-0.07	(-5.64)
	high	0.76	0.73	-0.03	(-3.00)
	HML	0.09	0.12		
	t-stat	(8.65)	(13.04)		
	Panel D: Partial Correlation- Double Sort 2				
		Liquidity Mismatch			
Sorting Variables		low	high	*HML*	*t-stat*
NAV sensitivity to Premium	low	0.61	0.64	0.05	(2.79)
	high	0.71	0.78	0.06	(4.86)
	HML	0.10	0.12		
	t-stat	(9.15)	(7.03)		

Table B.12: Price Discovery, Market Illiquidity and Expanded Rolling Window

Average 100-month rolling correlations of MSCI index returns on S&P 500 returns of 3 (or 4) monthly portfolios formed based on price discovery and limits to arbitrage proxies. Correlation is defined as $\rho_t^L = \frac{1}{N_t^L}\sum_{i=1}^{N_t^L} Corr(\Delta MSCI_i, \Delta S\&P500)$. Each portfolio is presorted based on information available during previous month. Panel A shows the sort into terciles by β_i from equation $R_{i,t}^{NAV} = \alpha + \beta_i \left(\frac{P_{i,t-1} - NAV_{i,t-1}}{NAV_{i,t-1}}\right) + \varepsilon$. I use a 100-month period to estimate β_i. Sort in panel B is based on Amihud's illiquidity ratio (Amihud (2002)) of the underlying markets ($ILLIQ$). Panel C shows the result of a double sort by median based on proxies from Panel A and Panel B. Panel D shows the result of a double sort by median based on price discovery as in panel A and illiquidity mismatch as defined in equation 2.12. t-statistics is based on Newey and West (1987) robust standard errors

	Panel A: Price Discovery				
Sorting Variable	low	medium	high	*HML*	*t-stat*
NAV sensitivity to Premium	0.55	0.67	0.79	0.24	(18.53)
	Panel B: Limits to Arbitrage				
Sorting Variable	low	medium	high	*HML*	*t-stat*
Amihud's Illiquidity Ratio	0.77	0.70	0.62	-0.08	(-12.05)
	Panel C: Double Sort 1				
		Amihud's Illiquidity Ratio			
Sorting Variables		low	high	*HML*	*t-stat*
NAV sensitivity to Premium	low	0.63	0.58	-0.05	(-7.09)
	high	0.78	0.75	-0.02	(-3.31)
	HML	0.15	0.18		
	t-stat	(25.77)	(37.44)		
	Panel D: Double Sort 2				
		Liquidity Mismatch			
Sorting Variables		low	high	*HML*	*t-stat*
NAV sensitivity to Premium	low	0.58	0.62	0.04	(2.45)
	high	0.75	0.78	0.04	(3.97)
	HML	0.16	0.16		
	t-stat	(20.99)	(9.61)		

Table B.13: Price Discovery, Market Illiquidity and DCC

Average DCC correlations of MSCI index returns on S&P 500 returns of 3 (or 4) monthly portfolios formed based on price discovery and limits to arbitrage proxies. Correlation estimates are computed using dynamic conditional correlation GARCH(1,1) model with constant mean. Each portfolio is presorted based on information available during previous month. Panel A shows the sort into terciles by β_i from equation $R_{i,t}^{NAV} = \alpha + \beta_i \left(\frac{P_{i,t-1} - NAV_{i,t-1}}{NAV_{i,t-1}} \right) + \varepsilon$. I use a 36-month period to estimate β_i. Sort in panel B is based on Amihud's illiquidity ratio (Amihud (2002)) of the underlying markets ($ILLIQ$). Panel C shows the result of a double sort by median based on proxies from Panel A and Panel B. Panel D shows the result of a double sort by median based on price discovery as in panel A and illiquidity mismatch as defined in equation 2.12. t-statistics is based on Newey and West (1987) robust standard errors

	Panel A: Price Discovery				
Sorting Variable	low	medium	high	*HML*	*t-stat*
NAV sensitivity to Premium	0.52	0.64	0.71	0.18	(18.12)
	Panel B: Limits to Arbitrage				
Sorting Variable	low	medium	high	*HML*	*t-stat*
Amihud's Illiquidity Ratio	0.70	0.64	0.59	-0.11	(-10.88)
	Panel C: Double Sort 1				
		Amihud's Illiquidity Ratio			
Sorting Variables		low	high	*HML*	*t-stat*
NAV sensitivity to Premium	low	0.62	0.54	-0.08	(-7.11)
	high	0.71	0.68	-0.03	(-3.49)
	HML	0.09	0.14		
	t-stat	(9.60)	(11.19)		
	Panel D: Double Sort 2				
		Liquidity Mismatch			
Sorting Variables		low	high	*HML*	*t-stat*
NAV sensitivity to Premium	low	0.54	0.60	0.06	(4.26)
	high	0.66	0.72	0.06	(3.71)
	HML	0.12	0.13		
	t-stat	(6.60)	(11.47)		

Table B.14: Cash ETFs and VIX

Predictive regression of average order imbalance for U.S. cash ETFs ($cashOI$) on percentage change in volatility index (ΔVIX), orthogonalised percentage change in local volatility index ($\Delta LVIX^o$), change in local interest rates (IR), lagged order imbalance (OI_{t-1}), as well as lags of ΔVIX and $\Delta LVIX^o$. In particular, I estimate the model below:

$$CashOI_{i,t} = \alpha + \sum_{k=1}^{5} \beta_k \Delta VIX_{t-k} + \sum_{k=1}^{3} \gamma_k \Delta LVIX^o_{i,t-k} + \gamma_5 \Delta LVIX^o_{i,t-5} + \delta_1 \Delta IR_{i,t-1} + \delta_2 CashOI_{i,t-1} + \varepsilon_{i,t}$$

Cash ETFs are defined as funds that invest in U.S. treasuries with less than 1 year of maturity. The number of lags in the model is determined using Akaike Information Criterion and Bayesian information criterion jointly. Coefficients and t-statistics (based on Newey and West (1987) robust standard errors) are presented. The lag for predictive regression is 1 day. Data is at a daily frequency and covers the period from 2006 or the first trading day of ETF (whichever is later) until the end of June 2018.***,**,* denote significance at the 1%, 5% and 10% level.

Country	Constant	ΔVIX_{t-1}	$\Delta LVIX^o_{t-1}$	ΔIR_{t-1}	$cashOI_{t-1}$	ΔVIX_{t-2}	$\Delta LVIX^o_{t-2}$	ΔVIX_{t-3}	$\Delta LVIX^o_{t-3}$	ΔVIX_{t-4}	ΔVIX_{t-5}	$\Delta LVIX^o_{t-5}$	Obs.	R^2
AUS	0.020**	0.289***	0.039	-0.065	0.301***	0.383***	0.002	0.233***	0.007	0.21***	0.257***	0.022	2,637	0.113
	(2.22)	(3.83)	(0.37)	(-0.27)	(13.28)	(3.84)	(0.02)	(2.74)	(0.07)	(2.58)	(3.37)	(0.22)		
AUT	0.039***	0.299***	0.05	0.191***	0.343***	0.368***	0.164	0.185**	0.096	0.227***	0.169*	0.06	2,886	0.140
	(4.17)	(3.94)	(0.42)	(3.22)	(13.01)	(4.15)	(1.24)	(2.31)	(0.88)	(2.68)	(1.9)	(0.55)		
BEL	0.096***	0.217	0.193	0.12	0.392***	0.451***	0.058	0.264	0.131	0.419**	0.045	0.184	966	0.176
	(5.58)	(1.38)	(0.77)	(0.72)	(8.83)	(2.85)	(0.28)	(1.62)	(0.68)	(2.47)	(0.22)	(1.06)		
BRA	0.039***	0.300***	-	-0.799***	0.344***	0.385***	-	0.229***	-	0.255***	0.176**	-	2,886	0.140
	(4.12)	(3.94)	-	(-4.55)	(13.05)	(4.27)	-	(3.3)	-	(3.09)	(2.00)	-		
CAN	0.007	0.290***	-0.037	0.034	0.276***	0.391***	-0.005	0.208***	0.114	0.179**	0.259***	0.016	2,196	0.097
	(0.74)	(3.59)	(-0.63)	(0.39)	(11.5)	(3.89)	(-0.06)	(2.76)	(1.56)	(2.20)	(3.24)	(0.25)		
CHL	0.039***	0.304***	-	-0.101***	0.343***	0.387***	-	0.229***	-	0.254***	0.173**	-	2,886	0.142
	(4.14)	(4.00)	-	(-3.02)	(13.02)	(4.28)	-	(3.31)	-	(3.07)	(1.97)	-		
CHN	0.035***	0.314***	-0.097	0.003	0.35***	0.478***	0.073	0.254***	-0.174	0.306***	0.197**	0.023	2,723	0.146
	(3.68)	(3.58)	(-0.61)	(0.08)	(13.05)	(5.65)	(0.46)	(3.12)	(-1.12)	(3.05)	(1.99)	(0.15)		
COL	0.015*	0.304***	-	0.007	0.299***	0.408***	-	0.244***	-	0.22***	0.248***	-	2,533	0.115
	(1.68)	(4.00)	-	(0.29)	(13.1)	(4.37)	-	(3.36)	-	(2.91)	(3.23)	-		
DNK	0.041***	0.303***	-	-0.002	0.344***	0.380***	-	0.224***	-	0.253***	0.162*	-	2,836	0.139
	(4.34)	(4.03)	-	(-0.21)	(12.88)	(4.17)	-	(3.18)	-	(3.03)	(1.85)	-		
FIN	0.039***	0.305***	0.041	0.019	0.343***	0.366***	0.146	0.187**	0.075	0.235***	0.168*	0.072	2,836	0.139
	(4.17)	(4.01)	(0.34)	(1.64)	(12.97)	(4.1)	(1.10)	(2.32)	(0.68)	(2.80)	(1.91)	(0.66)		
FRA	0.039***	0.307***	0.039	0.02*	0.341***	0.369***	0.063**	0.208***	0.051*	0.24***	0.169*	0.019	2,878	0.140
	(4.18)	(4.02)	(1.20)	(1.66)	(12.92)	(4.16)	(2.01)	(2.99)	(1.66)	(2.95)	(1.93)	(0.77)		
DEU	0.039***	0.304***	0.125	0.02*	0.343***	0.349***	0.143	0.189**	0.016	0.247***	0.168*	0.085	2,878	0.139
	(4.18)	(3.99)	(0.93)	(1.67)	(12.98)	(3.8)	(1.00)	(2.39)	(0.14)	(2.93)	(1.91)	(0.70)		
HKG	0.038***	0.303***	-0.016	0.202	0.342***	0.396***	-0.013	0.235***	-0.078	0.279***	0.181**	0.176*	2,886	0.140
	(4.11)	(4.00)	(-0.13)	(1.18)	(12.98)	(4.06)	(-0.11)	(3.13)	(-0.68)	(3.13)	(2.08)	(1.65)		
IND	0.022**	0.275***	-0.079	0.255***	0.304***	0.404***	0.132	0.217***	-0.06	0.222***	0.257***	0.021	2,678	0.116
	(2.45)	(3.67)	(-0.67)	(2.61)	(13.59)	(4.44)	(1.12)	(2.91)	(-0.54)	(2.91)	(3.43)	(0.2)		
IDN	0.038***	0.303***	-	0.294	0.343***	0.383***	-	0.228***	-	0.251***	0.175**	-	2,886	0.140
	(4.12)	(3.99)	-	(1.41)	(13.00)	(4.27)	-	(3.28)	-	(3.03)	(1.98)	-		
IRL	0.039***	0.299***	0.05	0.191***	0.343***	0.368***	0.164	0.185**	0.096	0.227***	0.169*	0.06	2,836	0.139
	(4.17)	(3.94)	(0.42)	(3.22)	(13.01)	(4.15)	(1.24)	(2.31)	(0.88)	(2.68)	(1.90)	(0.55)		

Continue on the next page

Table continued from previous page

Country	Constant	ΔVIX_{t-1}	$\Delta LVIX^o_{t-1}$	ΔIR_{t-1}	$cashOI_{t-1}$	ΔVIX_{t-2}	$\Delta LVIX^o_{t-2}$	ΔVIX_{t-3}	$\Delta LVIX^o_{t-3}$	ΔVIX_{t-4}	ΔVIX_{t-5}	$\Delta LVIX^o_{t-5}$	Obs.	R^2
ISR	0.039***	0.303***	-	0.086	0.343***	0.383***	-	0.226***	-	0.254***	0.175**	-	2,886	0.139
	(4.13)	(3.99)	-	(0.40)	(13)	(4.27)	-	(3.25)	-	(3.07)	(1.99)	-		
ITA	0.039***	0.305***	0.041	0.019*	0.343***	0.366***	0.146	0.187***	0.075	0.235***	0.168*	0.072	2,878	0.139
	(4.17)	(4.01)	(0.34)	(1.64)	(12.97)	(4.10)	(1.10)	(3.32)	(0.68)	(2.80)	(1.91)	(0.66)		
JPN	0.04***	0.326***	0.022	0.010	0.348***	0.369***	0.015	0.229***	0.114	0.209**	0.158*	-0.011	2,829	0.142
	(4.29)	(4.2)	(0.21)	(1.17)	(13.06)	(3.63)	(0.15)	(2.76)	(1.06)	(2.32)	(1.78)	(-0.11)		
MYS	0.039***	0.300***	-	0.904*	0.343***	0.389***	-	0.228***	-	0.254***	0.174**	-	2,886	0.140
	(4.12)	(3.95)	-	(1.77)	(13.01)	(4.29)	-	(3.29)	-	(3.08)	(1.97)	-		
MEX	0.033***	0.286***	0.272**	1.416	0.355***	0.459***	0.112	0.247***	0.119	0.300***	0.166	0.061	2,616	0.152
	(3.44)	(3.22)	(2.01)	(1.19)	(12.93)	(5.68)	(0.66)	(3.09)	(0.73)	(3.06)	(1.61)	(0.48)		
NLD	0.039***	0.306***	-0.008	0.019	0.343***	0.376***	0.120	0.191**	0.064	0.236***	0.168*	-0.054	2,878	0.139
	(4.16)	(4.00)	(-0.08)	(1.64)	(13.01)	(4.18)	(0.84)	(2.35)	(0.55)	(2.74)	(1.88)	(-0.42)		
NZL	0.039***	0.305***	-	-0.226	0.344***	0.382***	-	0.227***	-	0.257***	0.175**	-	2,886	0.140
	(4.13)	(4.02)	-	(-1.08)	(13.02)	(4.25)	-	(3.26)	-	(3.10)	(1.98)	-		
NOR	0.039***	0.302 ***	-	0.002	0.343 ***	0.384 ***	-	0.226 ***	-	0.254 ***	0.174*	-	2,886	0.139
	(4.11)	(3.98)	-	(0.00)	(13.00)	(4.27)	-	(3.26)	-	(3.07)	(1.98)	-		
PER	0.039***	0.302***	-	0.001***	0.343***	0.384***	-	0.226***	-	0.254***	0.174**	-	2,886	0.139
	(4.12)	(3.98)	-	(5.34)	(13.01)	(4.27)	-	(3.26)	-	(3.07)	(1.98)	-		
PHL	0.051***	0.322***	-	0.013	0.395***	0.52***	-	0.234**	-	0.324***	0.047	-	1,573	0.179
	(4.12)	(2.65)	-	(1.28)	(11.46)	(4.84)	-	(2.08)	-	(2.62)	(0.33)	-		
POL	0.039***	0.302***	-	1.645	0.342***	0.382***	-	0.226***	-	0.254***	0.172*	-	2,886	0.140
	(4.13)	(3.96)	-	(0.99)	(13.04)	(4.25)	-	(3.26)	-	(3.08)	(1.95)	-		
QAT	0.036***	0.303***	-	0.778	0.279***	0.166	-	0.148	-	0.17	0.327***	-	927	0.095
	(2.59)	(3.00)	-	(1.28)	(7.39)	(1.26)	-	(1.55)	-	(1.52)	(3.1)	-		
RUS	0.027***	0.291***	-0.148	-0.003	0.349***	0.534***	0.075	0.266***	-0.118	0.333***	0.182*	0.049	2,500	0.149
	(2.74)	(3.25)	(-1.01)	(-0.01)	(12.22)	(6.36)	(0.62)	(3.08)	(-1.18)	(3.31)	(1.72)	(0.55)		
SAU	0.039***	0.302***	-	-0.182	0.343***	0.385***	-	0.226***	-	0.254***	0.175**	-	2,886	0.139
	(4.12)	(3.97)	-	(-0.35)	(13.01)	(4.27)	-	(3.25)	-	(3.06)	(1.98)	-		
SGP	0.039***	0.302***	-	0.08	0.343***	0.383***	-	0.226***	-	0.254***	0.175**	-	2,886	0.139
	(4.12)	(3.99)	-	(0.28)	(13.01)	(4.26)	-	(3.24)	-	(3.08)	(1.98)	-		
ZAF	0.038***	0.305***	0.908***	-0.638**	0.322***	0.315***	0.648***	0.159**	0.394	0.211**	0.152*	0.423**	2,868	0.139
	(4.05)	(3.98)	(3.86)	(-2.02)	(12.73)	(3.7)	(3.13)	(2.27)	(1.59)	(2.52)	(1.69)	(2.14)		
KOR	0.039***	0.309***	-0.168	2.298***	0.342***	0.442***	0.024	0.235***	0.018	0.253***	0.176**	0.165	2,886	0.141
	(4.17)	(4.11)	(-1.31)	(2.97)	(13.08)	(4.34)	(0.19)	(3.11)	(0.15)	(2.94)	(1.99)	(1.53)		
ESP	0.039***	0.305***	0.041	0.019	0.343***	0.366***	0.146	0.187**	0.075	0.235***	0.168*	0.072	2,878	0.139
	(4.17)	(4.01)	(0.34)	(1.64)	(12.97)	(4.1)	(1.1)	(2.32)	(0.68)	(2.8)	(1.91)	(0.66)		
SWE	0.039***	0.300***	0.064	-0.009	0.344***	0.366***	0.229**	0.167**	0.104	0.225***	0.173**	0.019	2,885	0.140
	(4.17)	(3.93)	(0.61)	(-0.62)	(13.01)	(4.12)	(2.16)	(2.17)	(1.09)	(2.7)	(1.98)	(0.20)		
CHE	0.039***	0.289***	0.090	-0.006**	0.344***	0.35***	0.128	0.185**	0.088	0.204**	0.189**	-0.013	2,801	0.139
	(4.03)	(3.74)	(0.66)	(-2.12)	(12.73)	(3.63)	(0.94)	(2.26)	(0.66)	(2.26)	(2.09)	(-0.10)		
TWN	0.039***	0.302***	-	0.047	0.343***	0.384***	-	0.226***	-	0.254***	0.174**	-	2,886	0.139
	(4.12)	(3.98)	-	(0.10)	(13.01)	(4.27)	-	(3.25)	-	(3.07)	(1.98)	-		
THA	0.039***	0.302***	-	0.017**	0.343***	0.384***	-	0.226***	-	0.254***	0.174**	-	2,886	0.139
	(4.13)	(3.98)	-	(2.16)	(13.02)	(4.27)	-	(3.25)	-	(3.07)	(1.98)	-		
TUR	0.039***	0.301***	-	0.353	0.343***	0.383***	-	0.225***	-	0.254***	0.172*	-	2,886	0.139
	(4.13)	(3.95)	-	(0.43)	(12.96)	(4.25)	-	(3.23)	-	(3.07)	(1.95)	-		
ARE	0.006	0.293***	-	-0.024	0.272***	0.392***	-	0.218***	-	0.192**	0.258***	-	2,069	0.096
	(0.65)	(3.56)	-	(-1.07)	(10.9)	(3.83)	-	(2.75)	-	(2.33)	(3.17)	-		
GBR	0.039***	0.296***	0.134	0.018	0.343***	0.342***	0.125	0.184**	0.092	0.223***	0.169*	0.032	2,886	0.139
	(4.15)	(3.92)	(1.42)	(0.28)	(13.01)	(3.78)	(1.09)	(2.41)	(0.99)	(2.65)	(1.91)	(0.34)		

Table B.15: Panel Results: VIX Decomposition

Panel regressions with random effects for predictive regression of order imbalance for 41 country-level MSCI based iShares ETFs on percentage change in two component of VIX, namely, conditional variance (ΔVIX^{CV}) which is a measure of uncertainty and risk aversion (ΔVIX^{RA}), percentage change in orthogonalised local volatility index ($\Delta LVIX^{o}$), change in local interest rates (IR), lagged order imbalance (OI_{t-1}), dummy variable for a common language (L) and a distance to the U.S. (G), as well as lags of ΔVIX and $\Delta LVIX^{o}$. Specifically, I estimate the model below and nested specifications of the model:

$$OI_{i,t} = \alpha + \sum_{k=1}^{5} \beta_k^{CV} \Delta VIX_{t-k}^{CV} + \sum_{k=1}^{5} \beta_k^{RA} \Delta VIX_{t-k}^{RA} + \sum_{k=1}^{2} \gamma_k \Delta LVIX_{i,t-k}^{o} + \delta_1 \Delta IR_{i,t-1} + \delta_2 L_i + \delta_3 G_i + \delta_4 TOI_{i,t-1} + \varepsilon$$

The number of lags in the model is determined using Akaike Information Criterion and Bayesian information criterion jointly. Coefficients and z-statistics (based on robust standard errors) are presented. The lag for predictive regression is 1 day. Data is at a daily frequency and covers the period from 2006 or the first trading day of ETF (whichever is later) until the end of December 2016.***,**,* denote significance at the 1%, 5% and 10% level.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
VARIABLES	*TOI*	*TOI*	*TOI*	*smallOI*	*smallOI*	*smallOI*	*retailOI*	*retailOI*	*retailOI*	*largeOI*	*largeOI*	*largeOI*
ΔVIX_{t-1}^{CV}	-0.029***		-0.028***	-0.021***		-0.021***	-0.053***		-0.052***	-0.035***		-0.034***
	(-5.03)		(-4.85)	(-3.82)		(-3.97)	(-5.25)		(-5.16)	(-4.90)		(-4.94)
ΔVIX_{t-1}^{RA}		-0.002*	-0.002*		-0.002**	-0.002**		-0.005***	-0.005***		-0.001	-0.001
		(-1.72)	(-1.90)		(-2.00)	(-2.09)		(-3.93)	(-4.06)		(-0.96)	(-1.09)
$\Delta LVIX_{t-1}^{o}$	-0.052**	-0.074***	-0.053**	-0.032	-0.044**	-0.031	-0.013	-0.044	-0.013	-0.048	-0.079**	-0.049
	(-2.33)	(-3.32)	(-2.35)	(-1.59)	(-2.05)	(-1.49)	(-0.42)	(-1.37)	(-0.39)	(-1.39)	(-2.35)	(-1.41)
ΔIR_{t-1}	-0.005**	-0.005**	-0.005**	-0.002	-0.002	-0.002	-0.006**	-0.007**	-0.007**	-0.002	-0.002	-0.002
	(-2.20)	(-2.23)	(-2.23)	(-1.55)	(-1.58)	(-1.58)	(-2.36)	(-2.43)	(-2.39)	(-0.98)	(-0.97)	(-0.99)
OI_{t-1}	0.181***	0.182***	0.181***	0.207***	0.208***	0.207***	0.136***	0.136***	0.135***	0.122***	0.123***	0.122***
	(12.61)	(12.61)	(12.60)	(8.40)	(8.38)	(8.39)	(9.08)	(9.05)	(9.05)	(11.44)	(11.53)	(11.44)
L	0.004	0.004	0.004	-0.003	-0.003	-0.003	0.011	0.011	0.011	0.004	0.004	0.004
	(0.47)	(0.47)	(0.48)	(-0.78)	(-0.79)	(-0.78)	(0.69)	(0.68)	(0.68)	(0.42)	(0.42)	(0.42)
G	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000**	-0.000**	-0.000**	-0.000	-0.000	-0.000
	(-0.19)	(-0.19)	(-0.20)	(-0.59)	(-0.60)	(-0.60)	(-2.52)	(-2.52)	(-2.52)	(-0.08)	(-0.08)	(-0.09)

Continue on the next page

Table continued from previous page

VARIABLES	(1) *TOI*	(2) *TOI*	(3) *TOI*	(4) *smallOI*	(5) *smallOI*	(6) *smallOI*	(7) *retailOI*	(8) *retailOI*	(9) *retailOI*	(10) *largeOI*	(11) *largeOI*	(12) *largeOI*
$\Delta LVIX^{o}_{t-2}$	-0.090***	-0.102***	-0.088***	-0.059**	-0.066**	-0.059**	-0.110***	-0.115***	-0.102***	-0.101***	-0.121***	-0.099***
	(-3.17)	(-3.64)	(-3.14)	(-2.29)	(-2.47)	(-2.27)	(-3.55)	(-3.69)	(-3.30)	(-3.20)	(-3.92)	(-3.22)
ΔVIX^{CV}_{t-2}	-0.017***		-0.015***	-0.006		-0.005	-0.013*		-0.011	-0.029***		-0.027***
	(-3.51)		(-3.14)	(-1.49)		(-1.37)	(-1.70)		(-1.49)	(-4.27)		(-4.00)
ΔVIX^{CV}_{t-3}	-0.007		-0.005	-0.005		-0.003	-0.002		-0.002	-0.007		-0.005
	(-1.18)		(-0.85)	(-0.86)		(-0.67)	(-0.24)		(-0.18)	(-0.89)		(-0.60)
ΔVIX^{CV}_{t-4}	-0.025***		-0.025***	-0.014***		-0.015***	-0.027***		-0.029***	-0.025***		-0.025***
	(-3.89)		(-4.14)	(-2.77)		(-2.91)	(-2.84)		(-3.05)	(-2.67)		(-2.78)
ΔVIX^{CV}_{t-5}	-0.004		-0.004	-0.004		-0.005	-0.016*		-0.015	-0.015**		-0.015**
	(-0.56)		(-0.55)	(-0.82)		(-0.81)	(-1.69)		(-1.54)	(-2.08)		(-2.08)
ΔVIX^{RA}_{t-2}		-0.000	0.000		-0.002*	-0.001		-0.002	-0.001		-0.000	0.000
		(-0.16)	(0.11)		(-1.73)	(-1.47)		(-1.08)	(-0.79)		(-0.05)	(0.13)
ΔVIX^{RA}_{t-3}		-0.001	-0.001		-0.000	0.000		-0.004***	-0.004***		-0.001	-0.001
		(-1.51)	(-1.18)		(-0.01)	(0.16)		(-3.96)	(-3.60)		(-0.83)	(-0.51)
ΔVIX^{RA}_{t-4}		-0.004***	-0.003***		-0.002*	-0.002*		-0.002	-0.002		-0.004***	-0.004***
		(-3.98)	(-3.90)		(-1.81)	(-1.78)		(-1.45)	(-1.34)		(-4.21)	(-4.01)
ΔVIX^{RA}_{t-5}		-0.002	-0.001		-0.000	-0.000		-0.000	0.001		-0.002	-0.002
		(-1.53)	(-1.18)		(-0.30)	(-0.03)		(-0.09)	(0.40)		(-1.51)	(-1.24)
Constant	0.016*	0.015*	0.016*	0.006	0.006	0.006	0.034***	0.033**	0.034***	0.011	0.010	0.011
	(1.86)	(1.77)	(1.88)	(1.22)	(1.14)	(1.25)	(2.61)	(2.56)	(2.65)	(1.46)	(1.31)	(1.48)
Observations	50,441	50,441	50,441	45,356	45,356	45,356	47,237	47,237	47,237	48,325	48,325	48,325
Adjusted R^2	0.033	0.032	0.033	0.044	0.044	0.044	0.018	0.018	0.019	0.016	0.015	0.016

Appendix C

Supporting Documentation: Chapter 3

C.1 KMV Distance to Default Model

I follow Bharath and Shumway (2008) to construct the distance to default measure. The total value of the firm (V) can be assumed to follow the geometric Brownian motion :

$$dV = \mu V dt + \sigma_V V dW \tag{C.1}$$

where μ is the continuously compounded return, σ_V is the volatility of the total value of the firm and dW is a Wiener process. The value of firm's equity can be represented as a call option written on the underlying assets of the firm with a strike price F, that is equal to the face value of the debt issued by the firm with maturity T. When the value of the firm's assets is above the total amount of debt outstanding, the equity value is positive. In contrast, when the value of assets is lower then debt level, the firm is in bankruptcy and the payout to the equity holders is zero. Using Black-Scholes formula the value of the equity can be shown as:

$$E = V\mathcal{N}(d_1) - e^{-rT} F\mathcal{N}(d_2) \tag{C.2}$$

where $\mathcal{N}$ is the cumulative standard normal distribution function, d_1 and d_2 is:

$$d_1 = \frac{ln(V/F) + (r + 0.5\sigma_V^2)T}{\sigma_V\sqrt{T}} \tag{C.3}$$

$$d_2 = d_1 - \sigma_V\sqrt{T} \tag{C.4}$$

where r is the risk-free rate. All but two variables (V and σ_V) can be estimated in the equation C.2. Therefore, the second equation is required to solve the system of equations. Using Ito's lemma it can be shown that:

$$\sigma_E = \frac{V}{E}\mathcal{N}(d_1)\sigma_V \tag{C.5}$$

In order to solve these equations I start with the estimated value of volatility of firm's assets:

$$\sigma_V = \sigma_E\frac{E}{E+F} \tag{C.6}$$

Using equation C.2 the value of V can be estimated for the past year and the values of σ_V and μ can be re-estimated. The iterative procedure loops through the estimates of σ_V until the values converge. Once all parameters are estimated the distance to default measure can then be calculated as follows:

$$DD = \frac{\ln\frac{V}{F} + (\mu - 0.5\sigma_V^2)T}{\sigma_V\sqrt{T}} \tag{C.7}$$

The expected default frequency (EDF) is a more useful measure as it captures the implied probability of default by the firm:

$$EDF = \mathcal{N}(-DD) \tag{C.8}$$

C.2 Tables

Table C.1: List of Control Variables

This table shows the list of control variables used in this paper, as well as the source for each variable. The variables are constracted based on Green et al. (2017).

Acronym	Description	Source
acc	Accruals	Compustat
agr	Asset Growth	Compustat
bm	Book-to-Market Ratio	Compustat
grltnoa	Growth in L-T Net Operating Assets	Compustat
gma	Gross Profitability	Compustat
illiq	Amihud (2002) illiquidity	CRSP
invest	Investment Growth	Compustat
me	Market Capitalisation	CRSP
numan	Number of Analysts Covering Stock	I/B/E/S
op	Operating Profitability	Compustat
vol	Trading Volume orthogonalised to *me*	CRSP
vwap	Value-Weighted Average Price	CRSP

Table C.2: Distribution of Credit Scores

This table shows how the credit scores are assigned to each S&P long-term credit rating, as well as the average distribution of stocks across different ratings.

S&P Ratings	Credit Score	Average Percentage of Firms
		Investment grade
AAA	1	0.69
AA+	2	0.21
AA	3	1.01
AA-	4	1.64
A+	5	3.97
A	6	7.51
A-	7	6.75
BBB+	8	9.25
BBB	9	13.13
BBB-	10	10.50
		Non-investment grade
BB+	11	6.89
BB	12	9.09
BB-	13	11.41
B+	14	9.19
B	15	5.69
B-	16	2.29
CCC+	17	0.53
CCC	18	0.12
CCC-	19	0.04
CC	20	0.03
C	21	0.00
D	22	0.06

Bibliography

Abhyankar, Abhay, Ilias Filippou, Pedro Garcia-Ares, and Ozkan Haykir, 2019, Overcoming arbitrage limits: Option trading and momentum returns, Working Paper.

Acharya, Viral V, and Lasse Heje Pedersen, 2005, Asset pricing with liquidity risk, *Journal of Financial Economics* 77, 375–410.

Ackert, Lucy F., and Yisong S. Tian, 2008, Arbitrage, liquidity, and the valuation of exchange traded funds, *Financial Markets, Institutions and Instruments* 17, 331–362.

Adrian, Tobias, Daniel Stackman, and Erik Vogt, 2019, Global price of risk and stabilization policies, *IMF Economic Review* 67, 215–260.

Altman, Edward I, 1968, Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, *The Journal of Finance* 23, 589–609.

Amihud, Yakov, 2002, Illiquidity and stock returns: Cross-section and time-series effects, *Journal of Financial Markets* 5, 31–56.

Amihud, Yakov, Allaudeen Hameed, Wenjin Kang, and Huiping Zhang, 2015, The illiquidity premium: International evidence, *Journal of Financial Economics* 117, 350–368.

Asness, Clifford S, Andrea Frazzini, and Lasse Heje Pedersen, 2019, Quality minus junk, *Review of Accounting Studies* 24, 34–112.

Atanasov, Victoria, 2014, Common risk factors in equity markets, Technical report, Tinbergen Institute Discussion Paper.

Avramov, Doron, Tarun Chordia, Gergana Jostova, and Alexander Philipov, 2009, Credit ratings and the cross-section of stock returns, *Journal of Financial Markets* 12, 469–499.

Avramov, Doron, Tarun Chordia, Gergana Jostova, and Alexander Philipov, 2013, Anomalies and financial distress, *Journal of Financial Economics* 108, 139–159.

Baker, Malcolm, and Jeffrey Wurgler, 2006, Investor sentiment and the cross-section of stock returns, *The Journal of Finance* 61, 1645–1680.

Bandi, Federico M, Claudia E Moise, and Jeffrey R Russell, 2006, Market volatility, market frictions, and the cross-section of stock returns, University of Chicago and Case Western Reserve University.

Barber, Brad M., Terrance Odean, and Ning Zhu, 2009, Do retail trades move markets?, *Review of Financial Studies* 22, 151–186.

Barbon, Andrea, and Virginia Gianinazzi, 2019, Quantitative easing and equity prices: Evidence from the etf program of the bank of japan, Working Paper.

Bekaert, Geert, Michael Ehrmann, Marcel Fratzscher, and Arnaud Mehl, 2014, The global crisis and equity market contagion, *The Journal of Finance* 69, 2597–2649.

Bekaert, Geert, and Marie Hoerova, 2014, The vix, the variance premium and stock market volatility, *Journal of Econometrics* 183, 181–192.

Bekaert, Geert, Marie Hoerova, and Marco Lo Duca, 2013, Risk, uncertainty and monetary policy, *Journal of Monetary Economics* 60, 771–788.

Ben-David, Itzhak, Francesco Franzoni, and Rabih Moussawi, 2017, Exchange-traded funds, *Annual Review of Financial Economics* 9, 169–189.

Ben-David, Itzhak, Francesco Franzoni, and Rabih Moussawi, 2018, Do etfs increase volatility?, *The Journal of Finance* 73, 2471–2535.

Beneish, Messod Daniel, Charles MC Lee, and D Craig Nichols, 2015, In short supply: Short-sellers and stock returns, *Journal of Accounting and Economics* 60, 33–57.

Berger, Dave, Kuntara Pukthuanthong, and J Jimmy Yang, 2011, International diversification with frontier markets, *Journal of Financial Economics* 101, 227–242.

Bertone, Stephen, Imants Paeglis, and Rahul Ravi, 2015, (How) has the market become more efficient?, *Journal of Banking and Finance* 54, 72–86.

Bharath, Sreedhar T, and Tyler Shumway, 2008, Forecasting default with the merton distance to default model, *The Review of Financial Studies* 21, 1339–1369.

Bhattacharya, Ayan, and Maureen O'Hara, 2018, Can ETFs Increase Market Fragility? Effect of Information Linkages in ETF Markets, Working Paper.

BlackRock, 2011, *Revisiting the Flash Crash. A Year Has Passed, What Has Changed?*.

BlackRock, 2018, *February 2018 Case Study: ETF Trading in a High-Velocity Market*.

Boehmer, Ekkehart, Zsuzsa R Huszár, Yanchu Wang, and Xiaoyan Zhang, 2018, Are shorts equally informed? a global perspective, Working Paper.

Boehmer, Ekkehart, Charles M Jones, and Xiaoyan Zhang, 2008, Which shorts are informed?, *The Journal of Finance* 63, 491–527.

Boehmer, Ekkehart, Charles M. Jones, and Xiaoyan Zhang, 2017, Tracking Retail Investor Activity, Working Paper.

Boehmer, Ekkehart, and Juan Wu, 2012, Short selling and the price discovery process, *The Review of Financial Studies* 26, 287–322.

Boguth, Oliver, and Mikhail Simutin, 2018, Leverage constraints and asset prices: Insights from mutual fund risk taking, *Journal of Financial Economics* 127, 325–341.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, 2012, Salience theory of choice under risk, *Quarterly Journal of Economics* 127, 1243–1285.

Boutchkova, Maria, Hitesh Doshi, Art Durnev, and Alexander Molchanov, 2012, Precarious politics and return volatility, *The Review of Financial Studies* 25, 1111–1154.

Brogaard, Jonathan, Terrence Hendershott, and Ryan Riordan, 2014, High-frequency trading and price discovery, *The Review of Financial Studies* 27, 2267–2306.

Broman, Markus S, 2016, Liquidity, style investing and excess comovement of exchange-traded fund returns, *Journal of Financial Markets* 30, 27–53.

Brown, David C, Shaun William Davies, and Matthew Ringgenberg, 2019, ETF Flows, Non-Fundamental Demand, and Return Predictability, Working Paper.

Campbell, John Y, Jens Hilscher, and Jan Szilagyi, 2008, In search of distress risk, *The Journal of Finance* 63, 2899–2939.

Carhart, Mark M, 1997, On persistence in mutual fund performance, *The Journal of Finance* 52, 57–82.

Chacko, George, Sanjiv Das, and Rong Fan, 2016, An index-based measure of liquidity, *Journal of Banking and Finance* 68, 162–178.

Charoenwong, Ben, Randall Morck, and Yupana Wiwattanakantang, 2019, Asset prices and corporate responses to bank of japan etf purchases, Technical report, National Bureau of Economic Research.

Chen, Nai-fu, and Feng Zhang, 1997, Correlations, trades and stock returns of the pacific-basin markets, *Pacific-Basin Finance Journal* 5, 559–577.

Converse, Nathan, Eduardo Levy-Yeyati, and Tomas Williams, 2018, How ETFs Amplify the Global Financial Cycle in Emerging Markets, Working Paper.

Da, Zhi, Joseph Engelberg, and Pengjie Gao, 2011, In Search of Attention, *Journal of Finance* 66, 1461–1499.

Da, Zhi, Joseph Engelberg, and Pengjie Gao, 2015, The sum of all FEARS investor sentiment and asset prices, *Review of Financial Studies* 28, 1–32.

Da, Zhi, and Pengjie Gao, 2010, Clientele change, liquidity shock, and the return on financially distressed stocks, *Journal of Financial and Quantitative Analysis* 45, 27–48.

Da, Zhi, and Sophie Shive, 2018, Exchange traded funds and asset return correlations, *European Financial Management* 24, 136–168.

Davies, Shaun, 2019, Speculation sentiment, Working Paper.

DeGennaro, Ramon P, and Cesare Robotti, 2007, Financial market frictions, *Economic Review-Federal Reserve Bank of Atlanta* 92, 1.

Delcoure, Natalya, and Maosen Zhong, 2007, On the premiums of iShares, *Journal of Empirical Finance* 14, 168–195.

Dichev, Ilia D, 1998, Is the risk of bankruptcy a systematic risk?, *The Journal of Finance* 53, 1131–1147.

Drechsler, Itamar, and Qingyi Freda Drechsler, 2016, The shorting premium and asset pricing anomalies, Technical report, National Bureau of Economic Research.

Drechsler, Itamar, Alan Moreira, and Alexi Savov, 2018, Liquidity creation as volatility risk, Working Paper.

Engelberg, Joseph E, Adam V Reed, and Matthew C Ringgenberg, 2012, How are shorts informed?: Short sellers, news, and information processing, *Journal of Financial Economics* 105, 260–278.

Engelberg, Joseph E, Adam V Reed, and Matthew C Ringgenberg, 2018, Short-selling risk, *The Journal of Finance* 73, 755–786.

Engle, Robert, 2002, Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models, *Journal of Business & Economic Statistics* 20, 339–350.

Engle, Robert F, and Debojyoti Sarkar, 2006, Premiums-Discounts and Exchange Traded Funds, *The Journal of Derivatives* 13, 27–45.

Evans, Richard B, Oğuzhan Karakaş, Rabih Moussawi, and Michael Young, 2019, Phantom of the opera: Etfs and shareholder voting, Working Paper.

Fama, Eugene F, and Kenneth R French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Filippou, Ilias, Arie E. Gozluklu, and Hari Rozental, 2019, Etf arbitrage and international diversification, Working Paper.

Filippou, Ilias, Arie Eskenazi Gozluklu, and Mark P. Taylor, 2018, Global Political Risk and Currency Momentum, *Journal of Financial and Quantitative Analysis* 53, 2227–2259.

Filippou, Ilias, and Lucius Li, 2018, U.S. Political Sentiment, Working Paper.

Forbes, Kristin, 2012, The "big c": Identifying contagion, Technical report, National Bureau of Economic Research.

Forbes, Kristin J, and Francis E Warnock, 2012, Capital flow waves: Surges, stops, flight, and retrenchment, *Journal of International Economics* 88, 235–251.

Friederich, Sylvain, and Richard Payne, 2015, Order-to-trade ratios and market liquidity, *Journal of Banking & Finance* 50, 214–223.

Gagnon, Louis, and G Andrew Karolyi, 2010, Multi-market trading and arbitrage, *Journal of Financial Economics* 97, 53–80.

Gao, Pengjie, Christopher A Parsons, and Jianfeng Shen, 2017, Global relation between financial distress and equity returns, *The Review of Financial Studies* 31, 239–277.

Garlappi, Lorenzo, Tao Shu, and Hong Yan, 2008, Default risk, shareholder advantage, and stock returns, *The Review of Financial Studies* 21, 2743–2778.

Garlappi, Lorenzo, and Hong Yan, 2011, Financial distress and the cross-section of equity returns, *The Journal of Finance* 66, 789–822.

Glosten, Lawrence R, Suresh Nallareddy, and Yuan Zou, 2016, ETF trading and informational efficiency of underlying securities, Working Paper.

Goetzmann, William N., Zoran Ivkovic, and K. Geert Rouwenhorst, 2001, Day Trading International Mutual Funds: Evidence and Policy Solutions, *The Journal of Financial and Quantitative Analysis* 36, 287.

Goldstein, Morris, 1998, *The Asian financial crisis: Causes, cures, and systemic implications*, volume 55 (Peterson Institute).

Green, Jeremiah, John RM Hand, and X Frank Zhang, 2017, The characteristics that provide independent information about average us monthly stock returns, *The Review of Financial Studies* 30, 4389–4436.

Guo, Xu, and Chunchi Wu, 2019, Short interest, stock returns and credit ratings, Working Paper.

Hangströmer, Björn, and Nordén Lars, 2013, The diversity of high-frequency traders, *Journal of Financial Markets* 741–770.

Hendershott, Terrence, Charles M Jones, and Albert J Menkveld, 2011, Does algorithmic trading improve liquidity?, *The Journal of Finance* 66, 1–33.

Holden, Craig W., and Stacey Jacobsen, 2014, Liquidity measurement problems in fast, competitive markets: Expensive and cheap solutions, *Journal of Finance* 69, 1747–1785.

Hong, Harrison, and David A Sraer, 2016, Speculative betas, *The Journal of Finance* 71, 2095–2144.

Hou, Kewei, and Tobias J Moskowitz, 2005, Market frictions, price delay, and the cross-section of expected returns, *The Review of Financial Studies* 18, 981–1020.

Huang, Shiyang, Maureen O'Hara, and Zhuo Zhong, 2018, Innovation and informed trading: Evidence from industry etfs, Working Paper.

Huszár, Zsuzsa R, Ruth SK Tan, and Weina Zhang, 2017, Do short sellers exploit industry information?, *Journal of Empirical Finance* 41, 118–139.

Huszár, Zsuzsa R, Ruth SK Tan, and Weina Zhang, 2019, The world-wide source of industry information: The industry concentration of short sellers, Working Paper.

Israeli, Doron, Charles MC Lee, and Suhas A Sridharan, 2017, Is there a dark side to exchange traded funds? an information perspective, *Review of Accounting Studies* 22, 1048–1083.

Jørgensen, Kjell, Johannes Skjeltorp, and Bernt Arne Ødegaard, 2017, Throttling hyperactive robots–order-to-trade ratios at the oslo stock exchange, *Journal of Financial Markets* 37, 1–16.

Jotikasthira, Chotibhak, Christian Lundblad, and Tarun Ramadorai, 2012, Asset fire sales and purchases and the international transmission of funding shocks, *The Journal of Finance* 67, 2015–2050.

Kalay, Avner, Oğuzhan Karakaş, and Shagun Pant, 2014, The market value of corprate votes: Theory and evidence from option prices, *The Journal of Finance* 69, 1235–1271.

Karmaziene, Egle, and Valeri Sokolovski, 2015, Exchange traded funds and the 2008 short-sale ban, *Swedish House of Finance Research Paper* 14–05.

Lee, Charles M C, and Mark J Ready, 1991, Inferring Trade Direction from Intraday Data, *The Journal of Finance* 46, 733–746.

Lettau, Martin, and Ananth Madhavan, 2018, Exchange-traded funds 101 for economists, *Journal of Economic Perspectives* 32, 135–54.

Levy, Ariel, and Offer Lieberman, 2013, Overreaction of country etfs to us market returns: Intraday vs. daily horizons and the role of synchronized trading, *Journal of Banking & Finance* 37, 1412–1421.

Li, Frank Weikai, and Qifei Zhu, 2018, Shorting selling etfs, Working Paper.

Mackintosh, Phil, 2014, *ETF Insights: ETFs Rarely Enter the Arb Zone*, KCG Market Commentary.

Madhavan, Ananth, and Aleksander Sobczyk, 2016, Price Dynamics and Liquidity of Exchange-Traded Funds, *Journal Of Investment Management* 14, 1–17.

Malamud, Semyon, 2016, A dynamic equilibrium model of etfs, Working Paper.

Malinova, Katya, Andreas Park, and Ryan Riordan, 2018, Do retail traders suffer from high frequency traders, Working Paper.

Malkhozov, Aytek, Philippe Mueller, Andrea Vedolin, and Gyuri Venter, 2018, Funding illiquidity, funding risk, and global stock returns, Working Paper.

Mancini, Loriano, Angelo Ranaldo, and Jan Wrampelmeyer, 2013, Liquidity in the foreign exchange market: Measurement, commonality, and risk premiums, *Journal of Finance* 68, 1805–1841.

Marshall, Ben R, Nhut H Nguyen, and Nuttawat Visaltanachoti, 2013, Etf arbitrage: Intraday evidence, *Journal of Banking & Finance* 37, 3486–3498.

Menkhoff, Lukas, Lucio Sarno, Maik Schmeling, and Andreas Schrimpf, 2012, Carry trades and global foreign exchange volatility, *The Journal of Finance* 67, 681–718.

Merton, Robert C., 1974, On the pricing of corporate debt: The risk structure of interest rates, *The Journal of Finance* 29, 449–470.

Miffre, Joëlle, 2007, Country-specific ETFs: An efficient approach to global asset allocation, *Journal of Asset Management* 8, 112–122.

Miranda-Agrippino, Silvia, and Hélene Rey, 2015, World asset markets and the global financial cycle, Technical report, National Bureau of Economic Research.

Nagel, Stefan, 2005, Short sales, institutional investors and the cross-section of stock returns, *Journal of Financial Economics* 78, 277–309.

Newey, Whitney K., and Kenneth D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.

O'Hara, Maureen, and Mao Ye, 2011, Is market fragmentation harming market quality?, *Journal of Financial Economics* 100, 459–474.

Ohlson, James A, 1980, Financial ratios and the probabilistic prediction of bankruptcy, *Journal of Accounting Research* 109–131.

Pan, Kevin, and Yao Zeng, 2019, ETF arbitrage under liquidity mismatch, Working Paper.

Pasquariello, Paolo, 2014, Financial Market Dislocations, *The Review of Financial Studies* 27, 1868–1914.

Pástor, L'uboš, and Robert F Stambaugh, 2003, Liquidity risk and expected stock returns, *Journal of Political Economy* 111, 642–685.

Peress, Joel, and Daniel Schmidt, 2017, Noise traders incarnate: Describing a realistic noise trading process, Working Paper.

Petajisto, Antti, 2017, Inefficiencies in the pricing of exchange-traded funds, *Financial Analysts Journal* 73, 24–54.

Pukthuanthong, Kuntara, and Richard Roll, 2009, Global market integration: An alternative measure and its application, *Journal of Financial Economics* 94, 214–232.

Raddatz, Claudio, and Sergio L. Schmukler, 2012, On the international transmission of shocks: Micro-evidence from mutual fund portfolios, *Journal of International Economics* 88, 357–374.

Rapach, David E, Matthew C Ringgenberg, and Guofu Zhou, 2016, Short interest and aggregate stock returns, *Journal of Financial Economics* 121, 46–65.

Rapach, David E, Jack K Strauss, and Guofu Zhou, 2013, International stock return predictability: what is the role of the united states?, *The Journal of Finance* 68, 1633–1662.

Rey, Hélène, 2015, Dilemma not trilemma: the global financial cycle and monetary policy independence, Technical report, National Bureau of Economic Research.

Richie, Nivine, Robert T. Daigler, and Kimberly C. Gleason, 2008, The limits to stock index arbitrage: Examining S&P 500 futures and SPDRS, *Journal of Futures Markets* 28, 1182–1205.

Rozental, Hari, 2019, Financial frictions risk and etf premium, Working Paper.

Saffi, Pedro AC, and Kari Sigurdsson, 2010, Price efficiency and short selling, *The Review of Financial Studies* 24, 821–852.

SEC, 2019, Securities and exchange commission rule: Release nos. 33-10695, Technical report, Available at https://www.sec.gov/rules/final/2019/33-10695.pdf.

Shumway, Tyler, and Vincent A Warther, 1999, The delisting bias in crsp's nasdaq data and its implications for the size effect, *The Journal of Finance* 54, 2361–2379.

Skjeltorp, Johannes Atle, Elviraand Sojli, and Wing Wah Tham, 2015, Trading on algos, Working Paper.

Solnik, Bruno, Cyril Boucrelle, and Yann Le Fur, 1996, International market correlation and volatility, *Financial Analysts Journal* 52, 17–34.

Stambaugh, Robert F, Jianfeng Yu, and Yu Yuan, 2012, The short of it: Investor sentiment and anomalies, *Journal of Financial Economics* 104, 288–302.

Vanguard, 2016, *Explaining ETF premiums and discounts*, Available at https://www.vanguardcanada.ca/documents/etf-premiums-and-discounts.pdf.

Vassalou, Maria, and Yuhang Xing, 2004, Default risk in equity returns, *The Journal of Finance* 59, 831–868.

Wermers, Russ, and Jinming Xue, 2015, Intraday ETF Trading and the Volatility of the Underlying, Working Paper.

WisdomTree, 2019, *Intra-Day Pricing: How ETF shares are priced*, Investing and Trading, Available at https://www.wisdomtree.eu/en-gb/-/media/eu-media-files/other-documents/educational/intra-day-pricing-how-etf-shares-are-priced.pdf.

Zhang, Frank, 2010, High-frequency trading, stock volatility, and price discovery, Working Paper.