Fisher College of Business Working Paper Series

Charles A. Dice Center for Research in Financial Economics

Do ETFs Increase Volatility?

Itzhak Ben-David, The Ohio State University

Francesco Franzoni, University of Lugano and the Swiss Finance Institute

Rabih Moussawi, University of Pennsylvania

Dice Center WP 2011-20
Fisher College of Business WP 2011-03-20

Revision: April 2014
Original: December 2011

Do ETFs Increase Volatility?

Itzhak Ben-David

Fisher College of Business, The Ohio State University, and NBER

Francesco Franzoni

University of Lugano and the Swiss Finance Institute

Rabih Moussawi

Wharton Research Data Services, The Wharton School, University of Pennsylvania

June 2014

Abstract

A long-lasting debate in finance centers on the impact of derivatives on the efficiency of prices of the underlying securities. The paper contributes to this literature by studying whether exchange traded funds (ETFs)—an asset of rising importance—affect the non-fundamental volatility of the stocks in their baskets. Using identification strategies based on the mechanical variation in ETF ownership, including regression discontinuity, we show that stocks owned by ETFs exhibit significantly higher intraday and daily volatility. Variance-ratio tests, as well as price reversals, suggest that the mean-reverting component of stock prices is inflated by ETF ownership. We estimate that an increase of one standard deviation in ETF ownership is associated with an increase of 19% in intraday stock volatility. The driving channel appears to be arbitrage activity which propagates liquidity shocks from the ETF market to the underlying stocks.

Keywords: ETFs, stocks, volatility, mispricing, fund flow

JEL Classification: G12, G14, G15

We are especially grateful to Robin Greenwood (AFA discussant) and Dimitri Vayanos. We thank Yakov Amihud, George Aragon, Hank Bessembinder, Pierre Collin-Dufresne, Chris Downing, Andrew Ellul, Vincent Fardeau, Thierry Foucault, Rik Frehen, Denys Glushkov, Jungsuk Han, Harald Hau, Augustin Landier, Ananth Madhavan, David Mann, Rodolfo Martell, Albert Menkveld, Robert Nestor, Marco Pagano, Ludovic Phalippou, Scott Richardson, Anton Tonev, Tugkan Tuzun, Scott Williamson, Hongjun Yan, and participants at seminars and conferences at SAC Capital Advisors, the University of Lugano, the 4th Paris Hedge Funds Conference, the 5th Paul Woolley Conference (London School of Economics), the 8th Csef-IGIER Symposium (Capri), the 5th Erasmus Liquidity Conference (Rotterdam), the 1st Luxembourg Asset Pricing Summit, the Geneva Conference and Liquidity and Arbitrage, the 20th Annual Conference of the Multinational Finance Society, the 7th Rothschild Caesarea Conference, the Swedish House of Finance, the FIRS conference (Toronto), for helpful comments and suggestions. Ben-David acknowledges support from the Neil Klatskin Chair in Finance and Real Estate and from the Dice Center at the Fisher College of Business. An earlier version of this paper circulated under the title "ETFs, Arbitrage, and Shock Propagation".

1 Introduction

The question about the effect of derivatives on the quality of the underlying securities' prices has concerned the theoretical and empirical literature in finance for a long time. On one side of the debate, some authors have expressed the concern that liquidity shocks in derivatives markets can trickle down to the cash market adding noise to prices. For example, Stein (1987) makes the point that imperfectly informed speculators in futures market can destabilize spot prices. Among the supporters of the alternative view, Grossman (1988) argues that the existence of futures provides additional market-making power to absorb the impact of liquidity shocks. As a result, volatility in the spot market is reduced (see also Danthine (1978) and Turnovsky (1983)). This paper intends to contribute to this debate by bringing empirical evidence from the market for Exchange Traded Funds (ETFs).

With \$2.5 trillion of assets under management globally as of October 2013,[1] ETFs are rising steadily among the big players in the asset management industry. More importantly, this asset class is capturing an increasing share of transactions in financial markets. For example, in August 2010, exchange traded products accounted for about 40% of all trading volume in U.S. markets (Blackrock (2011)). This explosive growth has attracted the attention of regulators. The SEC has begun to review the role of ETFs in increasing volatility of the underlying securities. Regulators are wary of high frequency volatility because it can reduce participation of long-term investors.[2] The desire to address some open questions regarding this relatively unexplored asset class, as long as readily available data on ETF stocks ownership, flows, prices, and NAV, motivate us to choose the ETF market as a laboratory to study the impact of derivatives on security prices.

Using exogenous variation in ETF ownership we test whether ETFs lead to an increase in the non-fundamental volatility of the securities in their baskets. The main empirical finding of

[1] See http://www.hedgefundfundamentals.com/wp-content/uploads/2012/08/HFF_Hedge_Funds_101_10-2013FINAL.pdf

[2] Regulators have taken into consideration the potential illiquidity of ETFs, which manifested during the Flash Crash of May 6, 2010, when 65% of the cancelled trades were ETF trades. Also relevant is the potential for counterparty risk, which seems to be operating in the cases of both synthetic replication (as the swap counterparty may fail to deliver the index return) and physical replication (as the basket securities are often loaned out). Concerns have been expressed that a run on ETFs may endanger the stability of the financial system (Ramaswamy (2011)). With regard to the SEC ETF-related concerns, see "SEC Reviewing Effects of ETFs on Volatility" by Andrew Ackerman, Wall Street Journal, 19 October 2011, and "Volatility, Thy Name is E.T.F.", by Andrew Ross Sorkin, New York Times, October 10, 2011. With regard to the SEC focus on short-term volatility, see the SEC Concept release No. 34-61358.

the paper is a causal link going from ETF ownership to stock volatility. At least part of this volatility effect can be traced to the impact of ETF arbitrage on the mean-reverting component of stock prices. Hence, the evidence supports the hypothesis that ETFs increase noise in stock prices.

The theoretical channel for the effect that we identify relies on limited arbitrage and clientele effects. If arbitrage is limited, a liquidity shock can propagate from the ETF market to the underlying securities and add noise to prices. To illustrate this effect, consider the example of a large liquidity sell order of ETF shares by an institutional trader. As captured by the models of Greenwood (2005) and Gromb and Vayanos (2010), arbitrageurs buy the ETF and hedge this position by selling the underlying portfolio. Arbitrageurs with limited risk-bearing capacity require a compensation in terms of positive expected returns to take the other side of the liquidity trade. Hence, the selling activity leads to downward price pressure on the underlying portfolio. Through this channel, the repeated arrival of liquidity shocks in the ETF market adds a new layer of non-fundamental volatility in the prices of the underlying securities. An additional assumption to obtain this result is that, in the absence of ETFs, liquidity trades would not hit the underlying security with the same intensity. Rather, it has to be the case that ETFs attract a new clientele of high-turnover investors that impound liquidity shocks at a higher rate.[3] This conjecture seems warranted in light of Amihud and Mendelson's (1987) model, which predicts that short-horizon investors self-select into more liquid assets, such as ETFs.

For robustness, we rely on two different identification strategies to obtain the main empirical result. First, we exploit cross-sectional and time-series variation in ETF ownership of stocks. ETFs tend to hold stocks in the same proportion as in the index that they track. The identification comes from the fact that variation in ETF ownership, across stocks and over time, depends on factors that are exogenous with respect to our dependent variables of interest (volatility and turnover). Specifically, the same stock appears with different weights in different indexes. Furthermore, the fraction of ETF ownership in a firm depends also on the size of the ETF (its assets under management) relative to that of the company. As a result, while it is possible that flows into ETFs are correlated with fundamental information regarding the

[3] E.g., hedge funds prefer using ETFs as a hedging vehicle ("Hedge Fund Monitor" by Goldman Sachs, November 2013).

underlying stocks (e.g., sector-related news), it is unlikely that fundamental reasons produce an effect on volatility that is stronger for stocks with higher ETF ownership.

For the second identification strategy, we draw on recent research by Chang, Hong, and Liskovich (2013). These authors implement a regression discontinuity design that exploits the mechanical rule allocating stocks to the Russell 1000 (top 1000 stocks by size) and Russell 2000 (next 2000 stocks by size) indexes. Due to the big difference in index weights, the top stocks in the Russell 2000 receive significantly larger amounts of passive money than the bottom stocks in the Russell 1000. Hence, in a close proximity of the cutoff, a switch to either index generates a great amount of exogenous variation in ETF ownership, which we use to identify the effect of ETFs on volatility. This procedure identifies in a clean way the effect of interest, but the estimates are local effects. For this reason, we choose to emphasize the more conservative magnitudes that result from the first identification strategy.

Our first set of results shows that intraday volatility increases with ETF ownership. For S&P 500 stocks, a one standard deviation change in ETF ownership is associated with a 19% standard deviation increase in intraday volatility. The effect on volatility also survives in daily returns and is not explained by mutual fund ownership, including that by index funds. The estimates are generally less economically significant for smaller stocks, consistent with ETF arbitrageurs concentrating on a subset of more liquid stocks to build the replicating portfolio.

The increase in volatility is not necessarily a negative phenomenon if it results from enhanced price discovery which makes prices more reactive to fundamental information. This case corresponds to an improvement of price efficiency. To test whether this effect is behind the observed increase in volatility, we measure the impact of ETFs on the mean-reverting component of stock prices. Using intraday variance ratios as in O'Hara and Ye (2011), we show that price efficiency deteriorates for stocks with higher ETF ownership at the fifteen second frequency, which captures the investment horizon of ETF arbitrageurs. At the daily frequency, ETF flows trigger price reversals suggesting a persistence of liquidity shocks at lower frequencies as well. In sum, ETFs appear to inflate the mean-reverting component of stock prices which suggests a deterioration in price efficiency, both intraday and at the daily frequency.

To bring further evidence on the driving channel for the volatility effect, we document that volatility increases at times when arbitrage is more likely to occur, that is, when the divergence between the ETF price and the NAV is large. We also find that ETF flows impact the volatility of the underlying stocks and this effect is stronger for stocks with high ETF ownership. Further supporting the arbitrage channel, we show that the volatility effect is more pronounced among stocks with lower limits of arbitrage, as captured by bid-ask spreads and share lending fees.

The hypothesis that ETFs attract a new clientele of high-turnover investors yields the testable prediction that turnover should also increase with ETF ownership. The evidence suggests that this is the case. In particular, a one-standard deviation increase in ETF ownership is associated with an increase of 19% of a standard deviation in daily turnover. Also, the higher turnover is linked to the same arbitrage channels that are driving the volatility effect. This finding corroborates the view that the high turnover clientele of ETFs is inherited by the underlying stocks as a result of arbitrage.

Our study is related to several strands of the literature. Earlier studies that examine whether the existence of derivatives increase the volatility of the fundamental asset focused on the link between futures and equities. The proposed economic channel in this literature is the same as the one that we test in this paper: non-fundamental shocks in the futures market filter to the equity market via arbitrage trades, thus increasing the volatility in the equity market. In a cross-sectional analysis, Bessembinder and Seguin (1992) find that high trading volume in the futures market is associated with lower equity volatility. However, consistent with the idea that non-fundamental shocks in the futures market are passed down to the equity market, they find that unexpected futures-trading volume is positively correlated with equity volatility. Chang, Cheng, and Pinegar (1999) document that the introduction of futures trading increased the volatility of stocks in the Nikkei index stocks. Roll, Schwartz, and Subrahmanyam (2007) find evidence of Granger causality between prices in the futures and equity markets: price shocks are transmitted from the futures market to the equity market and vice versa. Relative to this literature, our evidence is more conclusive in finding a significant impact of ETF ownership on the volatility of the underlying assets.

Several studies test whether ETFs have a destabilizing effect on markets. Cheng and Madhavan (2009) and Trainor (2010) investigate whether the daily rebalancing of leveraged and inverse ETFs increases stock volatility and find mixed evidence. Bradley and Litan (2010) voice concerns that ETFs may drain the liquidity of already illiquid stocks and commodities, especially if a short squeeze occurs and ETF sponsors rush to create new ETF shares. Madhavan (2011) relates market fragmentation in ETF trading to the Flash Crash of 2010. In work that is more recent than our paper, Da and Shive (2013) find that ETF ownership has a positive effect on the comovement of stocks in the same basket. This result is a direct implication of our finding. We show that ETF ownership increases stock volatility via the propagation of liquidity shocks. Because the stocks in the same basket are going to be affected by the same liquidity shocks, their covariance increases as a result.

This paper also relates to the empirical and theoretical literature studying the effect of institutions on asset prices. There is mounting evidence of the effect of institutional investors on expected returns (Shleifer (1986), Barberis, Shleifer, and Wurgler (2005), Greenwood (2005), Coval and Stafford (2007), and Wurgler (2011) for a survey) and on correlations of asset returns (Anton and Polk (2014), Chang and Hong (2011), Greenwood and Thesmar (2011), Lou (2011), and Jotikasthira, Lundblad, and Ramadorai (2012)). Cella, Ellul, and Giannetti (2013) show that institutional investors' portfolio turnover is an important determinant of stock price resiliency following adverse shocks. In the context of momentum strategies, Lou and Polk (2013) make the related claim that arbitrageurs can have a destabilizing impact on stock prices. Related to our empirical evidence, Basak and Pavlova (2013a, 2013b) make the theoretical point that the inclusion of an asset in an index tracked by institutional investors increases the non-fundamental volatility in that asset's prices.

The theoretical framework for the shock propagation effect that we describe is based on the literature on shock propagation with limited arbitrage. Shock propagation can occur via a number of different channels, including portfolio rebalancing by risk-averse arbitrageurs (e.g., Greenwood (2005)), wealth effects (e.g., Kyle and Xiong (2001)), and liquidity spillovers (e.g., Cespa and Foucault (2012)). The mechanism that most closely describes our empirical evidence is the one by Greenwood (2005).

The paper proceeds as follows. Section 2 provides institutional details on ETF arbitrage and the theoretical framework for the effects that we study. Section 3 describes the data. Section 4 provides the main evidence of the effects of ETF ownership on stock volatility and turnover.

Section 5 provides evidence on role of arbitrage in driving the main effect on volatility. Section 6 concludes.

2 ETF Arbitrage: Institutional Details and Theoretical Framework

2.1 Mechanics of Arbitrage

Exchange traded funds (ETFs) are investment companies that typically focus on one asset class, industry, or geographical area. Most ETFs track an index, very much like passive index mutual funds. Unlike index funds, ETFs are listed on an exchange and trade throughout the day. ETFs were first introduced in the late 1980s and became popular with the issuance in January 1993 of the SPDR (Standard & Poor's Depository Receipts, known as "Spider"), which is an ETF that tracks the S&P 500 (which we label "SPY," from its ticker). In 1995, another SPDR, the S&P MidCap 400 Index (MDY) was introduced, and subsequently the number of ETFs exploded to more than 1,600 by the end of 2012, spanning various asset classes and investment strategies.

To illustrate the growing importance of ETFs in the ownership of common stocks, we present descriptive statistics for S&P 500 and Russell 3000[4] stocks in Table 1. Due to the expansion of this asset class, ETF ownership of individual stocks has increased dramatically over the last decade. For S&P 500 stocks, the average fraction of a stock's capitalization held by ETFs has risen from 0.27% in 2000 to 3.78% in 2012. The table shows that the number of ETFs that follow the S&P500 index grew from 2 to about 50 during the same period. The average assets under management (AUM) for ETFs holding S&P 500 stocks in 2012 was $5bn. The statistics for the Russell 3000 stocks paint a similar picture.

In our analysis, we focus on ETFs that are listed on U.S. exchanges and whose baskets contain U.S. stocks. The discussion that follows applies strictly to these "plain vanilla" exchange traded products that do physical replication, that is, they hold the securities of the basket that they aim to track. We omit from our sample leveraged and inverse leveraged ETFs that use derivatives to deliver the performance of the index, which represent at most 2.3% of the assets in

[4] The Russell 3000 includes the largest 3000 stocks by market capitalization, reconstituted at the beginning of June each year.

the sector (source: BlackRock). These more complex products are studied by Cheng and Madhavan (2009), among others.

Similar to closed-end funds, retail and institutional investors can trade ETF shares in the secondary market.[5] However, unlike closed-end funds, new ETF shares can be created and redeemed. Because the price of ETF shares is determined by the demand and supply in the secondary market, it can diverge from the value of the underlying securities (the NAV). Some institutional investors (called "authorized participants," APs), which are dealers that have signed an agreement with the ETF provider, can trade bundles of ETF shares (called "creation units," typically 50,000 shares) with the ETF sponsor. An AP can create new ETF shares by transferring the securities underlying the ETF to the ETF sponsor. These transactions constitute the primary market for ETFs. Similarly, the AP can redeem ETF shares and receive the underlying securities in exchange. For some funds, ETF shares can be created and redeemed in cash.[6]

To illustrate the arbitrage process through creation/redemption of ETF shares, we distinguish the two cases of (i) ETF premium (the price of the ETF exceeds the NAV) and (ii) ETF discount (the ETF price is below the NAV). In the case of an ETF premium, APs have an incentive to buy the underlying securities, submit them to the ETF sponsor, and ask for newly created ETF shares in exchange. Then the AP sells the new supply of ETF shares on the secondary market. This process puts downward pressure on the ETF price and, potentially, leads to an increase in the NAV, reducing the premium. In the case of an ETF discount, APs buy ETF units in the market and redeem them for the basket of underlying securities from the ETF sponsor. Then the APs can sell the securities in the market. This generates positive price pressure on the ETF and possibly negative pressure on the NAV, which reduces the discount.

Creating/redeeming ETF shares has limited costs in most cases, especially for equity-focused funds. These costs include the fixed creation/redemption fee plus the costs of trading the underlying securities. Petajisto (2013) describes the fixed creation/redemption costs as ranging in absolute terms from $500 to $3,000 per creation/redemption transaction, irrespective of the

[5] Unlike premia and discounts in closed-end funds (e.g., Lee, Shleifer, and Thaler (1991), Pontiff (1996)), price divergence between the ETF and the NAV can be more easily arbitraged away thanks to the possibility of continuously creating and redeeming ETF shares. As a result ETF premia/discounts are order of magnitudes smaller than for closed-end funds.

[6] Creation and redemption in cash is especially common with ETFs on foreign assets or for illiquid assets, e.g., fixed income ETFs.

number of units involved. This fee would amount to about 3.4 bps for a single creation unit in the SPY (that is, 50,000 shares worth about $8.8 million as of October 2013), or 0.6 bps for five creation units. During our sample period (2000–2012), share creation/redemption occurs, on average, on 71% of the trading days. For the largest ETF, the SPY, flows into and out of the fund occurred almost every day in 2012 (99.2% of the trading days).

Arbitrage can also be undertaken by market participants who are not APs and without creation/redemption of ETF shares. Because both the underlying securities and ETFs are traded, investors can buy the inexpensive asset and short sell the more expensive one. For example, in the case of an ETF premium, traders buy the underlying securities and short sell the ETF. They hold the positions until prices converge, at which point they close down the positions to realize the arbitrage profit. Conversely, in the case of an ETF discount, traders buy the ETF and short sell the individual securities. ETF sponsors facilitate arbitrageur activity by disseminating NAV values at a 15-second frequency throughout the trading day. They do so because the smooth functioning of arbitrage is what brings about the low tracking error of these instruments. As a result of the low trading costs and availability of information, arbitraging ETFs against the NAV has become popular among hedge funds and high-frequency traders in recent years (Marshall, Nguyen, and Visaltanachoti (2010)). ETF prices can also be arbitraged against other ETFs (Marshall, Nguyen, and Visaltanachoti (2010)) or against futures contracts (Richie, Daigler, and Gleason (2008)).[7]

These institutional details, with some modifications, also apply to synthetic ETFs, which are more prevalent in Europe. These products replicate the performance of the index using total return swaps and other derivatives. As a result, creation and redemption are handled in cash. However, the secondary market arbitrage still involves transactions in the underlying securities. So, the potential for propagation of demand shocks from the ETF market to the underlying securities via arbitrage is also present among synthetic ETFs.

[7] To be precise, although these trading strategies involve claims on the same cash flows, they may not be arbitrages in the strict sense because they can involve some amount of risk. In particular, market frictions can introduce noise into the process. For example, execution may not be immediate, shares may not be available for short selling, or mispricing can persist for longer than the arbitrageurs' planned horizon for the trade. In the remainder of the paper, when we refer to ETF arbitrage, we are implying the broader definition of "risky arbitrage."

Finally, although we limit our analysis to ETFs that track equity indexes, the arbitrage process is an inherent characteristic of all types of ETFs. As a consequence, one should expect the effects of ETFs that we describe in this paper to play out for all types of underlying assets.

2.2 Theoretical Framework

The main testable hypothesis of the paper is that the arbitrage between ETFs and the securities in their baskets propagates liquidity shocks from the ETF market to the prices of the basket securities. As a consequence, non-fundamental volatility of the underlying securities increases due to ETF ownership.

We use Greenwood's (2005) model with risk-averse market makers to explain the channel of shock transmission that generates this testable conjecture. The market makers in the model can be thought of as the Authorized Participants or, more generally, the arbitrageurs in the ETF market. We apply this model to two assets with identical fundamentals: the ETF and the basket of underlying securities (whose market value is the NAV of the ETF). To illustrate, we imagine a situation in which the ETF price and the NAV are aligned at the level of the fundamental value of the underlying securities, as in Figure 1a. Then, a non-fundamental shock, such as an exogenous increase in demand, hits the ETF market. This type of shock could happen, for example, if a large institution receives inflows and scales up its existing ETF allocation. Arbitrageurs absorb the liquidity demand by shorting the ETF. Because they are risk averse, the arbitrageurs require compensation for the (negative) inventory in the ETF that they are taking on. Hence, the ETF price has to rise (Figure 1b). At the same time, to hedge their short ETF position, arbitrageurs take a long position in the securities in the ETF basket. Again, to compensate the arbitrageurs for the risk they take, the prices of the basket securities have to rise, as in Figure 1c. Eventually, when other sources of liquidity materialize or uncertainty is resolved, prices revert back to fundamentals (Figure 1d). In sum, shock transmission results from the trading of risk-averse investors who require compensation for holding assets in the two markets. To provide the investors with the required risk premium, prices have to adjust in both markets.

In Greenwood's (2005) model, the long and short hedging trades happen simultaneously (i.e., the movements in Figures 1b and 1c happen at the same time). Moreover, given that there is a unique market maker, two assets with identical payoffs always end up having the same price, and no discrepancy between the ETF price and the NAV can be present at any time. As a result, a strict adherence to the model would prevent the ETF price from ever deviating from the NAV. Although this simple theoretical framework allows us to describe the mechanism for liquidity shock transmission, we need a richer model to capture the fact that in reality the ETF price and the NAV can diverge for some time.

Cespa and Foucault (2012) provide a useful framework with multiple investor classes and some degree of market fragmentation. They assume three types of traders: liquidity demanders, who submit market orders in one of two markets, and two types of liquidity suppliers: market makers, who specialize in one asset class, and cross-market arbitrageurs, who trade securities in both markets. Arbitrageurs respond to misalignments in the prices of the assets in the two markets. The model is static in the sense that all investor classes trade in the same period. As a result, even with this model, price discrepancies between two identical assets cannot emerge. However, one can conceive a dynamic extension of the Cespa and Foucault (2012) framework in which trades occur sequentially. In the first period, there is a liquidity shock in one of the two assets that is accommodated by market makers via a price adjustment. In the next period, the market makers for the second asset observe the price realization of the first asset and adjust their own price. Cross-market arbitrageur trading occurs in the second period, bringing about price convergence between the two assets. In this dynamic framework, the prices of two identical assets can temporarily differ (in the first period). In this modified framework, arbitrageurs' risk aversion and hedging trades are still crucial for the transmission of liquidity shocks between the two markets.

The mechanism that we have just described generates predictions that partly overlap with those from an alternative scenario positing gradual price discovery after a shock to fundamentals. According to this alternative view, prices behave similarly to the description in Figure 1, but the trigger is a fundamental shock rather than a liquidity shock. Specifically, it is possible that price discovery takes place in the ETF market first, for example, because it is more liquid. Then, when fundamental information gets to the market, ETF prices adjust immediately, but the underlying securities' prices remain temporarily fixed ("stale pricing"). The slow adjustment of the NAV generates a sequence of price moves that resembles those in Figure 1. This situation is illustrated in Figure 2. The initial equilibrium (Figure 2a) is perturbed by a shock to the fundamental value of the ETF components (Figure 2b). If price discovery takes place in the ETF market, the ETF price moves first (Figure 2c) and the prices of the underlying securities move with a delay (Figure 2d).

Because stale pricing could be a relevant phenomenon, especially for the more illiquid underlying securities, one needs to assess whether liquidity shock propagation does take place. The crucial distinction between the liquidity shock propagation mechanism (Figure 1) and the alternative scenario with stale pricing (Figure 2) is that non-fundamental shocks induce a reversal in stock prices (Figure 1d). This does not happen if the initial shock is a fundamental one, as in the price discovery scenario. Hence, to disentangle the two hypotheses, in the empirical analysis we test for price reversals after arbitrage activity.

The hypothesis that ETF ownership increases volatility faces the challenge of clearly specifying the counterfactual. An alternative hypothesis is that, if ETFs were not available, the same investors would directly trade the underlying securities. According to this argument, ETFs are simply another vehicle through which the same clientele trades in the underlying securities. Grossman (1988) makes a related point about futures. He argues that the volatility of the prices of the underlying assets would be even higher in the absence of futures, because future markets, being more liquid, are better suited to absorb non-fundamental shocks.

Hence, the hypothesis that ETF ownership increases the rate of arrival of liquidity shocks needs the complementary assumption that ETFs attract a new clientele of investors, who would not otherwise trade the underlying securities. Theoretical support for this conjecture comes from Amihud and Mendelson (1986) and Constantinides (1986), who propose that investors with shorter holding periods self-select into assets with lower trading costs. Atkins and Dyl (1997) find support for this conjecture by showing that securities with lower bid-ask spread have higher trading volume. These theories and empirical evidence suggest that, due to the low trading costs of ETFs, a new clientele of high-frequency investors can materialize around the newly created securities. This clientele would not trade the less-liquid underlying assets if ETFs were not present.

Ultimately, whether low transaction costs of ETFs attract a clientele of high-frequency traders that increase the exposure of the underlying securities to non-fundamental shocks is an empirical question. A unique prediction of the new-clientele hypothesis is that ETF ownership is related to higher turnover in the underlying securities. This consideration motivates us to use turnover as an additional dependent variable, besides volatility, in our empirical tests.

3 Data

We use Center for Research in Security Prices (CRSP), Compustat, Bloomberg, and OptionMetrics data to identify ETFs traded on the major U.S. exchanges and to extract returns, prices, and shares outstanding. To identify ETFs, we first draw information from CRSP for all securities that have the historical share code of 73, which exclusively defines ETFs in the CRSP universe. We then screen all U.S.-traded securities in the Compustat XpressFeed and OptionMetrics data, identifying ETFs using the security-type variables, and merge this sample with the CRSP ETF sample.[8] Our initial sample consists of 1,883,124 daily observations for 1,673 ETFs between 1993 and 2012. Because very few ETFs traded in the 1990s, we restrict the sample to the 2000–2012 period. Among other statistics, Table 1 reports stock-level averages of the number of ETFs and of the AUM of the ETFs, broken down by the S&P 500 and Russell 3000 universes. The table shows that the number of ETFs holding the average stock increased dramatically since the year 2000, for both S&P 500 and Russell 3000 stocks. In 2000, there were two ETFs per stock in both universes, on average, compared to 49 and 27 in 2012 for the average S&P 500 and Russell 3000 stock, respectively. Furthermore, as the total market capitalization of ETFs increased, the average ownership of ETFs per stock increased from 0.3% in 2000 to 3.8% in 2012.

We use total shares outstanding at day-end to compute the daily market capitalization of each ETF and to measure the net share creations/redemptions of each ETF at the daily level. Because CRSP shares outstanding figures are stale during the month, we assessed the accuracy of three databases that provide shares outstanding data at a daily frequency: Bloomberg, Compustat, and OptionMetrics. Thanks to direct validation by BlackRock, we concluded that Bloomberg is more accurate and timely in updating ETF shares outstanding when newly created or redeemed shares are cleared with the Depository Trust & Clearing Corporation (DTCC). On many occasions, Compustat and OptionMetrics shares outstanding data lag Bloomberg by up to

[8] Note that at the time of the first draft of this paper in 2011, the CRSP-Compustat merged product did not correctly link ETF securities in the CRSP and Compustat universes. For this reason, we use historical CUSIP and ticker information to map securities in the CRSP, Compustat, and OptionMetrics databases.

three and sometimes five days. Therefore, Bloomberg is our primary source for shares outstanding and the related net flow measures. We use Compustat and OptionMetrics to complement the ETF series when there are gaps in the Bloomberg data.

We then obtain net asset value (NAV), in addition to fund styles (objectives) and other characteristics, from the CRSP Mutual Fund and Morningstar databases. We restrict our sample to ETFs that invest primarily in U.S. domestic equity stocks, because they are not plagued with stale pricing issues (global equity or bond ETFs) or other issues (short bias, volatility, and futures-based ETFs, commodities, etc.). Therefore, we exclude leveraged, short equity ETFs, and all ETFs that invest in international or non-equity securities, or in futures and physical commodities. We also eliminate active and long/short ETFs as well as dedicated short bias funds and focus on plain vanilla U.S. domestic long equity ETFs. To do so, we use both CRSP Style Codes and Lipper prospectus objective codes in the CRSP Mutual Fund Database and restrict our sample to the fund objectives that span broad-based U.S. Diversified Equity funds and U.S. sector ETFs that invest in equities (e.g., U.S. companies investing in oil and natural resources vs. those investing in oil or commodity futures).[9] We end up with 660 U.S. equity ETFs, for which we obtain quarterly holdings information using Thomson-Reuters Mutual Fund holdings database. ETFs are subject to Investment Company Act reporting requirements, and similar to mutual funds, they have to disclose their portfolio holdings at the end of each fiscal quarter.[10] We use these data to align ETF ownership every month using the most recently reported holdings. Then, for every stock, we sum the total ownership by various ETFs to construct our ETF holdings measure. We also use Thomson-Reuters Mutual Fund holdings database to compute the ownership by index funds, active funds, and total mutual fund ownership excluding ETFs. To do that, we use the index fund flag in CRSP Mutual Fund database, and merge it with Thomson-

[9] The Lipper Asset Code is not sufficient to accurately filter for U.S. domestic equity funds, because the Equity Funds code comprises of a wide array of U.S. and global funds that implement various direct investment or alternative/inverse strategies. Instead, we use Lipper Objective Code classifications that are assigned by Lipper to a specific population of equity funds and are based on how the fund invests by looking at the actual holdings of the fund to determine market cap and style versus a benchmark. We restrict our sample to the following Lipper Objective Codes: Board Based U.S. Equity: S&P 500 Index Objective Funds, Mid-Cap Funds, Small-Cap Funds, Micro-Cap Funds, Capital Appreciation Funds, Growth Funds, Growth and Income Funds, and Equity Income Funds ('CA' ,'EI' ,'G' ,'GI' ,'MC' ,'MR' ,'SG' ,'SP' respectively). We also include Sector Funds that invest in U.S. companies: Basic Materials, Consumer Goods, Consumer Services, Financial Services, Health/Biotechnology, Industrials, Natural Resources, Real Estate, Science and Technology, Telecommunications, Specialty/Miscellaneous Funds, and Utilities (BM, CG, CS, FS, H, ID, NR, RE, TK, TL, S, and UT, respectively).

[10] We find that Thomson Mutual Fund Ownership data is more reliable and more complete than CRSP Mutual Fund Holdings until mid-2010.

Reuters holdings data using WRDS MFLinks. Similar to how ETF ownership is calculated, we compute monthly index and active fund ownership by using the most recently reported holdings.

We use Trade and Quote database (TAQ) data to compute stock-level volatility at a daily frequency from second-by-second data. For each stock, we compute a return in each second during the day using the last trade price at the end of each second during market hours (between 9:30 am and 4:00 pm). Then, we compute the standard deviation of those second-by-second returns as the intraday volatility measure.[11] Daily turnover is computed as CRSP volume divided by shares outstanding.

We follow the methodology in O'Hara and Ye (2011) and use TAQ data to compute the variance ratio as the absolute value of the 15-second log returns divided by three times the variance of 5-second log returns minus one. We decide on the 15-second return interval as the base case since ETF intraday indicative values used by arbitrageurs are typically disseminated every 15 seconds.

Some ETFs are traded until 4:15 pm (Engle and Sarkar (2006)), but the major U.S. stock markets close at 4:00 pm. Thus, to ensure that ETF prices and the NAV are computed at the same time, we obtain 4:00 pm ETF prices from the TAQ feed as the last trade in the ETF at or before 4:00 pm. Then, we compute ETF mispricing as the difference between the ETF share price and the NAV of the ETF portfolio at 4:00 pm. Mispricing is expressed as a fraction of the ETF price.[12] Part of our analysis is carried out at a monthly frequency. To this end, we compute volatility at a monthly frequency from the standard deviation of daily returns within the month.

We extract stock lending fees from the Markit Securities Finance (formerly Data Explorers) database. The database contains about 85% of the OTC security lending market, with historical data going back to 2002. In constructing the aggregate security loan fee, Markit extracts the agreed fees from contract-level information and constructs a fee value that is the volume weighted average of each contract-level security loan fee. We use the variable that reports the average lending fee over the prior seven days.

[11] We also compute intraday volatility using intraday returns based on NBBO midpoints, and the results are similar.
[12] The label "mispricing" does not mean to imply that either the ETF or the NAV are correctly priced, while the other is not. We are just complying with the standard jargon in the industry and taking a shortcut with respect to the more cumbersome label of "discount/premium."

Table 2 reports summary statistics for the variables that we use in the regressions. Panel A presents summary statistics for the day-stock level sample. Panel B presents summary statistics for the month-stock level sample. Panel C presents a correlation table for the daily sample. Panel D presents summary statistics for the variables used in the returns regressions at the daily frequency. We further describe these variables in later sections.

4 The Effect of ETF Ownership on Volatility and Turnover

4.1 Identification from Cross-Sectional and Time-Series Variation in ETF Ownership

The focus of our tests is whether ETF ownership leads to an increase in the volatility of the underlying securities. The first source of identification is the variation in ETF ownership across stocks and over time.

ETF ownership of stock *i* at time *t* is defined as the sum of the dollar value of holdings by all ETFs investing in the stock divided by the stock's capitalization. In formulas

$$ETF\ owernship_{i,t} = \frac{\sum_{j=1}^{J} w_{i,j,t} AUM_{j,t}}{Mkt\ Cap_{i,t}}, \tag{1}$$

where *J* is the set of ETFs holding stock *i*; $w_{i,j,t}$ is the weight of the stock in the portfolio of ETF *j*; and $AUM_{j,t}$ is the assets under management of ETF *j*.

From Equation (1), it appears that variation in ETF ownership across stocks and over time primarily comes from three sources. First, stocks are typically part of multiple indices (e.g., a stock might be part of the S&P 500, the S&P 500 Value, the Russell 3000, and sector indices). Second, there is variation in ETFs' assets under management; thus, the dollar amount that the ETFs invest across stocks varies. Third, there is variation in weighting schemes. The S&P 500 and many other indexes are capitalization-weighted, but the Dow Jones is price-weighted. Our identifying assumption is that variation in ETF ownership resulting from these three sources is exogenous with respect to our dependent variables of interest, stock volatility and turnover, especially when stock-level controls (such as market capitalization and liquidity) are included in the regression. Conditioning on a given universe, such us the S&P 500 and the Russell 3000, characteristics like volatility and turnover play no role in determining the sub-index to which a stock belongs (e.g., S&P 500 Growth or Value or sector indices).

One could argue that investors' demand for ETFs, which determines AUM, may relate to fundamental information, which also affects volatility and turnover. However, the way these AUM translate into demand for individual stocks is arguably exogenous, because it depends on the way in which indices are computed. Given these considerations, we believe that the identifying assumption is well founded.

To further ensure that our results are driven by exogenous variation in ETF ownership, in our preferred specifications we include stock-level fixed effects. In these regressions, the variation in ETF ownership is for the same stock over time while controlling for unobservable characteristics that are potentially correlated with the dependent variable.

A caveat to this design is a potentially mechanical relation between ETF ownership and volatility due to the relation between ETF ownership and stock size. Specifically, based on Equation (1), we can anticipate that there is a mechanical negative correlation between ETF ownership and stock market capitalization. This can happen if the weights at the numerator do not grow fast enough with capitalization to compensate for the increase in the denominator. Given that market capitalization is negatively correlated with volatility (Table 2, Panel C), which is one of the main dependent variables of interest in our analysis, the negative relation between ownership and size (Table 2, Panel C) could induce a spurious positive relation between ownership and volatility. To filter out this mechanical link, we include controls for market capitalization in all of our analyses.

Overall, these arguments suggest that there is exogenous variation in ETF ownership that can be used to identify the effect of ETFs on volatility. We isolate this exogenous component of ETF ownership by controlling for stock size and fixed effects.

4.2 ETF Ownership, Intraday Volatility, and Turnover

We start by looking at whether ETF ownership has an impact on intraday volatility, which is the frequency at which arbitrage takes place. Using daily stock-level observations, we regress intraday volatility, computed using second-by-second returns from TAQ, on prior-day

ETF ownership as well as on prior-day controls for size and liquidity. The controls for liquidity are the inverse of the stock price, the Amihud (2002) measure of price impact, and the bid-ask spread expressed as a percentage. We also include day fixed effects in all regressions and add stock fixed effects in even numbered columns. Standard errors are clustered at the stock level.

Also stemming from the liquidity-shock-propagation hypothesis is the implication that the securities in the ETF baskets inherit the high-turnover clientele of the ETFs. Hence, to test this prediction, we regress turnover on ETF ownership in specifications that mirror those for volatility. Turnover is computed as the CRSP dollar volume divided by market capitalization.

First, we limit our sample to the S&P 500 stock universe. The volatility results are presented in Table 3, Columns (1) and (2). The regressions show that intraday volatility is significantly related to ETF ownership. Column (2) indicates that a one standard deviation increase in ETF ownership is associated with higher volatility by 19% of a standard deviation.[13] The effect seems economically important.

In Columns (3) and (4) of Table 3, we explore whether ETF ownership also affects stock turnover. The estimates reveal a positive and significant relation between ETF ownership and turnover. Column (4) shows that a one standard deviation increase in ETF ownership is associated with higher turnover by about 19% of a standard deviation.[14] Again, the effect seems economically large and supports the view that ETFs attract a high-turnover clientele which is passed down to the underlying securities.

In Columns (5) to (8), we repeat these tests for the sample of Russell 3000 stocks. After controlling for stock fixed effects, we again find a significant relation between ETF ownership and stock volatility. In both turnover specifications, the estimates are statistically significant. In this sample, however, the effects are substantially smaller than for large stocks. For example, Column (6) shows that a one standard deviation increase in ETF ownership raises intraday volatility by about 8% of a standard deviation. Quite plausibly, arbitrageurs are less likely to rely on small stocks to replicate ETF baskets. Hence, small stocks' prices and volume are less impacted by ETF ownership.

[13] (0.243 * 0.014) / 0.018 = 0.1890.
[14] (11.631 * 0.014) / 0.853 = 0.1909.

The results in Table 3 provide our first evidence that stock volatility is significantly related to ETF ownership. We consider variation in ETF ownership as exogenous with respect to the dependent variables, especially after controlling for stock characteristics and fixed effects. Hence, we feel that we can attribute a causal interpretation to the estimates in Table 3. Further corroboration on the causal link between ETF ownership and volatility comes from the alternative identification strategy in Section 4.5 which uses a regression discontinuity design.

4.3 Lower Frequency Effect and Controls for Mutual Fund Ownership

Our results in Table 3 show that ETF ownership is associated with higher return volatility within the day. However, a legitimate concern is that while it is possible that ETFs affect the microstructure of trading for the underlying securities, these effects are washed out over longer horizons. To examine this possibility, we study whether the effects that we identify are a short-lived phenomenon (e.g., induced by high-frequency traders) or whether these effects also exist at frequencies that are relevant for long-term investors. We define our explanatory variables at the monthly frequency and construct the dependent variable, volatility, using the daily return observations within a month. In this way, we can study whether ETF ownership impacts the volatility of daily returns.

Table 4 shows a regression of daily stock volatility in a given month on the average ETF ownership of the stock within the month. We use stock-level controls to absorb effects that could induce a mechanical link between ownership and our dependent variable. To this purpose, we include (the logarithm of) the market capitalization of the stock as well as the same controls for liquidity as in Table 3. We cluster standard errors both at the date and the stock levels. In addition, date and stock fixed effects are included in all the specifications.

In Columns (1) to (3), we limit the sample to S&P 500 stocks, and in Columns (4) to (6), we extend it to Russell 3000 stocks. The regressions in Columns (1) and (4) show that stock volatility is positively related to ETF ownership and that the effect is stronger for large stocks. In Column (1), a one standard deviation increase in stock ownership for S&P 500 stocks (1.44%) is associated with a 20 bps increase in daily volatility, which represents 16% of a standard deviation of the dependent variable.[15] The economic significance is therefore large. Extending the universe to smaller stocks (Column (4)), the effect is diluted, amounting to about 5% of a standard deviation.[16] This finding confirms the evidence for intraday volatility in Table 3.

The prior results may raise the question about the extent to which ETF ownership captures a separate effect from the ownership of other institutional investors. Among these, open end funds are the most similar to ETFs because they are also exposed to daily flows. In Columns (2) and (5), we include a control for total mutual fund ownership, which is measured using quarterly holdings and end-of-prior-month stock capitalization. The coefficient on mutual fund ownership is positive and significant suggesting that other institutions may be affecting volatility in a similar way to ETFs. However, the estimate is at least an order of magnitude smaller than the effect of ETF ownership, which remains intact. This finding suggests that the arbitrage channel, which is specific to ETFs, plays the dominant role in affecting stock volatility.

Among open-ended funds, index funds are the closest to ETFs as they are also passively tracking a basket of stocks. While active funds retain some discretion on the amount of trading, index funds are forced to scale up and down their portfolios in response to daily flows. Columns (3) and (6) include separate controls for active and index fund ownership. While both positive and significant, the coefficient for index funds is significantly larger, consistent with our priors. The estimates for ETF ownership is only slightly impacted by these controls and remains by far the most important effect.

Overall, the evidence suggests that the effect of ETFs on volatility persists beyond the intraday horizon. The daily volatility we study in this section is relevant for investors, such as mutual funds, that do not trade at high frequencies but still reallocate their portfolio on a daily basis. Also important, ETF ownership impacts volatility well beyond the general effect of ownership by mutual funds. While index funds resemble ETFs in that they adjust their ownership daily to track a basket of stocks, not being listed on an exchange they are not exposed to high-frequency arbitrage. The significantly larger impact of ETF ownership on volatility likely reflects this peculiarity of ETFs.[17]

---

[15] (0.146 * 1.440) / 1.290 = 0.1630.

[16] (0.044 * 1.730) / 1.490 = 0.0511.

[17] Although we include stock fixed effects in our main regressions, concerns about identification may linger in case unobservable stock characteristics determine ETF ownership and stock volatility in sub-periods of a stock's

4.4 Identifying the Impact on Non-Fundamental Volatility: Variance Ratios and Price Reversals

The finding that higher ETF ownership is associated with increased volatility is not necessarily evidence in favor of the hypothesis that ETFs increase the noise in the prices of the underlying securities. For example, Amihud and Mendelson (1987) provide a simple model in which the volatility of trading prices is positively related to the speed at which prices adjust to fundamentals. If ETF arbitrage makes prices adjust more promptly to fundamentals, the model yields the prediction that the *fundamental* volatility of the underlying securities goes up. This increase in volatility differs from the prediction of the hypothesis that is tested in this paper, which instead focuses on *non-fundamental* volatility, or noise in the definition of Black (1986).

O'Hara and Ye (2011) use variance ratios to measure price efficiency in intra-day data. On each day *t*, stock *i*'s variance ratio is constructed as:

$$VR_{i,t} = \left| \frac{Var(r_{k,i,t})}{k \cdot Var(r_{1,i,t})} - 1 \right| \tag{2}$$

where the numerator is the variance of *k*-period log returns on day *t* and the denominator is *k* times the variance of single period log returns on day *t* (also see Lo and MacKinlay (1988)). As argued by these authors, in an efficient market, the variance ratio should be closer to zero as prices are expected to follow a random walk and the fraction in Equation (2) approaches one. This device can be used to test the impact of ETFs on non-fundamental volatility. If ETFs add noise to prices, the variance ratios should increase with ETF ownership. In other words, the liquidity shocks originating in the ETF market boost the mean-reverting component in the prices of the underlying securities.

In our application, we measure single-period returns from transaction prices at five-second intervals and choose $k = 3$, so that multi-period returns are measured over fifteen-second intervals. While this choice of time-interval is to some extent arbitrary, it can be justified based on the observation that information about NAV is disseminated by ETF sponsors at fifteen

---

appearance in the sample. To address this concern, we run the regressions in first differences. The main results remain significant, as shown in Appendix Table A1.

second intervals to facilitate high-frequency arbitrage. This frequency is therefore relevant to capture the effect of arbitrageurs on the underlying stock prices.

Table 5 reports estimates from regressions of the stock-level variance ratio on ETF ownership in the prior day and the same set of controls as in the previous table. The results point unambiguously to a positive and significant relation between ETF ownership and variance ratios. This evidence suggests that, at this frequency, prices of stocks with higher ETF ownership are farther away from a random walk and, therefore, contain more noise. For S&P 500 stocks, the economic magnitude is large and in line with the effects from Table 3. Based on Column (2) of Table 5, a one-standard deviation increase in ETF ownership is associated with an increase of about 12% of a standard deviation in the variance ratio.[18] Consistent with the results in Table 3, the effect is reduced, but still significant, when the universe is extended to smaller stocks.

Another way to test whether ETFs add noise to prices is to focus on price reversals. In Section 2.2, we argue that if the initial price impact is reversed, the trigger is a non-fundamental shock (as in Figure 1). Instead, if prices remain at the new level the initial shock results from new fundamental information (as in Figure 2). Disentangling these two scenarios is another way to test whether ETF arbitrage induces a mean-reverting component in stock prices. Further, by casting the analysis at the daily frequency, we test whether the noise in prices survives beyond the intra-day horizon used for the variance ratio tests in Table 5.

We use stock-level ETF flows at the daily frequency as a conditioning variable to identify prices reversals. As explained above, ETF flows (redemptions and creations) are the result of APs' arbitrage activity. Stock level flows are defined as the weighted average of the daily flows in the ETFs that own the stock. The weights are the fraction of ownership in the stock by each ETF. Daily ETF flows are measures as a fraction of prior day assets under management.

On the day in which flows occurs, we expect a price move in the same direction as the flows, irrespective of whether the motive for trade is fundamental or non-fundamental. To the extent that at least part of the originating shock is non-fundamental, a reversal should occur in the next days. To capture this behavior we regress returns at different horizons on stock level flows. We include the usual stock-level controls and time fixed effects. The standard errors are

[18] (1.123 * 0.014) / 0.134 = 0.1173.

clustered at the stock level to correct for the autocorrelation of residuals induced by overlapping observations for multiday returns.

The evidence in Table 6 is broadly consistent with the transmission of non-fundamental shocks. In Column (1), we observe that the first-day effect of ETF flows is positive and significant.[19] In the twenty trading days following the flows, prices partially revert suggesting that the initial shock adds a mean-reverting component to stock prices. In terms of magnitude, after twenty days, about 45% of the initial shock has reverted. Extending the horizon farther out to forty days does not increase the magnitude of reversals (not reported), which suggests that on average flows convey fundamental and non-fundamental shocks in roughly equal shares.

In sum, the evidence in this section suggests that the positive link between ETF ownership and volatility, which we report in Tables 3 and 4, is consistent with an increase in non-fundamental volatility. Specifically, ETFs appear to add a mean-reverting component to stock prices both intraday (Table 5) and at the daily frequency (Table 6).

4.5 Identification Using Regression Discontinuity Design

The identification based on cross-sectional and time-series variation in ETF ownership, which underlies the results in Tables 3 and 4, can be flawed if the stock level controls fail to capture characteristics that co-determine ETF ownership and volatility. While this is unlikely because our preferred specifications include stock level fixed effects, in this section we corroborate our main results with an alternative approach to identification.

Chang, Hong, and Liskovich (2013) devise an identification strategy that exploits the exogenous variation in the membership of the Russell 1000 and Russell 2000 indexes and is cast within the regression discontinuity (RD) framework. The identifying assumption in RD is that the individuals (in our case, the firms) have imprecise control over the treatment variable (in our case, the assignment to an index). If this is the case, the treatment is "as good as" randomly assigned around the cutoff (Lee and Lemieux (2010)).

[19] The magnitude of the effect is economically important. For the S&P 500 sample (Column (1)), a one-standard deviation move in ETF flows is associated with a contemporaneous return of about 12 bps (=12.418*0.01). This seems like a large effect given that the mean daily return in the sample is 0.056%.

This argument nicely fits the Russell index assignment experiment. The Russell 1000 index is comprised of the top 1000 stocks by market capitalization, while the Russell 2000 includes the next 2000 stocks. The index is reconstituted in June of each year only based on end-of-May stock capitalization, hence no discretion is involved in index assignment. Index composition is kept constant for the rest of the year. For stocks in a close proximity of the cutoff, changes in index membership can be considered as exogenous events as they result from random variation in stock prices at the end of May.

Chang, Hong, and Liskovich (2013) corroborate the validity of RD design in the context of the Russell 1000/Russell 2000 experiment. In particular, they show that firms in a close neighborhood of the cutoff are similar in terms of the distribution of some baseline covariates (Returns in the month before reconstitution, Assets, EPS), which is a necessary condition for the validity of the RD design. Also, they test for manipulation of the ranking variable using McCrary's (2008) methodology and reject this hypothesis. These authors also show that, although the amount of passive assets benchmarked to the Russell 1000 is 2 to 3.5 larger than those tracking the Russell 2000, the weights of the top stocks in the Russell 2000 are about 10 times larger than those for the bottom stocks in the Russell 1000. As a result, there is a significantly larger amount of passive money tracking the top Russell 2000 stocks.

We exploit the latter finding to identify discontinuous variation in ETF ownership around the cutoff between the two indexes. In particular, we focus on stocks that switch between indexes and use the event of a switch as an instrument for ETF ownership. Then, we regress our outcome variable, stock volatility, on instrumented ETF ownership. To identify the effect of interest, we rely on the insight from RD that variation in ETF ownership around the cutoff can be considered as exogenous. One additional identifying assumption that needs to be satisfied is excludability, that is, the requirement that the instrument affects the outcome variable only through the treatment variable. In our context, this translates into the condition that a switch in index membership only affects volatility through ETF ownership. Below, we discuss instances in which this assumption may not be satisfied and propose some solutions.

Following Chang, Hong, and Liskovich (2013), we cast our analysis at the monthly frequency. The first index reconstitution in our sample occurs in May 2000, so we include all months between June 2000 and December 2012. Although we use actual switches between indexes as our instrument, we still need to restrict the sample to stocks in a close neighborhood of the cutoff. Rules for defining the cutoff changed during the sample period. The cutoff used to be simply the 1000th position in terms of market capitalization until 2006. Starting with the 2007 reconstitution, Russell Inc. adopted a banding rule whereby stocks only switch from their current index if they move beyond a 5% range around the market capitalization percentile of the 1000th stock. To account for this effect, we use membership data and market capitalization directly provided by Russell Inc. to compute index switches and cutoffs for every year in the sample. Finally, Chang, Hong, and Liskovich (2013) argue that the optimal bandwidth is about 100 stocks around the cutoff. To be conservative, we consider bandwidths of 50, 100, 150, and 200 stocks on each side of the cutoff.

We carry out a two-stage least square estimation. In each stage, we run our regressions on two separate groups of stocks: those that in May before index reconstitution are in the Russell 1000 and those that are in the Russell 2000. The sample composition is kept constant for all the months between June (the first month after index reconstitution) and May of the next year. The first stage consists of a regression of ETF ownership on an indicator variable for whether the stock switches index membership in June. For the Russell 1000 sample, the indicator variable flags stocks that switch to the Russell 2000. Vice versa, for the Russell 2000 sample the dummy captures a switch to the Russell 1000. In the second stage, for the same two separate groups of stocks, we regress daily volatility on the fitted value of ETF ownership from the first stage.

In all specifications, we add the usual set of controls for size and liquidity. Although a well-specified RD design does not require the inclusion of covariates for identification, Lee and Lemieux (2010) suggest that covariates could help in improving estimation efficiency. Note, however, that while have time fixed effects, we do not include stock fixed effects because identification in RD is inherently cross-sectional. Standard errors are clustered at the month level.

Table 7, Panel A, has the first stage regressions. We separately consider stocks that belong to the Russell 1000 before index reconstitution (Columns (1), (3), (5) and (7)) and stocks belonging to the Russell 2000 before index reconstitution (Columns (2), (4), (6) and (8)). The independent variable of interest is an indicator for whether the stock switches to the other index. The dependent variable (ETF ownership) is measured in each month following the index reconstitution. To illustrate the setting, consider Column (1). The sample includes stocks that are in Russell 1000 in May (prior to the reconstitution). We use the end-of-May cutoff to determine the stocks that are included in the sample (± 50 stocks around the cutoff). The stocks are kept in the sample in all months between June (after reconstitution) and May of next year. The indicator variable flags the stocks that switch to the Russell 2000 after reconstitution. The estimate in Column (1) suggests that ETF ownership in the twelve months after reconstitution increases for the stocks switching to the Russell 2000, which is expected given that the amount of passive money tracking these stocks is significantly larger when they change index.

Across bandwidths, the estimates unambiguously point that ETF ownership increases for stocks switching to the Russell 2000 and decreases for stocks moving to the Russell 1000. The magnitudes are similar for moves in either direction (especially for bandwidths above 50) and suggest that top Russell 2000 members have ETF ownership larger by, on average, about 45 bps than bottom Russell 1000 stocks. These results confirm the evidence in Chang, Hong, and Liskovich (2013) that the amount of passive money tracking the top Russell 2000 stocks is significantly larger.

Table 7, Panel B, reports the second stage estimates of the effect of ETF ownership on volatility. Mirroring the layout in Panel A, the instruments are indicators for a switch to either index and the sample is also restricted to members of either index before reconstitution. The effect of ETF ownership on volatility is significant across most samples and bandwidths. The estimates are larger for stocks that are included in the Russell 2000 than for switchers to the Russell 1000. To explain these different magnitudes, one could speculate that, once they appear onto the radar screen of arbitrageurs as top members of the Russell 2000, stocks do not immediately leave arbitrageurs' portfolios as they could still prove to be valid hedging instruments.

The magnitudes are considerably larger than those from prior tables (e.g., compare to Table 4, which is also based on monthly data). However, it is known that the RD estimates can be interpreted as a weighted average treatment effect, where the weights are the relative ex ante probability that the value of an individual's assignment variable will be in the neighborhood of the threshold (Lee and Lemieux (2010)). In other words, the effect we measure is tilted towards stocks that are highly likely to switch indexes, that is, stocks that can experience a drastic change from being not considered in arbitrageurs' strategies to having top weights in their replicating portfolios. For this reason, we prefer to emphasize the magnitudes from the prior tables as they are more general and more conservative.

Another good practice in RD is to include controls for polynomials of the ranking variable, which is market capitalization in our application. A sign of a well-specified experiment is the fact that the estimates are stable when different degrees of the polynomials are included (Lee and Lemieux (2010)). In Panels A and B, we control for a linear specification of the ranking variable. Panel C replicates the IV estimation with a quadratic polynomial. Reassuringly, the estimates are in the same ballpark as in Panel B.

Finally, we come back to the validity of the exclusion restriction in our context. This could be violated, for example, if ownership by other institutions tracking the index affects volatility beyond ownership by ETFs. We replicate our analysis including ownership by active funds and index funds and results are unaffected (Appendix Table A2). Another case of violation is where inclusion among the top stocks in the Russell 2000 makes a firm more visible. In that case, prices could react more quickly to fundamental information and returns could become more volatile (see the argument in Section 4.3). According to this story, price efficiency would be increased. As in Section 4.3, we use variance ratios (computed using intra-day data and averaged over the month) to measure noise in prices and replicate the RD estimation. Results in the Appendix Table A3 show that for the set of switchers to the Russell 2000, variance ratios are increasing in ETF ownership, which suggests that that price efficiency actually decreases after inclusion.

Overall, the RD design provides us with additional confidence in the causal interpretation of the effect of ETF ownership on stock-level non-fundamental volatility. For the rest of the analysis, we go back to using the pooled regressions, as they provide more conservative estimates of the effect on the entire sample of stocks.

5 Exploring the Arbitrage Channel

As discussed in Section 2, we want to investigate whether ETFs propagate demand shocks to the underlying securities. If this were the case, a new layer of liquidity shocks would hit the basket securities. In Section 4, we provide evidence consistent with this conjecture by showing that stocks with higher ETF ownership display higher volatility and turnover. In this section, we study more closely the arbitrage channel.

5.1 Stock Volatility, ETF Ownership, and Arbitrage Activity

Arbitrage occurs in two ways. At high frequencies, arbitrageurs take long and short positions in ETFs and the underlying baskets and wait for price convergence. At lower frequencies, Authorized Participants create and redeem ETF shares to profit from mispricing. In both cases, arbitrageurs and APs react to price discrepancies between the ETF price and the NAV (ETF mispricing). Hence, in our first set of tests, we use stock-level mispricing as a proxy for arbitrage trading. Then, in a second set of tests, we focusing more closely on APs' activities and measure arbitrage trading using creation and redemption of ETF shares (i.e., ETF flows).

5.1.1 Arbitrage Trades following ETF Mispricing

In this analysis the main explanatory variable is stock-level absolute mispricing, which is a signal for arbitrage profitability and therefore it proxies for the potential amount of arbitrage volume. For stock *i* on day *t*, absolute mispricing is defined as the weighted average of the mispricing of the ETFs holding the stock:

$$abs\left(ETF\ mispricing_{i,t}\right) = \frac{\sum_{j=1}^{J}\left|Mispricing_{j,t}\right| * ETF\ owernship_{i,j,t}}{\sum_{j=1}^{J} ETF\ owernship_{i,j,t}}, \tag{3}$$

where J is the set of ETFs holding stock i at time t, and $Mispricing_{j,t}$ is the difference between the ETF price and its NAV, scaled by the ETF price and measured using closing prices. The absolute value is motivated by the fact that arbitrage responds to both positive and negative levels of mispricing.

The question is whether the main effect of ownership on volatility is larger for stocks that are more exposed to arbitrage, as proxied by mispricing. Then, our regression specification is:

$$Volatility_{i,t} = \alpha + \beta_1\, abs(ETF\ mispricing_{i,t-1}) * ETF\ ownership_{i,t-1}$$
$$+\beta_2\, abs(ETF\ mispricing_{i,t-1}) + \beta_3\, ETF\ ownership_{i,t-1} \qquad (4)$$
$$+\beta_4\, Controls_{i,t-1} + Stock\ FE + Day\ FE + \varepsilon_{i,t}.$$

We run a similar specification using stock turnover as the dependent variable. We use the same controls as in Table 3, and standard errors are clustered at the stock level. The main variable of interest in equation (4) is the interaction between ownership and mispricing.[20]

Table 8, Panel A, presents the regressions. In Column (1), we observe that intraday volatility increases with the ETF ownership, as previously found. The new result is that the effect is significantly stronger for stocks with high ETF mispricing, which is reflected in the slope on the interaction. For stocks that have close to zero ETF ownership, the effect of ETF mispricing is minimal. A one standard deviation increase in abs(ETF mispricing) is associated with an increase of 0.4% of a standard deviation in volatility.[21] However, if ETF ownership is at its mean (1.9%), the effect is much larger: a one standard deviation increase in abs(ETF mispricing) is associated with an increase of 85.4% of a standard deviation in volatility.[22]

In Column (2) we repeat the analysis, this time controlling for lagged intraday volatility, and its interaction with ETF ownership. This specification addresses the concern that mispricing is mechanically higher for ETFs that contain more volatile stocks in their baskets. The significant relation between volatility and lagged mispricing remains significant after controlling for one-day lagged volatility and its interaction with ETF ownership. This result suggests that a mechanical link between volatility and mispricing is not driving our results.

The effect on intraday turnover is large as well (Column (3)). For ETF ownership approaching zero, a one standard deviation increase in lagged absolute mispricing is associated with higher intraday turnover by 0.3% of a standard deviation.[23] However, when ETF ownership is at its mean, intraday turnover is higher by 26.3% of a standard deviation.[24] Also in this case, the results remain statistically significant after controlling for the previous' day turnover and its interaction with ETF ownership (Column (4)).

[20] We use lagged end-of-day mispricing in our tests because it proxies for arbitrage that takes place during day t. Using day-t mispricing instead does not materially affect the results.
[21] 0.006 * 0.013 / 0.018 = 0.0043.
[22] (0.006 * 0.013 + 42.035* 0.013 * 0.019) / 0.018 = 0.8543.
[23] (0.207 * 0.013) / 0.853 = 0.0032.
[24] (0.207 * 0.013 + 896.893 * 0.013 * 0.019) / 0.853 = 0.2628.

Comparing the coefficients of the variables of interest (ETF ownership, and its interaction with the absolute value of the mispricing) indicates that about 40% to 50% of the correlation between volatility and mispricing is driven by factors other than a mechanical relationship.

Although the results for the S&P 500 sample are very strong both statistically and economically, the corresponding results for the Russell 3000 are not significantly different from zero, confirming the prior evidence of a weaker effect on smaller stocks. Overall, these results suggest that the arbitrage of ETF mispricing is an important channel to explain the impact of ETF ownership on volatility, especially for large stocks.

5.1.2 A Direct Measure of Arbitrage Activity by APs

Next, we more directly test the impact of ETF arbitrage through creation and redemption activity by APs. We measure stock-level flows using the following definition:

$$abs\left(ETF\ Flows_{i,t}\right) = \sum_{j=1}^{J} \frac{\sum_{j=1}^{J} \left|\frac{Fund\ flows_{j,t}}{AUM_{j,t-1}}\right| * ETF\ owernship_{i,j,t}}{\sum_{j=1}^{J} ETF\ owernship_{i,j,t}}. \qquad (5)$$

For each stock *i* and day *t*, we sum the product of the percentage of flows into the ETFs that own the stock and the percentage ownership of the ETF in the stock. For example, if ETF *j* experiences a flow of 1% and owns 10% of stock *i*, the stock is likely to experience a demand for 1% * 10% = 0.1% of its shares. Because both positive (share creation) and negative (share redemption) flows represent arbitrage activity (that potentially can increase volatility), in equation (5) we take the absolute value of the flows.

Our specification resembles equation (4), but we replace $abs\left(ETF\ Flows_{i,t}\right)$ with $abs\left(ETF\ mispricing_{i,t-1}\right)$. Table 8, Panel B, presents the results of the regressions. We first consider the S&P 500 sample (Columns (1) through (4)). The main effect of ETF ownership on stock volatility (Column (1)) remains positive and significant. Moreover, the effect is magnified for stocks with higher flows. When ETF ownership is at its mean, a one standard deviation increase in absolute ETF flows translates into volatility that is higher by 3.7% of a standard deviation.[25]

The effect of ETF ownership interacted with flows on stock turnover is similar in magnitude and significance (Column (3)). For the mean value of ETF ownership, a one standard deviation increase in absolute ETF flows is associated with turnover higher by 6.6% of a standard deviation.[26]

Columns (5) through (8) present similar regressions for the Russell 3000 sample. Here, the results are in the same direction as in the S&P 500 sample. They are weaker for volatility and stronger for turnover. For stocks at the mean level of ownership, a one standard deviation increase in absolute ETF flows translates into an increase of 1.2% of a standard deviation in intraday volatility (Column (5))[27] and of 12.3% of a standard deviation in intraday turnover (Column (7).[28] In most specifications, controlling for the lagged dependent variable and its interaction with flows reduces the main effect, but does not impact the statistical significance.

In sum, our findings support the conjecture that ETF ownership also increases volatility and turnover through the channel of share creation/redemption by market makers (APs). The economic importance of this channel seems smaller in magnitude than the effect originating from ETF mispricing arbitrage. This is consistent with the fact that share creation/redemption activity occurs at lower frequencies than intraday arbitrage. Hence, intraday volatility is less impacted.

5.2 Limits to Arbitrage

To further test whether the main effects operate through the arbitrage channel, we introduce interactions with proxies for limits to arbitrage. The prior is that arbitrage trading should be less important when limits to arbitrage are more binding. We use two proxies for limits to arbitrage: the stock-level bid-ask spread and stock lending fees.

Because ETF arbitrage involves a roundtrip transaction in the stock, a large stock-level bid-ask spread reduces the profitability of arbitrage trades and the incidence of arbitrage trading

[25] (-0.009 * 0.013 + 3.197 * 0.019 * 0.013) / 0.018 = 0.0373.
[26] (-0.090 * 0.013 + 232.101 * 0.019 * 0.013) / 0.853 = 0.0534.
[27] (0 * 0.080 + 0.141 * 0.021 * 0.080) / 0.020 = 0.0118.
[28] (-0.129 * 0.080 + 70.306 * 0.021 * 0.080) / 0.875 = 0.1227.

in a given stock. The prediction is, therefore, that the volatility and turnover of stocks with high bid-ask spreads are less sensitive to proxies of ETF arbitrage. In Table 9, Panel A, we split the sample according to the median percentage bid-ask spread in the cross-section of stocks in the prior day and re-run the analysis from Table 3. The panel shows that overall the sensitivity of both volatility and turnover to the interaction of absolute mispricing and ETF ownership is higher for stocks with a low spread. For the same level of mispricing and ETF ownership, the impact on intraday volatility and turnover is lower for high bid-ask spread stocks. The only exception to this pattern comes from the turnover of S&P 500 stocks.

We find additional evidence consistent with the idea that arbitrageurs trade on ETF mispricing when we examine the effects on intraday volatility and turnover of ETF flows (Table 9, Panel B). Similar to Panel A, we regress intraday volatility and turnover on ETF ownership interacted with absolute ETF flows, as well as main effects, controls, and fixed effects. We are interested in the way the coefficient on the interaction varies across columns. The sample is split by bid-ask spread, with odd columns containing stocks with below-median spreads and even columns containing stocks with above-median spreads. The results show that in most regression pairs, the effects are stronger for the low bid-ask spread sample than for the high bid-ask spread sample. These results are consistent with the idea that APs are reluctant to create/redeem shares when the costs of the transactions are too high.

Next, we use stock lending fees as a proxy for limits-to-arbitrage. When the lending cost is high, arbitrageurs are less likely to engage in arbitrage transactions, because the transaction costs associated with short selling shares are higher, hence reducing the profitability of trades. Also, a high lending fee can reflect a shortage in shares for lending, meaning that some arbitrageurs may simply not be able to carry out the trade (Cohen, Diether, and Malloy (2007)). The prior is that the effects of arbitrage trades on intraday volatility and turnover are expected to be stronger when lending fees are lower.

Table 9, Panel C, presents evidence of this effect. For both intraday volatility and turnover, the effect of absolute mispricing is weaker, for a given level of ETF ownership, when lending fees are higher (even-numbered columns). In other words, when stock lending fees are high, ETF ownership does not increase intraday volatility as much for a given level of mispricing. To provide evidence that APs' trades are also affected by the cost of shorting, we split the sample by lending fees and repeat the tests for fund flows. The results are presented in Table 9, Panel D. Again, we are interested in the coefficient on the interaction between ETF ownership and the absolute measure of fund flows. Consistent with the prior, the results show that in all specifications the effect is stronger for the subsample that has low lending fees (odd-numbered columns).

Overall, these results seem to corroborate the relevance of the arbitrage channel for the impact of ETF ownership on stock volatility and turnover. Whenever arbitrage is more costly, as signaled by a higher bid-ask spread or steeper stock lending fees, the impact of the arbitrage proxies is reduced.

6 Conclusion

ETFs have enjoyed rising popularity with both retail and institutional investors. This success seems warranted given that ETFs provide an unprecedented source of diversification at low cost and high liquidity. However, the evidence in this paper seems to point out an unintended effect of this relatively new asset class on the prices of their underlying securities.

We present results showing that arbitrage activity between ETFs and the underlying securities leads to an increase in stock volatility. Moreover, consistent with a deterioration of pricing efficiency, ETF ownership and flows appear to make prices diverge from random walks, both intra-day and daily. These findings lend support to the conjecture that liquidity shocks in the ETF market are propagated via arbitrage trades to the prices of underlying securities, adding a new layer of non-fundamental volatility.

While the effects that we point out are obtained in the universe of ETFs written on U.S. equity, no theoretical reason seems to prevent a generalization of these conclusions to other underlying asset classes and other types of derivatives. For this reason, our work contributes to the ongoing debate on the effect of derivatives on the quality and behavior of the prices of the underlying securities. Moreover, our evidence corroborates the results of a recent literature showing a role of index trading in generating non-fundamental volatility and comovement (Basak and Pavlova (2013a, 2013b)).

References

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

Amihud, Yakov, and Haim Mendelson, 1986, Asset pricing and the bid-ask spread, *Journal of Financial Economics* 17, 223–249.

Amihud, Yakov, and Haim Mendelson, 1987, Trading mechanisms and stock returns: An empirical investigation, *Journal of Finance* 42(3), 533–553.

Anton, Miguel, and Christopher Polk, 2014, Connected stocks, *Journal of Finance* 69(3), 1099–1128.

Atkins, Allen B., and Edward A. Dyl, 1997, Transactions Costs and Holding Periods for Common Stocks, *Journal of Finance* 52 (1), 309–325.

Barberis, Nicholas, Andrei Shleifer, and Jeffrey Wurgler, 2005, Comovement, *Journal of Financial Economics* 75(2), 283–317.

Basak, Suleyman, and Anna Pavlova, 2013a, Asset Prices and Institutional Investors, *American Economic Review* 103(5), 1728–1758.

Basak, Suleyman, and Anna Pavlova, 201b3, A model of financialization of commodities, Working paper, London Business School.

Bessembinder, Hendrik, and Paul J. Seguin, 1992, Futures-trading activity and stock price volatility, *Journal of Finance* 47 (5), 2015–2034.

Black, Fisher, 1986, Noise, *Journal of Finance* 41(3), 529–543.

Blackrock, 2011, ETF landscape, Industry review, 2011–H1.

Borkovec, Milan, Ian Domowitz, Vitaly Serbin, and Henry Yegerman, 2010, Liquidity and price discovery in exchange-traded funds, Investment Technology Group Report.

Bradley, Harold, and Robert E. Litan, 2010, Choking the recovery: Why new growth firms aren't going public and unrecognized risks of future market disruptions, Working paper, Kauffman Foundation.

Cella, Cristina, Andrew Ellul, and Mariassunta Giannetti, 2013, Investors' horizons and the amplification of market shocks, *Review of Financial Studies* 26(7), 1607–1648.

Cespa, Giovanni, and Thierry Foucault, 2012, Illiquidity shock propagation and liquidity crashes, Working paper, HEC Paris.

Chang, Eric C., Joseph W. Cheng, and J. Michael Pinegar, 1999, Does futures trading increase stock market volatility? The case of the Nikkei stock index futures markets, *Journal of Banking & Finance* 23, 727–753.

Chang, Yen-Cheng, Harrison Hong, and Inessa Liskovich, 2013, Regression Discontinuity and the Price Effects of Stock Market Indexing, NBER working paper 19290.

Cheng, Minder, and Ananth Madhavan, 2009, The dynamics of leveraged and inverse exchange-traded funds, *Journal of Investment Management* 7 (4).

Cohen, Lauren, Karl B. Diether, and Christopher J. Malloy, 2007, Supply and demand shifts in the shorting market, *Journal of Finance* 62(5), 2061–2096.

Constantinides, George M., 1986, Capital Market Equilibrium with Transaction Costs, *Journal of Political Economy* 94(4), 842–862.

Coval, Joshua, and Erik Stafford, 2007, Asset fire sales (and purchases) in equity markets, *Journal of Financial Economics* 86(2), 479–512.

Da, Zhi, and Sophie Shive, 2013, When the bellwether dances to noise: Evidence from exchange-traded funds, Working Paper, University of Notre Dame.

Danthine, Jean-Pierre, 1978, Information, Futures Prices, and Stabilizing Speculation, *Journal of Economic Theory* 17, 79–98.

Engle, Robert F., and Debojyoti Sarkar, 2006, Premiums-Discounts and exchange-traded funds, *ETFs and Indexing* 1, 35–53.

Greenwood, Robin, 2005, Short- and long-term demand curves for stocks: Theory and evidence on the dynamics of arbitrage. *Journal of Financial Economics* 75(3), 607–649.

Greenwood, Robin, and David Thesmar, 2011, Stock price fragility, *Journal of Financial Economics* 102(3), 471–490.

Gromb, Denis, and Dimitri Vayanos, 2010, Limits of arbitrage: the state of the theory, *Annual Review of Financial Economics*, 2, 251–275.

Grossman, Sanford Jay, 1989, An Analysis of the Implications for Stock and Futures Price Volatility of Program Trading and Dynamic Hedging Strategies, NBER Working paper 2357.

Jones, Charles M., and Paul J. Seguin, 1997, Transaction costs and price volatility: evidence from commission deregulation, *The American Economic Review* 87 (4), 728–737.

Jotikasthira, Chotibhak, Christian Lundblad, and Tarun Ramadorai, 2012, Asset fire sales and purchases and the international transmission of financial shocks, *Journal of Finance* 67(6), 2015–2050.

Kyle, Albert S., and Wei Xiong, 2001, Contagion as a wealth effect, *Journal of Finance* 56(4), 1401–1440.

Lee, Charles M. C., Andrei Shleifer, and Richard H. Thaler, 1991, Investor sentiment and the closed-end fund puzzle, *Journal of Finance* 46(1), 75–109.

Lee, David S., and Thomas Lemieux, 2010, Regression discontinuity designs in economics, *Journal of Economic Literature* 48(2), 281–355.

Lo, Andrew W., and Archie Craig MacKinlay,1988, Stock market prices do not follow random walks: Evidence from a simple specification test, *Review of Financial Studies* 1(1), 41–66.

Lou, Dong, 2011, A flow–based explanation of returns predictability, Working paper, London School of Economics.

Lou, Dong, and Christopher Polk, 2013, Comomentum: Inferring arbitrage activity from return correlations, Working paper, London School of Economics.

McCrary, Justin, 2008, Manipulation of the running variable in the regression discontinuity design: A density test, *Journal of Econometrics* 142(2), 698–714.

Madhavan, Ananth, 2011, Exchange-traded funds, market structure and the Flash Crash, Working paper, BlackRock Inc.

Marshall, Ben R., Nhut H. Nguyen, and Nuttawat Visaltanachoti, 2010, ETF arbitrage, Working paper, Massey University.

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

Petajisto, Antti, 2013, Inefficiencies in the pricing of exchange-traded funds, Working paper, NYU Stern School of Business.

Pontiff, Jeffrey, 1996, Costly arbitrage: Evidence from closed-end funds, *Quarterly Journal of Economics*, 1135–1151.

Ramaswamy, Srichander, 2011, Market structures and systemic risks of exchange-traded funds, Working paper, Bank of International Settlements.

Richie, Nivine, Robert Daigler, and Kimberly C. Gleason, 2008, The limits to stock index arbitrage: Examining S&P 500 futures and SPDRs, *Journal of Futures Markets* 28(12), 1182–1205.

Roll, Richard, Eduardo Schwartz, and Avanidhar Subrahmanyam, 2007, Liquidity and the law of one price: The case of the futures-cash basis, *Journal of Finance* 62 (5), 2201–2234.

Shleifer, Andrei, 1986, Do demand curves for stocks slope down? *Journal of Finance* 41, 579–590.

Stiglitz, Joseph E, 1989, Using tax policy to curb speculative short-term trading, *Journal of Financial Services Research*, 3(2–3), 101–115.

Stein, Jeremy C., 1987, Informational externalities and welfare-reducing speculation, *Journal of Political Economy* 95, 1123–1145.

Summers, Lawrence H . and Summers, Victoria P., 1989, When financial markets work too well: a cautious case for a securities transactions tax, *Journal of Financial Services Research*, 3(2–3), 261–286.

Trainor, Jr., William J., 2010, Do leveraged ETFs increase volatility? *Technology and Investment* 1(3), 215–220.

Turnovsky, Stephen J., 1983, The Determination of Spot and Futures Prices with Storable Commodities, *Econometrica* 51, 1363–1387.

Wurgler, Jeffrey, 2011, On the economic consequences of index-linked investing, in Gerald Rosenfeld, Jay W. Lorsch, and Rakesh Khurana, eds.: *Challenges to Business in the Twenty–First Century* (American Academy of Arts and Sciences, Cambridge, MA).

Appendix. Variable Description

Variable	Description	Source
<u>Daily Sample</u>		
ETF ownership	The sum of the ownership by all ETFs holding the stock, using the most recent quarterly investment company reports for equity ETFs.	Thomson-Reuters
Index (active) mutual Fund ownership	The sum of the ownership by all index (active) mutual funds holding the stock, using the most recent quarterly investment company reports.	Thomson-Reuters, CRSP Mutual Fund, and MFLinks
log(Mktcap)	The logged market capitalization of the stock (in $ millions) at the end of the day.	CRSP
1/Price	The inverse of the nominal share price at the end of the day.	CRSP
Amihud ratio	Absolute return scaled by dollar volume in $million.	CRSP
Bid-ask spread	The quoted spread divided by the bid-ask midpoint.	CRSP
Intraday volatility	Standard deviation of second-by-second intraday returns.	TAQ
Variance Ratio	The ratio of 15-second log return variance divided by 3 times the 5-second log return variance minus 1.	TAQ
Daily turnover	Total share volume scaled by period-end shares outstanding, after adjusting both volume and shares outstanding for splits and similar events.	CRSP
abs(ETF mispricing)	Stock-day level measure. Weighted average of the absolute percentage difference between the ETF Price and the NAV across the ETFs holding the stock (using the ETF price and NAV at 4:00 pm). The weight is ETF ownership of the stock.	TAQ, Bloomberg, Compustat
abs(ETF flows)	Stock-day level measure. Weighted average of the absolute percentage change in ETF shares outstanding across the ETFs holding the stock. The weight is ETF ownership of the stock.	Bloomberg, Compustat
Ret(t_1, t_2)	The total return of the stock between the close of t_1 and the close of t_2.	CRSP
Lending Fee	Loan fee aggregated at the security level, 7-day average.	Markit
<u>Monthly Sample</u>		
ETF mispricing volatility (within the month)	Standard deviation of day-end ETF mispricing (using the ETF price and NAV at 4:00 pm).	TAQ, Bloomberg, Compustat
ETF flow volatility (within the month)	Standard deviation of the relative change in daily ETF shares outstanding during the month.	Bloomberg, Compustat
log(Mktcap)	The logged market capitalization of the stock (in $ millions) at the end of the month.	CRSP
1/Price	The inverse of the nominal share price at the end of the month.	CRSP
Amihud	Absolute return scaled by dollar volume in $million, average.	CRSP
Bid-ask spread	The quoted spread divided by the bid-ask midpoint.	CRSP

Table 1. ETF Ownership Statistics

The table presents descriptive statistics for ETF ownership of stocks. For each year, across months and stocks, we average the number of ETFs, their assets under management (AUM), the weight of each stock in the ETF, and the percentage of each stocked owned by ETFs. We present statistics for S&P 500 stocks (left columns) and for Russell 3000 stocks (right columns).

	S&P 500				Russell 3000			
Year	#ETFs	Average ETF AUM ($m)	Average stock weight in ETF (%)	Average ownership of ETF in firm (%)	#ETFs	Average ETF AUM ($m)	Average stock weight in ETF (%)	Average ownership of ETF in firm (%)
2000	2.45	5627.93	0.64	0.27	2.41	5129.91	0.53	0.30
2001	13.45	2173.41	0.42	0.63	8.91	1053.93	0.16	0.37
2002	15.47	2798.87	0.45	0.88	10.18	1185.35	0.14	0.71
2003	15.95	3542.45	0.45	1.00	10.42	1465.49	0.14	0.85
2004	21.40	3451.84	0.47	1.06	14.30	1702.26	0.14	1.11
2005	24.74	3756.30	0.49	1.37	15.73	2040.02	0.16	1.37
2006	25.80	4337.34	0.51	1.68	16.81	2447.86	0.17	1.85
2007	36.04	4082.81	0.64	1.97	22.60	2438.93	0.24	2.17
2008	50.61	2980.85	0.69	2.69	30.26	1789.13	0.28	2.81
2009	53.19	2733.88	0.67	3.11	31.30	1710.54	0.26	3.41
2010	52.04	3261.34	0.68	3.16	30.08	2311.04	0.27	3.60
2011	52.77	3977.15	0.67	3.52	28.87	2937.45	0.27	3.77
2012	48.59	5026.84	0.68	3.78	26.93	3434.84	0.26	3.82
Average	30.43	3547.27	0.57	1.90	20.01	2045.99	0.21	2.10

Table 2. Summary Statistics

The table presents summary statistics for the variables used in the study. Panels A and B show summary statistics for the stock-day and for the stock-month samples, respectively. Panel C reports correlations for the daily sample, while Panel D shows summary statistics for the return regressions (returns are in percentages). All panels distinguish between the S&P 500 and the Russell 3000 samples.

Panel A: Daily Frequency Sample Statistics

S&P 500

	N	Mean	Std Dev	Min	Median	Max
Intraday volatility (%)	1,480,640	0.022	0.018	0.004	0.016	0.147
Intraday turnover (%)	1,480,640	0.970	0.853	0.031	0.700	6.230
ETF ownership	1,480,640	0.019	0.014	0.000	0.016	0.092
abs(ETF mispricing)	1,480,640	0.002	0.013	0.000	0.001	3.960
abs(ETF flows)	1,480,640	0.008	0.025	0.000	0.005	7.370
Variance Ratio	1,440,053	0.179	0.134	0.000	0.151	0.582
log(Mktcap ($m))	1,480,640	9.270	1.130	5.040	9.170	13.400
1/Price	1,480,640	0.041	0.038	0.001	0.031	0.870
Amihud	1,480,640	0.0004	0.0009	0.0000	0.0002	0.0315
Bid-ask spread	1,480,640	0.003	0.006	0.000	0.001	0.098

Russell 3000

	N	Mean	Std Dev	Min	Median	Max
Intraday volatility (%)	7,712,862	0.025	0.020	0.004	0.019	0.147
Intraday turnover (%)	7,712,862	0.874	0.875	0.029	0.596	6.230
ETF ownership	7,712,862	0.021	0.018	0.000	0.016	0.092
abs(ETF mispricing)	7,712,862	0.009	0.055	0.000	0.001	42.300
abs(ETF flows)	7,712,862	0.013	0.080	0.000	0.006	87.600
Variance Ratio	7,586,475	0.110	0.109	0.000	0.076	0.582
log(Mktcap ($m))	7,712,862	7.000	1.540	0.616	6.760	13.400
1/Price	7,712,862	0.081	0.117	0.000	0.050	40.000
Amihud	7,712,862	0.020	0.055	0.000	0.003	0.965
Bid-ask spread	7,712,862	0.004	0.006	0.000	0.002	0.379

Panel B: Monthly Frequency Sample Statistics

S&P 500

	N	Mean	Std Dev	Min	Median	Max
Daily stock volatility (%)	51,349	2.080	1.290	0.612	1.730	10.800
ETF ownership (%; average wi	51,349	2.110	1.440	0.050	1.760	9.360
ETF mispricing volatility (within	51,349	0.003	0.003	0.000	0.002	0.021
ETF flows volatility (within the	51,349	0.045	0.045	0.001	0.033	0.433

Russell 3000

	N	Mean	Std Dev	Min	Median	Max
Daily stock volatility (%)	311,079	2.610	1.490	0.612	2.240	10.800
ETF ownership (%; average wi	311,079	2.320	1.730	0.017	1.880	9.380
ETF mispricing volatility (within	311,079	0.003	0.003	0.000	0.003	0.021
ETF flows volatility (within the	311,079	0.062	0.055	0.001	0.047	0.435

Table 2. Summary Statistics (Cont.)

Panel C: Correlations for the Daily Sample

S&P 500

	Intraday volatility	Intraday turnover	ETF ownership	abs(ETF mispricing)	abs(ETF flows)	log(Mktcap)	1/Price	Amihud
Intraday volatility	1.000							
Intraday turnover	0.390	1.000						
ETF ownership	-0.011	0.375	1.000					
abs(ETF mispricing)	0.046	-0.006	-0.071	1.000				
abs(ETF flows)	0.026	0.023	-0.011	0.047	1.000			
log(Mktcap)	-0.086	-0.217	-0.067	-0.022	0.008	1.000		
1/Price	0.436	0.141	-0.030	0.013	-0.003	-0.391	1.000	
Amihud	0.175	-0.076	-0.192	0.031	0.010	-0.484	0.393	1.000
Bid-ask spread	0.213	-0.151	-0.409	0.048	0.016	-0.167	0.199	0.403

Russell 3000

	Intraday volatility	Intraday turnover	ETF ownership	abs(ETF mispricing)	abs(ETF flows)	log(Mktcap)	1/Price	Amihud
Intraday volatility	1.000							
Intraday turnover	0.271	1.000						
ETF ownership	-0.070	0.180	1.000					
abs(ETF mispricing)	-0.014	-0.037	-0.075	1.000				
abs(ETF flows)	0.015	0.004	-0.016	0.184	1.000			
log(Mktcap)	-0.273	0.119	0.006	-0.068	-0.035	1.000		
1/Price	0.440	-0.044	-0.025	0.003	0.006	-0.393	1.000	
Amihud	0.271	-0.210	-0.157	0.011	0.007	-0.436	0.419	1.000
Bid-ask spread	0.279	-0.177	-0.326	0.082	0.008	-0.256	0.312	0.478

Panel D: Summary Statistics for Return Regressions

S&P 500

	N	Mean	Std Dev	Min	Median	Max
Ret(t)	1,415,085	0.056	2.115	-9.459	0.019	10.403
Ret(t+1,t+5)	1,415,085	0.206	4.518	-19.930	0.221	21.345
Ret(t+1,t+10)	1,415,085	0.384	6.130	-23.891	0.460	25.242
Ret(t+1,t+20)	1,415,085	0.742	8.618	-31.460	0.955	33.696
net(ETF Flows)	1,415,085	0.001	0.010	-0.060	0.000	0.063

Russell 3000

	N	Mean	Std Dev	Min	Median	Max
Ret(t)	6,954,903	0.060	2.478	-9.459	0.000	10.405
Ret(t+1,t+5)	6,954,903	0.178	5.240	-19.931	0.151	21.348
Ret(t+1,t+10)	6,954,903	0.351	7.052	-23.895	0.357	25.259
Ret(t+1,t+20)	6,954,903	0.688	9.958	-31.461	0.767	33.698
net(ETF Flows)	6,954,903	0.001	0.014	-0.060	0.000	0.063

Table 3. ETF Ownership, Intraday Stock Volatility, and Turnover (Daily Sample)

The table reports estimates from OLS regressions of intraday volatility and daily turnover on ETF ownership and controls. In Columns (1) to (4), the sample consists of S&P 500 stocks, and in Columns (5) to (8) the sample consists of Russell 3000 stocks. The frequency of the observations is daily. Intraday stock volatility is computed using second-by-second data from the TAQ database, and daily turnover is computed as daily volume from CRSP divided by shares outstanding. Variable descriptions are provided in the Appendix. Standard errors are clustered at the stock level. *t*-statistics are presented in parentheses. ***, **, and * represent statistical significance at the 1%, 5%, and 10% levels, respectively.

Sample:	S&P 500				Russell 3000			
Dependent variable:	Intraday volatility		Intraday turnover		Intraday volatility		Intraday turnover	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ETF ownership (t-1)	0.333***	0.243***	18.869***	11.631***	-0.009	0.069***	7.624***	4.026***
	(9.613)	(7.461)	(7.976)	(8.773)	(-1.360)	(8.883)	(14.875)	(10.027)
log(Mktcap (t-1))	0.003***	0.004***	-0.171***	-0.194***	-0.001***	-0.003***	0.034***	0.077***
	(8.781)	(5.356)	(-10.524)	(-5.552)	(-12.372)	(-10.781)	(6.106)	(9.068)
1/Price (t-1)	0.219***	0.195***	2.826***	1.202**	0.059***	0.032***	0.534***	-0.044
	(20.998)	(12.929)	(6.106)	(2.263)	(26.912)	(12.631)	(12.861)	(-1.048)
Amihud (t-1)	-0.243	-0.333	-158.086**	123.183***	0.015***	0.020***	-2.551***	-1.141***
	(-0.554)	(-1.038)	(-7.861)	(-7.548)	(6.206)	(8.656)	(-26.777)	(-15.669)
Bid-ask spread (t-1)	-0.124	-0.119*	-9.143***	-7.636***	-0.033	-0.006	-12.764***	-10.096***
	(-1.496)	(-1.872)	(-4.773)	(-5.516)	(-1.211)	(-0.264)	(-12.396)	(-13.161)
Stock fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
Day fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,472,346	1,472,346	1,472,346	1,472,346	7,687,652	7,687,652	7,687,652	7,687,652
Adjusted R^2	0.425	0.466	0.282	0.464	0.367	0.451	0.123	0.381

Table 4. ETF Ownership and Daily Stock Volatility (Monthly Sample)

The table reports estimates from OLS regressions of daily volatility on ETF ownership, ownership by mutual funds, and controls. In Columns (1) to (3), the sample consists of S&P 500 stocks, and in Columns (4) to (6), the sample consists of Russell 3000 stocks. The frequency of the observations is monthly. Daily stock volatility is computed using daily returns within a month. Variable descriptions are provided in the Appendix. Standard errors are clustered at the date and stock levels. *t*-statistics are presented in parentheses. ***, **, and * represent statistical significance at the 1%, 5%, and 10% levels, respectively.

Dependent variable:	Daily stock volatility (computed within the month)					
Sample:	S&P 500			Russell 3000		
	(1)	(2)	(3)	(4)	(5)	(6)
ETF ownership (average within the month)	0.146***	0.142***	0.137***	0.044***	0.036***	0.033***
	(8.039)	(7.939)	(7.675)	(7.542)	(6.190)	(5.810)
log(Mktcap (t-1))	-0.164***	-0.177***	-0.181***	-0.244***	-0.270***	-0.266***
	(-2.988)	(-3.282)	(-3.381)	(-11.448)	(-12.480)	(-12.294)
1/Price (t-1)	6.464***	6.345***	6.286***	2.866***	2.917***	2.916***
	(7.153)	(7.141)	(7.083)	(11.872)	(12.125)	(12.149)
Amihud (t-1)	85.958***	94.728***	93.795***	0.879***	1.012***	0.979***
	(4.170)	(4.521)	(4.539)	(2.885)	(3.322)	(3.220)
Bid-ask spread (t-1)	28.025***	28.522***	29.511***	4.092	6.442*	6.599*
	(2.904)	(2.965)	(3.098)	(1.154)	(1.833)	(1.883)
Mutual fund ownership		0.010***			0.009***	
		(3.557)			(8.179)	
Index mutual fund ownership			0.027***			0.021***
			(3.063)			(6.018)
Active mutual fund ownership			0.008***			0.008***
			(2.699)			(6.262)
Stock fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	49,947	49,947	49,947	292,699	292,699	292,699
Adjusted R^2	0.630	0.631	0.631	0.560	0.561	0.561

Table 5. ETF Ownership and Price Efficiency: Variance Ratios

The table reports estimates from OLS regressions of Variance Ratios on ETF ownership and controls. In Columns (1) and (2), the sample consists of S&P 500 stocks, and in Columns (3) and (4) the sample consists of Russell 3000 stocks. The frequency of the observations is daily. The variance ratio is computed as the absolute value of the ratio of the variance of fifteen-second log returns on day *t* and 3 times the variance of five-second log returns on day *t*, minus 1 using data from the TAQ database. Variable descriptions are provided in the Appendix. Standard errors are clustered at the stock level. *t*-statistics are presented in parentheses. ***, **, and * represent statistical significance at the 1%, 5%, and 10% levels, respectively.

Dependent variable:	Variance Ratio			
Sample:	S&P 500		Russell 3000	
	(1)	(2)	(3)	(4)
ETF ownership (average within the month)	1.620***	1.123***	0.122***	0.277***
	(8.578)	(6.888)	(2.938)	(8.824)
log(Mktcap (t-1))	0.049***	0.016***	0.026***	0.012***
	(20.862)	(3.435)	(33.885)	(13.377)
1/Price (t-1)	1.234***	0.531***	0.064***	0.016***
	(16.271)	(6.538)	(4.767)	(3.291)
Amihud (t-1)	-10.250***	-11.381***	-0.007	-0.009
	(-4.320)	(-6.969)	(-0.699)	(-1.622)
Bid-ask spread (t-1)	-1.466***	-1.554***	-0.742***	-0.856***
	(-4.063)	(-5.951)	(-5.855)	(-10.776)
Stock fixed effects	Yes	Yes	Yes	Yes
Day fixed effects	No	Yes	No	Yes
Observations	1,422,282	1,422,282	7,657,065	7,657,065
Adjusted R^2	0.245	0.352	0.149	0.284

Table 6. Price Reversals

The table reports estimates from OLS regressions of one- and multi-day returns on ETF flows and controls. In Columns (1) to (4), the sample consists of S&P 500 stocks, and in Columns (5) to (8), the sample consists of Russell 3000 stocks. The frequency of the observations is daily. Returns are in percent. Variable descriptions are provided in the Appendix. Standard errors are clustered at the stock level. *t*-statistics are presented in parentheses. ***, **, and * represent statistical significance at the 1%, 5%, and 10% levels, respectively.

Sample:	S&P 500				Russell 3000			
Dependent variable:	Ret(t)	Ret(t+1,t+5)	Ret(t+1,t+10)	Ret(t+1,t+20)	Ret(t)	Ret(t+1,t+5)	Ret(t+1,t+10)	Ret(t+1,t+20)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ETF flows (t)	12.418***	-2.749***	-4.464***	-5.643***	5.138***	-1.224***	-0.231	-1.882***
	(19.318)	(-4.681)	(-5.509)	(-5.291)	(41.902)	(-6.697)	(-0.945)	(-5.846)
log(Mktcap (t-1))	0.013***	-0.034***	-0.067***	-0.121***	0.013***	0.004	0.012**	0.025**
	(6.179)	(-4.489)	(-4.344)	(-3.978)	(12.430)	(1.569)	(2.131)	(2.328)
1/Price (t-1)	-0.895***	1.039***	2.070***	5.625***	-0.546***	-0.403***	-0.686***	-0.627***
	(-9.405)	(2.840)	(2.802)	(3.954)	(-19.299)	(-6.926)	(-6.296)	(-2.937)
Amihud (t-1)	28.849***	-19.909	-39.320	-37.478	0.096***	-1.350***	-2.539***	-4.308***
	(7.452)	(-1.630)	(-1.553)	(-0.825)	(2.679)	(-12.341)	(-12.227)	(-10.718)
Bid-ask spread (t-1)	2.106***	2.357	5.982	12.265	2.455***	-1.359	-0.520	0.006
	(3.577)	(1.039)	(1.338)	(1.450)	(6.926)	(-1.185)	(-0.246)	(0.001)
Day fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,427,667	1,427,667	1,427,667	1,427,667	7,084,196	7,084,196	7,084,196	7,084,196
Adjusted R^2	0.326	0.300	0.279	0.279	0.279	0.249	0.223	0.221

Table 7. Regression Discontinuity Design around the Russell 1000/Russell 2000 cutoff

The table reports estimates from a design exploiting the discontinuity in ETF ownership around the cutoff between the Russell 1000 and Russell 2000 indexes. The frequency of the data is monthly. Panel A has the regressions of ETF ownership on a dummy for inclusion in the Russell 2000 for stocks in the Russell 1000 before index reconstitution (Columns (1), (3), (5) and (7)) and regressions of ETF ownership on a dummy for inclusion in the Russell 1000 for stocks in the Russell 2000 before index reconstitution (Columns (2), (4), (6) and (8)). Stocks are ranked in terms of market capitalization and different ranges of this rank around the cutoff are used for inclusion in the sample: 50 stocks on each side (Columns (1) and (2)), 100 stocks on each side (Columns (3) and (4)), 150 stocks on each side (Columns (5) and (6)), and 200 stocks on each side (Columns (7) and (8)). The same stocks enter the sample from June after index reconstitution to May of the next year, except if delistings occur. Panel B regresses daily stock volatility (computed within a month) on instrumented ETF ownership. The instruments are either a dummy for the inclusion in the Russell 2000 for stocks in the Russell 1000 before reconstitution (Columns (1), (3), (5) and (7)) or a dummy for the inclusion in the Russell 1000 for stocks in the Russell 2000 before reconstitution (Columns (2), (4), (6) and (8)). The same bandwidths around the cutoff are considered to restrict the sample as in Panel A. The regressions in this panel, as well as in Panel A, include a linear specification of the ranking variable (not reported). Panel C replicates the analysis in Panel B including instead a quadratic specification of the ranking variable (not reported). The first stage is modified accordingly. Variable descriptions are provided in the Appendix. Standard errors are clustered at the date level. *t*-statistics are presented in parentheses. ***, **, and * represent statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: First Stage Regressions

Dependent variable:	ETF Ownership							
Sample:	± 50 stocks around cuttoff		± 100 stocks around cuttoff		± 150 stocks around cuttoff		± 200 stocks around cuttoff	
	R1000	R2000	R1000	R2000	R1000	R2000	R1000	R2000
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
In Russell 2000	0.294***		0.508***		0.493***		0.475***	
	(4.466)		(9.553)		(10.558)		(9.898)	
In Russell 1000		-0.638***		-0.402***		-0.488***		-0.457***
		(-8.585)		(-7.660)		(-9.268)		(-9.488)
log(Mktcap (t-1))	-0.748***	-0.526***	-0.666***	-0.373***	-0.617***	-0.276***	-0.525***	-0.282***
	(-7.047)	(-7.922)	(-8.702)	(-7.667)	(-8.985)	(-5.689)	(-9.061)	(-6.472)
1/Price (t-1)	-1.941***	-1.744***	-1.749***	-2.208***	-1.962***	-1.976***	-1.589***	-1.854***
	(-8.005)	(-4.890)	(-6.286)	(-6.948)	(-7.516)	(-7.350)	(-6.108)	(-8.321)
Amihud (t-1)	-0.764	-27.490***	-3.725**	-3.215***	-4.982***	-4.527***	-8.048***	-6.204***
	(-0.791)	(-8.000)	(-2.377)	(-4.001)	(-2.846)	(-3.949)	(-3.581)	(-4.190)
Bid-ask spread (t-1)	-13.560***	10.834***	-10.248***	2.723	-10.777***	-3.933	-9.636***	-4.581*
	(-3.480)	(4.733)	(-3.167)	(1.210)	(-3.419)	(-1.525)	(-3.514)	(-1.708)
Month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,745	4,854	7,328	9,947	11,191	15,090	15,347	19,807
Adjusted R^2	0.457	0.673	0.467	0.628	0.481	0.608	0.481	0.594

Table 7. Regression Discontinuity Design around the Russell 1000/Russell 2000 cutoff (Cont.)

Panel B: Second Stage Regressions, First Degree Polynomial

Polynomial:	Linear specification							
Dependent variable:	Daily stock volatility (computed within the month)							
Sample:	± 50 stocks around cuttoff		± 100 stocks around cuttoff		± 150 stocks around cuttoff		± 200 stocks around cuttoff	
Instrument:	In R2000	In R1000	In R2000	In R1000	In R2000	In R1000	In R2000	In R1000
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ETF ownership (instrumented)	1.993	0.218***	0.647***	0.355***	0.824***	0.215**	0.726***	0.192***
	(1.636)	(2.680)	(2.925)	(2.698)	(4.083)	(2.604)	(4.355)	(2.945)
log(Mktcap (t-1))	-0.016	0.417	-0.844***	0.157	-0.823***	-1.096	-0.910***	-0.879
	(-0.019)	(0.099)	(-4.628)	(0.142)	(-4.882)	(-1.074)	(-6.270)	(-0.914)
1/Price (t-1)	7.321***	-24.091***	5.177***	-23.376***	6.318***	-22.458***	5.542***	-25.322***
	(3.125)	(-6.760)	(8.478)	(-6.311)	(10.470)	(-4.695)	(12.552)	(-5.754)
Amihud (t-1)	-4.387**	-0.003**	-3.105*	-0.002***	-3.413**	-0.002***	-1.248	-0.002***
	(-1.987)	(-2.529)	(-1.918)	(-3.088)	(-2.096)	(-4.491)	(-0.706)	(-6.463)
Bid-ask spread (t-1)	16.191	0.000	-1.999	-0.001**	2.980	-0.000	3.638	0.000
	(0.796)	(0.074)	(-0.254)	(-2.059)	(0.447)	(-0.340)	(0.718)	(0.503)
Month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,745	4,854	7,328	9,947	11,191	15,090	15,347	19,807

Table 7. Regression Discontinuity Design around the Russell 1000/Russell 2000 cutoff (Cont.)

Panel C: Second Stage Regressions, Second Degree Polynomial

Polynomial:	Quadratic specification							
Dependent variable:	Daily stock volatility (computed within the month)							
Sample:	± 50 stocks around cuttoff		± 100 stocks around cuttoff		± 150 stocks around cuttoff		± 200 stocks around cuttoff	
Instrument:	In R2000	In R1000	In R2000	In R1000	In R2000	In R1000	In R2000	In R1000
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ETF ownership	1.521	0.308***	0.711	0.248**	0.644***	0.297***	0.688***	0.220***
	(1.032)	(3.134)	(1.089)	(2.577)	(3.258)	(3.577)	(3.858)	(2.874)
log(Mktcap (t-1))	-0.315	-1.184***	-0.801*	-1.378***	-0.940***	-1.242***	-0.937***	-1.372***
	(-0.299)	(-7.975)	(-1.921)	(-11.452)	(-5.431)	(-10.485)	(-5.998)	(-12.269)
1/Price (t-1)	6.500**	4.632***	5.295***	4.049***	5.890***	4.901***	5.427***	4.412***
	(2.148)	(7.803)	(4.191)	(8.160)	(9.790)	(10.811)	(12.100)	(11.173)
Amihud (t-1)	-4.532**	3.108	-2.944	-0.142	-4.163***	-0.642	-1.459	-0.660
	(-2.273)	(0.724)	(-1.280)	(-0.135)	(-2.701)	(-0.617)	(-0.793)	(-0.649)
Bid-ask spread (t-1)	16.171	-24.640***	-1.546	-23.222***	1.363	-21.857***	4.224	-25.181***
	(0.611)	(-6.821)	(-0.155)	(-6.383)	(0.213)	(-4.464)	(0.821)	(-5.737)
Month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,745	4,854	7,328	9,947	11,191	15,090	15,347	19,807

Table 8. Stock Volatility, ETF Ownership, and Arbitrage

The table reports estimates from OLS regressions of intraday volatility and daily turnover on ETF ownership, variables that proxy for ETF arbitrage, and controls. In Columns (1) to (4), the sample consists of S&P 500 stocks, and in Columns (5) to (8), the sample consists of Russell 3000 stocks. The frequency of the observations is daily. Intraday stock volatility is computed using second-by-second data from the TAQ database, and daily turnover is computed as daily volume from CRSP divided by shares outstanding. In Panel A, the variable of interest is the interaction of lagged absolute ETF mispricing and ETF ownership. In Panel B, the variable of interest is the interaction of lagged absolute ETF fund flows and ETF ownership. Variable descriptions are provided in the Appendix. Standard errors are clustered at the stock level. *t*-statistics are presented in parentheses. ***, **, and * represent statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Effect of ETF Mispricing on Volatility and Turnover

Sample:	S&P 500				Russell 3000			
Dependent variable:	Intraday volatility		Intraday turnover		Intraday volatility		Intraday turnover	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ETF ownership (t-1)	0.186***	0.090***	10.371***	2.811***	0.068***	0.052***	4.005***	0.334**
	(5.814)	(5.691)	(8.038)	(4.876)	(8.633)	(10.844)	(9.949)	(2.263)
× abs(ETF mispricing (t-1))	42.035***	18.879***	896.893***	321.296***	-0.113	-0.085	-2.660	-1.574
	(9.876)	(8.530)	(6.860)	(6.300)	(-0.417)	(-0.678)	(-0.350)	(-0.543)
abs(ETF mispricing (t-1))	0.006***	0.000	0.207**	0.076	-0.005	-0.002	-0.085	-0.031
	(2.749)	(0.139)	(2.459)	(1.533)	(-0.943)	(-0.895)	(-0.811)	(-0.727)
log(Mktcap (t-1))	0.004***	0.002***	-0.198***	-0.073***	-0.003***	-0.001***	0.071***	0.027***
	(5.351)	(5.412)	(-5.658)	(-5.586)	(-11.660)	(-12.542)	(8.253)	(7.951)
1/Price (t-1)	0.193***	0.090***	1.145**	0.416**	0.032***	0.018***	-0.062	-0.054***
	(12.832)	(10.536)	(2.148)	(2.053)	(12.693)	(19.253)	(-1.454)	(-2.679)
Amihud (t-1)	-0.306	-0.145	-122.456***	-47.584***	0.020***	0.009***	-1.153***	-0.497***
	(-0.960)	(-0.952)	(-7.536)	(-7.654)	(8.404)	(8.336)	(-15.860)	(-16.466)
Bid-ask spread (t-1)	-0.096	-0.033	-7.187***	-2.762***	0.004	-0.007	-9.967***	-4.319***
	(-1.595)	(-1.195)	(-5.328)	(-5.481)	(0.187)	(-0.680)	(-13.096)	(-13.772)
Dependent variable (t-1)		0.531***		0.578***		0.519***		0.558***
		(41.455)		(69.477)		(116.564)		(170.937)
× ETF ownership (t-1)		-0.123		0.889***		-0.734***		1.132***
		(-0.216)		(3.397)		(-4.635)		(11.621)
Stock fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,471,139	1,455,418	1,471,139	1,455,418	7,679,072	7,530,046	7,679,072	7,530,046
Adjusted R^2	0.470	0.614	0.465	0.659	0.452	0.591	0.381	0.597

Table 8. Stock Volatility, ETF Ownership, and Arbitrage (Cont.)

Panel B: Effects of Fund Flows on Volatility and Turnover

Sample:	S&P 500				Russell 3000			
Dependent variable:	Intraday volatility		Intraday turnover		Intraday volatility		Intraday turnover	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ETF ownership (t-1)	0.229***	0.102***	10.305***	2.908***	0.068***	0.052***	3.328***	0.146
	(7.003)	(6.392)	(7.996)	(4.875)	(8.846)	(11.013)	(8.269)	(0.991)
× abs(ETF flows (t-1))	3.197***	1.321***	232.101***	75.085***	0.141*	0.066	70.306***	23.407***
	(5.861)	(5.371)	(5.988)	(5.823)	(1.688)	(1.609)	(8.298)	(7.954)
abs(ETF flows (t-1))	-0.009***	-0.003***	-0.090	-0.041	-0.000*	-0.000*	-0.129***	-0.043***
	(-4.521)	(-3.428)	(-1.491)	(-1.555)	(-1.893)	(-1.669)	(-3.466)	(-3.281)
log(Mktcap (t-1))	0.004***	0.002***	-0.198***	-0.073***	-0.003***	-0.001***	0.073***	0.027***
	(5.240)	(5.336)	(-5.709)	(-5.634)	(-11.581)	(-12.448)	(8.520)	(8.178)
1/Price (t-1)	0.194***	0.089***	1.120**	0.410**	0.032***	0.018***	-0.063	-0.054***
	(12.769)	(10.525)	(2.130)	(2.044)	(12.692)	(19.244)	(-1.490)	(-2.717)
Amihud (t-1)	-0.302	-0.138	-121.598***	-47.386***	0.020***	0.009***	-1.137***	-0.491***
	(-0.951)	(-0.908)	(-7.525)	(-7.633)	(8.458)	(8.386)	(-15.699)	(-16.310)
Bid-ask spread (t-1)	-0.112*	-0.037	-7.565***	-2.890***	0.003	-0.008	-9.946***	-4.311***
	(-1.792)	(-1.322)	(-5.532)	(-5.644)	(0.119)	(-0.756)	(-13.088)	(-13.757)
Dependent variable (t-1)		0.529***		0.579***		0.519***		0.558***
		(41.375)		(69.764)		(116.566)		(171.125)
× ETF ownership (t-1)		0.171		0.836***		-0.742***		1.101***
		(0.313)		(3.264)		(-4.678)		(11.279)
Stock fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,471,139	1,455,418	1,471,139	1,455,418	7,679,072	7,530,046	7,679,072	7,530,046
Adjusted R^2	0.467	0.613	0.466	0.659	0.452	0.591	0.381	0.597

Table 9. Evidence from Limits-to-Arbitrage

The table reports estimates from OLS regressions of intraday volatility and daily turnover on ETF ownership, variables that proxy for ETF arbitrage, and controls. In Columns (1) to (4), the sample consists of S&P 500 stocks, and in Columns (5) to (8), the sample consists of Russell 3000 stocks. The frequency of the observations is daily. Intraday stock volatility is computed using second-by-second data from the TAQ database, and daily turnover is computed as daily volume from CRSP divided by shares outstanding. In Panels A and C, the variable of interest is the interaction of lagged absolute ETF mispricing and ETF ownership. In Panels B and D, the variable of interest is the interaction of lagged absolute ETF fund flows and ETF ownership. The sample is split by the lagged bid-ask spread (Panels A and B) or the lagged stock lending fee (Panels C and D). Variable descriptions are provided in the Appendix. Standard errors are clustered at the stock level. *t*-statistics are presented in parentheses. ***, **, and * represent statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Mispricing and Bid-Ask Spread

Sample:	S&P 500				Russell 3000			
Dependent variable:	Intraday volatility		Intraday turnover		Intraday volatility		Intraday turnover	
Bid-ask spread (t-1):	Low	High	Low	High	Low	High	Low	High
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ETF ownership (t-1)	0.142***	0.168***	10.017***	8.286***	0.094***	0.066***	3.564***	4.195***
	(4.833)	(4.169)	(7.050)	(5.462)	(8.944)	(7.257)	(6.480)	(9.764)
× abs(ETF mispricing (t-1))	50.828***	17.244***	750.869***	764.789***	0.736***	-0.197	21.775**	-11.773*
	(12.241)	(5.869)	(5.204)	(5.591)	(3.767)	(-0.955)	(2.227)	(-1.956)
abs(ETF mispricing (t-1))	0.003	0.001	0.204**	-0.266	-0.016***	-0.003	-0.429***	-0.007
	(1.388)	(0.189)	(2.149)	(-1.487)	(-4.880)	(-0.753)	(-3.018)	(-0.118)
log(Mktcap (t-1))	0.005***	0.002**	-0.186***	-0.295***	0.000*	-0.005***	-0.037***	0.083***
	(4.044)	(2.541)	(-5.012)	(-7.028)	(1.780)	(-15.029)	(-2.982)	(8.354)
1/Price (t-1)	0.082***	0.190***	-0.985	0.363	0.062***	0.026***	-1.590***	0.028
	(3.555)	(12.505)	(-1.327)	(0.679)	(10.606)	(10.371)	(-6.486)	(0.804)
Amihud (t-1)	-0.467	-0.222	-213.891***	-98.507***	0.048***	0.014***	-3.218***	-0.937***
	(-0.866)	(-0.524)	(-7.192)	(-6.234)	(7.150)	(6.353)	(-10.700)	(-16.194)
Bid-ask spread (t-1)	-0.641***	0.119**	-14.885***	-3.711**	-0.685***	0.081***	-10.460***	-6.150***
	(-5.013)	(2.500)	(-5.906)	(-2.471)	(-8.990)	(3.718)	(-4.387)	(-9.479)
Stock fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	735,570	735,569	735,570	735,569	3,839,536	3,839,536	3,839,536	3,839,536
Adjusted R^2	0.488	0.522	0.544	0.436	0.407	0.474	0.401	0.362

Table 9. Evidence from Limits to Arbitrage (Cont.)

Panel B: Fund Flows and Bid-Ask Spread

Sample:	S&P 500				Russell 3000			
Dependent variable:	Intraday volatility		Intraday turnover		Intraday volatility		Intraday turnover	
Bid-ask spread (t-1):	Low	High	Low	High	Low	High	Low	High
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ETF ownership (t-1)	0.205***	0.169***	9.918***	7.760***	0.099***	0.064***	3.023***	3.562***
	(6.813)	(4.262)	(6.834)	(5.344)	(9.495)	(7.130)	(5.581)	(8.188)
× abs(ETF flows (t))	4.231***	2.648***	275.046***	197.370***	-0.096	0.239**	74.942***	55.122***
	(7.632)	(4.580)	(9.243)	(4.609)	(-0.979)	(2.456)	(12.255)	(6.068)
abs(ETF flows (t))	-0.010***	-0.004**	-0.095	-0.037	0.000*	-0.001***	0.018	-0.133***
	(-4.051)	(-2.004)	(-1.606)	(-0.404)	(1.691)	(-2.949)	(1.536)	(-9.671)
log(Mktcap (t-1))	0.005***	0.002**	-0.186***	-0.294***	0.001*	-0.005***	-0.034***	0.084***
	(3.997)	(2.547)	(-5.094)	(-7.066)	(1.898)	(-14.978)	(-2.725)	(8.478)
1/Price (t-1)	0.086***	0.190***	-0.949	0.338	0.062***	0.026***	-1.590***	0.026
	(3.633)	(12.430)	(-1.284)	(0.637)	(10.610)	(10.370)	(-6.478)	(0.764)
Amihud (t-1)	-0.484	-0.204	-213.397***	-97.238***	0.047***	0.014***	-3.147***	-0.928***
	(-0.868)	(-0.481)	(-7.212)	(-6.185)	(7.068)	(6.410)	(-10.486)	(-16.103)
Bid-ask spread (t-1)	-0.683***	0.116**	-15.357***	-3.831**	-0.695***	0.080***	-10.444***	-6.132***
	(-5.086)	(2.440)	(-5.970)	(-2.569)	(-9.083)	(3.689)	(-4.357)	(-9.464)
Stock fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	735,568	735,571	735,568	735,571	3,839,536	3,839,536	3,839,536	3,839,536
Adjusted R^2	0.482	0.522	0.545	0.438	0.407	0.474	0.401	0.362

Table 9. Evidence from Limits to Arbitrage (Cont.)

Panel C: Mispricing and Lending Fees

Sample:	S&P 500				Russell 3000			
Dependent variable:	Intraday volatility		Intraday turnover		Intraday volatility		Intraday turnover	
Rebate rate:	Low	High	Low	High	Low	High	Low	High
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ETF ownership (t-1)	0.085***	0.026*	6.587***	5.601***	0.037***	0.044***	2.813***	1.969***
	(5.327)	(1.854)	(5.845)	(5.076)	(8.026)	(7.162)	(6.827)	(4.164)
× abs(ETF mispricing (t-1))	21.480***	18.856***	1,467.626***	783.211***	2.221***	-0.536***	324.516***	-2.942
	(5.072)	(4.503)	(5.320)	(3.807)	(2.606)	(-2.767)	(3.950)	(-0.557)
abs(ETF mispricing (t-1))	-0.157*	-0.224***	-16.278***	-8.800**	-0.035*	0.002***	-7.344***	0.019
	(-1.657)	(-2.952)	(-2.904)	(-2.084)	(-1.772)	(3.862)	(-3.835)	(1.079)
log(Mktcap (t-1))	0.000	0.001	-0.464***	-0.566***	-0.004***	-0.006***	-0.007	0.033*
	(0.525)	(1.214)	(-11.998)	(-10.087)	(-14.197)	(-16.838)	(-0.391)	(1.772)
1/Price (t-1)	0.187***	0.204***	0.038	1.338	0.035***	0.015***	-0.416***	-0.064
	(13.798)	(15.204)	(0.058)	(1.254)	(12.870)	(5.077)	(-4.354)	(-1.058)
Amihud (t-1)	-0.491	-0.662	-273.758***	-407.072***	-0.010***	-0.018***	-1.437***	-1.435***
	(-0.693)	(-0.848)	(-5.798)	(-6.412)	(-3.445)	(-5.086)	(-10.874)	(-10.162)
Bid-ask spread (t-1)	1.777***	2.183***	47.114***	44.975***	1.026***	1.372***	-13.974***	-15.849***
	(3.117)	(4.903)	(4.595)	(3.088)	(11.298)	(13.376)	(-6.540)	(-5.944)
Stock fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	366,618	366,618	366,618	366,618	2,088,566	2,088,563	2,088,566	2,088,563
Adjusted R^2	0.518	0.582	0.504	0.524	0.477	0.520	0.458	0.428

Table 9. Evidence from Limits to Arbitrage (Cont.)

Panel D: Fund Flows and Lending Fees

Sample:	S&P 500				Russell 3000			
Dependent variable:	Intraday volatility		Intraday turnover		Intraday volatility		Intraday turnover	
Rebate rate:	Low	High	Low	High	Low	High	Low	High
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ETF ownership (t-1)	0.107***	0.047***	7.899***	6.312***	0.034***	0.040***	2.321***	1.200**
	(6.442)	(3.128)	(7.013)	(5.739)	(7.370)	(6.541)	(6.005)	(2.510)
× abs(ETF flows (t))	0.953	0.263	98.639**	48.965	0.684***	0.375***	100.294***	83.037***
	(1.639)	(0.753)	(2.485)	(1.234)	(7.212)	(4.292)	(12.066)	(6.767)
abs(ETF flows (t))	0.039**	0.046***	1.079	2.848***	-0.002	-0.001***	-0.856***	-0.250***
	(2.560)	(4.979)	(0.977)	(2.781)	(-0.612)	(-3.837)	(-3.278)	(-7.415)
log(Mktcap (t-1))	0.000	0.001	-0.467***	-0.564***	-0.004***	-0.006***	-0.005	0.036*
	(0.404)	(1.257)	(-12.144)	(-10.088)	(-14.128)	(-16.803)	(-0.305)	(1.948)
1/Price (t-1)	0.187***	0.204***	0.015	1.323	0.035***	0.015***	-0.411***	-0.064
	(13.694)	(15.195)	(0.023)	(1.248)	(12.857)	(5.076)	(-4.336)	(-1.059)
Amihud (t-1)	-0.474	-0.627	-272.455***	-404.583***	-0.010***	-0.018***	-1.430***	-1.433***
	(-0.665)	(-0.795)	(-5.891)	(-6.440)	(-3.409)	(-5.074)	(-10.902)	(-10.234)
Bid-ask spread (t-1)	1.764***	2.154***	46.079***	42.960***	1.026***	1.374***	-14.022***	-15.321***
	(3.068)	(4.840)	(4.403)	(2.967)	(11.271)	(13.377)	(-6.622)	(-5.800)
Stock fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	366,618	366,618	366,618	366,618	2,088,566	2,088,563	2,088,566	2,088,563
Adjusted R^2	0.518	0.582	0.503	0.524	0.477	0.520	0.459	0.429

Appendix Table A1. Regression in First Differences

The table reports estimates from OLS regressions of changes in volatility on changes in ETF ownership. In Column (1), the sample consists of S&P 500 stocks, and in Column (2), the sample consists of Russell 3000 stocks. The frequency of the observations is monthly. Daily stock volatility is computed using daily returns within a month. The changes are computed month on month. Variable descriptions are provided in the Appendix. Standard errors are clustered at the date and stock levels. *t*-statistics are presented in parentheses. ***, **, and * represent statistical significance at the 1%, 5%, and 10% levels, respectively.

Dependent variable:	Change in volatility	
Sample:	S&P 500	Russell 3000
	(1)	(2)
Change in ETF ownership	0.055**	0.060***
	(2.227)	(6.509)
log(Mktcap (t-1))	-0.022***	-0.013***
	(-6.629)	(-15.439)
1/Price (t-1)	-0.147	-0.014
	(-1.345)	(-0.525)
Amihud (t-1)	-37.318***	-0.656***
	(-3.369)	(-9.297)
Bid-ask spread (t-1)	-9.046***	-4.648***
	(-3.823)	(-4.987)
Month fixed effects	Yes	Yes
Observations	48,466	293,581
Adjusted R^2	0.278	0.173

Appendix Table A2. Regression Discontinuity Design with Mutual Fund Controls (Second Stage)

The table reports estimates from a design exploiting the discontinuity in ETF ownership around the cutoff between the Russell 1000 and Russell 2000 indexes, controlling for ownership by active and index funds. The frequency of the data is monthly. The table has the second stage regressions of stock volatility on ETF ownership. ETF ownership is instrumented using a dummy for inclusion in the Russell 2000 for stocks in the Russell 1000 before index reconstitution (Columns (1), (3), (5) and (7)) and regressions of ETF ownership on a dummy for inclusion in the Russell 1000 for stocks in the Russell 2000 before index reconstitution (Columns (2), (4), (6) and (8)). Stocks are ranked in terms of market capitalization and different ranges of this rank are used for inclusion in the sample: 50 stocks on each side (Columns (1) and (2)), 100 stocks on each side (Columns (3) and (4)), 150 stocks on each side (Columns (5) and (6)), and 200 stocks on each side (Columns (7) and (8)). The same stocks enter the sample from June after index reconstitution to May of the next year, except if delistings occur. The regressions include a linear specification of the ranking variable (not reported). Variable descriptions are provided in the Appendix. Standard errors are clustered at the date level. *t*-statistics are presented in parentheses. ***, **, and * represent statistical significance at the 1%, 5%, and 10% levels, respectively.

Polynomial:	Linear specification							
Dependent variable:	Daily stock volatility (computed within the month)							
Sample:	± 50 stocks around cuttoff		± 100 stocks around cuttoff		± 150 stocks around cuttoff		± 200 stocks around cuttoff	
Instrument:	In R2000	In R1000	In R2000	In R1000	In R2000	In R1000	In R2000	In R1000
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ETF ownership (instrumented)	2.173	0.307***	0.676***	0.486***	0.868***	0.307***	0.776***	0.255***
	(1.459)	(3.489)	(3.026)	(3.657)	(3.925)	(3.935)	(4.063)	(3.852)
log(Mktcap (t-1))	-0.008	-1.166***	-0.843***	-1.332***	-0.829***	-1.266***	-0.917***	-1.378***
	(-0.009)	(-8.582)	(-4.517)	(-10.685)	(-4.737)	(-10.619)	(-5.992)	(-12.260)
1/Price (t-1)	7.763***	4.453***	5.306***	4.608***	6.494***	5.114***	5.698***	4.647***
	(2.712)	(7.840)	(8.428)	(9.286)	(10.096)	(12.058)	(12.024)	(13.073)
Amihud (t-1)	-4.081*	14.874***	-2.500	1.766	-3.473**	0.819	-0.790	1.664
	(-1.806)	(3.628)	(-1.636)	(1.636)	(-2.326)	(0.796)	(-0.481)	(1.620)
Bid-ask spread (t-1)	20.131	-24.448***	-0.736	-22.276***	2.938	-20.396***	3.157	-22.142***
	(0.867)	(-6.822)	(-0.097)	(-5.596)	(0.446)	(-3.966)	(0.616)	(-4.718)
Index mutual fund ownership	-15.150	0.784	-1.448	-1.790	-5.655**	-0.577	-5.799**	0.064
	(-1.153)	(0.517)	(-0.548)	(-1.014)	(-2.013)	(-0.457)	(-2.352)	(0.053)
Active mutual fund ownership	4.006*	2.469***	1.254***	1.977***	1.338***	1.975***	1.174***	2.179***
	(1.769)	(7.511)	(2.821)	(7.103)	(4.186)	(8.703)	(4.910)	(10.058)
Month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,725	4,838	7,292	9,907	11,137	15,017	15,281	19,714

Appendix Table A3. Regression Discontinuity Design for Variance Ratios (Second Stage)

The table reports estimates from a design exploiting the discontinuity in ETF ownership around the cutoff between the Russell 1000 and Russell 2000 indexes. The frequency of the data is monthly. The table has the second stage regressions of stock-level Variance Ratio on ETF ownership. The variance ratio is computed as the absolute value of the ratio of the variance of fifteen-second log returns on day *t* and 3 times the variance of five-second log returns on day *t* minus 1 using data from the TAQ database, then it is averaged within a month. ETF ownership is instrumented using a dummy for inclusion in the Russell 2000 for stocks in the Russell 1000 before index reconstitution. Stocks are ranked in terms of market capitalization and different ranges of this rank are used for inclusion in the sample: 50 stocks on each side (Column (1)), 100 stocks on each side (Column (2)), 150 stocks on each side (Column (3)), and 200 stocks on each side (Column (4)). The same stocks enter the sample from June after index reconstitution to May of the next year, except if delistings occur. The regressions include a linear specification of the ranking variable (not reported). Variable descriptions are provided in the Appendix. Standard errors are clustered at the date level. *t*-statistics are presented in parentheses. ***, **, and * represent statistical significance at the 1%, 5%, and 10% levels, respectively.

Polynomial:	Linear specification			
Dependent variable:	Intraday Variance Ratio (averaged within the month)			
Sample:	± 50 stocks around cuttoff	± 100 stocks around cuttoff	± 150 stocks around cuttoff	± 200 stocks around cuttoff
Instrument:	In R2000	In R2000	In R2000	In R2000
	(1)	(2)	(3)	(4)
ETF ownership (instrumented)	0.075*	0.025**	0.027***	0.021***
	(1.743)	(2.493)	(3.657)	(3.344)
log(Mktcap (t-1))	0.056*	0.025***	0.025***	0.016***
	(1.844)	(3.374)	(4.606)	(3.082)
1/Price (t-1)	0.456***	0.437***	0.486***	0.456***
	(5.541)	(9.978)	(11.227)	(12.329)
Amihud (t-1)	-0.701***	-0.917***	-0.996***	-0.996***
	(-6.120)	(-6.249)	(-6.745)	(-6.775)
Bid-ask spread (t-1)	-0.550	-1.503***	-1.613***	-1.418***
	(-0.834)	(-4.042)	(-4.556)	(-5.230)
Month fixed effects	Yes	Yes	Yes	Yes
Observations	3,745	7,328	11,191	15,347

Figure 1: Illustration of the Propagation of Non-fundamental Shocks Via Arbitrage

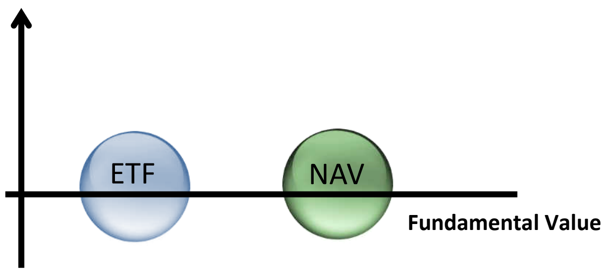

Figure 1a. Initial equilibrium

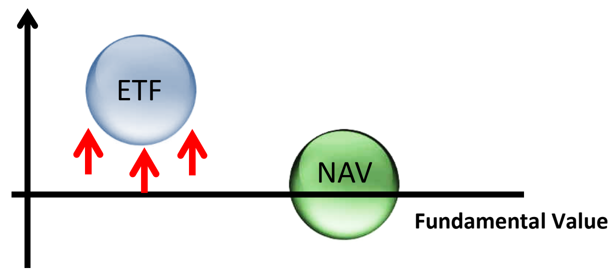

Figure 1b. Non-fundamental shock to ETF

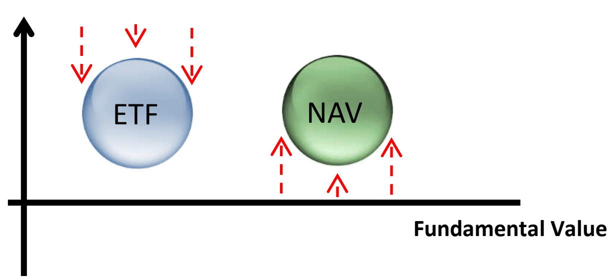

Figure 1c. Initial outcome of arbitrage: the non-fundamental shock is propagated to the NAV, and the ETF price starts reverting to the fundamental value.

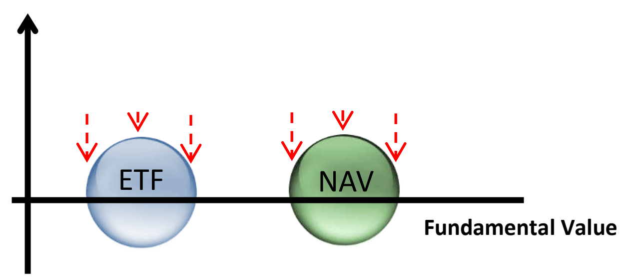

Figure 1d. Re-establishment of equilibrium: after some time, both the ETF price and the NAV revert to the fundamental value.

Figure 2: Illustration of the Propagation of a Fundamental Shock with Price Discovery Occurring in the ETF Market

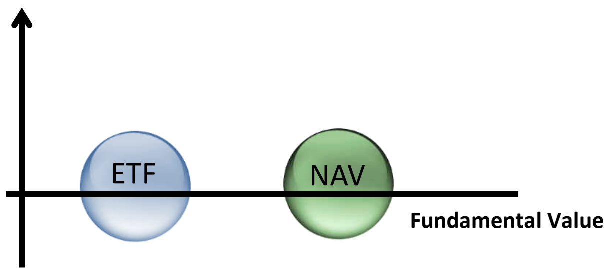

Figure 2a. Initial equilibrium

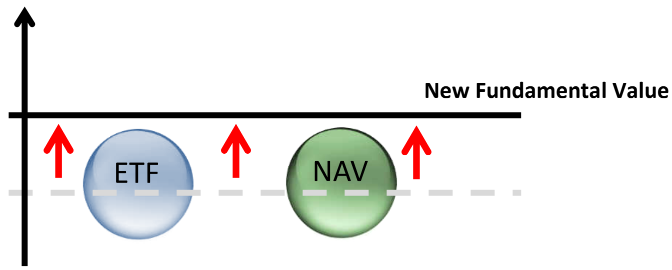

Figure 2b. Shock to fundamental value

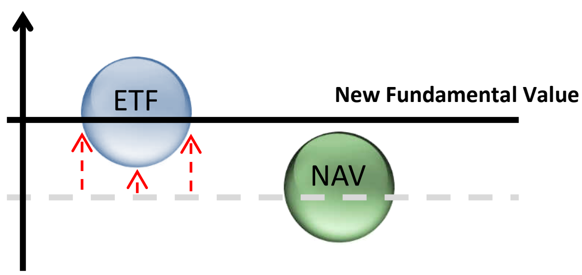

Figure 2c. Price discovery takes place in the ETF market. The ETF price moves to the new fundamental value.

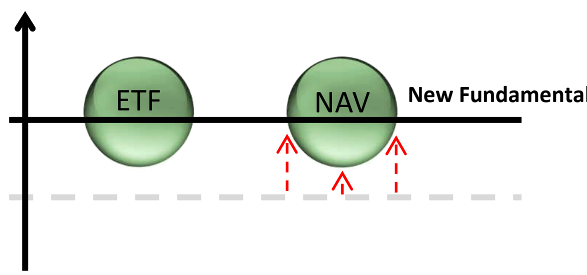

Figure 2d. After a delay, the NAV catches up with the new fundamental.