# Do ETFs Increase Volatility?

Itzhak Ben-David

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

Francesco Franzoni

*University of Lugano (USI) and the Swiss Finance Institute*

Rabih Moussawi

*Villanova School of Business, Villanova University*
*and WRDS, The Wharton School, University of Pennsylvania*

March 2015

**Abstract**

Due to their exceptional liquidity, ETFs are likely to be a catalyst for noise traders. This noise can propagate to the underlying securities through the arbitrage channel. Therefore, we explore whether ETFs increase the non-fundamental volatility of the securities in their baskets. We exploit exogenous changes in index membership, and find that stocks with higher ETF ownership display significantly higher volatility. ETF ownership is also related to significant departures of stock prices from a random walk at the intraday and daily frequencies. Additional time-series evidence suggests that ETFs introduce new noise into the market, as opposed to just reshuffling existing noise across securities.

Keywords: ETFs, volatility, arbitrage, fund flows

JEL Classification: G12, G14, G15

---

We are especially grateful to Martin Oehmke (NBER discussant), 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, Joan Hombert, 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 NBER Summer Institute (Asset Pricing), Insead, HEC Paris, Cambridge Judge Business School, Villanova University, University of Lugano (USI), 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 Center for Financial Policy Conference at the University of Maryland, Jacobs Levy's Quantitative Financial Research Conference at the Wharton School, the Geneva Conference on 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), and SAC Capital Advisors 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

Passive investing is gaining popularity in the asset management industry. In 1980, almost all mutual funds followed active strategies, but by the end of 2014, 30% of assets were in passive allocations (Morningstar, 2015). Exchange Traded Funds (ETFs) are playing the leading role in the rise of passive investing. These vehicles were non-existent in the U.S. in 1993, when the first ETF tracking the S&P 500 was launched. At the end of 2014, they had a market capitalization of $2 trillion, which is almost half of the passive mutual fund industry.[1] Some authors argue that the shift to passive investing is welfare improving, given the drop in intermediation fees and the improvement in portfolio diversification that index funds provide (French, 2008). Furthermore, Stambaugh (2014) argues that the rise in passive investing is symptomatic of improved market efficiency, as profit opportunities for active managers are shrinking.

However, because of their peculiar characteristics, ETFs do not conform to the traditional view of passive funds as buy and hold investors. For example, ETFs provide intraday liquidity to their investors. As a result, they attract high-frequency demand, which translates into price pressure on the underlying securities, due to the arbitrage relation between the ETF and its basket. This trading activity is potentially destabilizing for the underlying securities' prices because it likely reflects non-informational motives. (Arguably, informed traders exploit their advantage by exchanging individual securities, as opposed to index products, such as ETFs.) To compound this effect, the lower trading costs of ETFs relative to the underlying securities can increase the rate of arrival of demand shocks to the market. Specifically, trading strategies that were too expensive without ETFs suddenly become affordable thanks to these instruments. Noise trading can therefore leave a bigger footprint on security prices because of ETFs, suggesting that ETFs may pose new challenges to the efficient pricing of the underlying securities.

Despite the ways in which ETFs differ from traditional passive funds, and despite their prominent role in today's investment space, there has been virtually no work exploring ETF's potential to impound noise in the underlying securities.[2] This paper aspires to fill this gap by

---

[1] ETFs, along with other exchange traded products (ETPs), have reached $2.8 trillion of assets under management (AUM) globally as of December 2014 (BlackRock, December 2014). Also important, ETPs are involved in an increasing share of transactions in equity markets. For example, in August 2010, exchange traded products accounted for about 40% of all trading volume in U.S. markets.

[2] A few papers test whether ETFs have a destabilizing effect, but most of them focus on specific types of ETFs or specific events. Cheng and Madhavan (2009) and Trainor (2010) investigate whether the daily rebalancing of

providing the first large sample evidence on the role of ETFs in propagating noise. We study whether the prices of the securities with higher ownership by ETFs display higher volatility and are more likely to depart from a random walk. The analysis focuses on plain vanilla ETFs that physically replicate U.S. stock indexes, which hold the large majority of assets in the industry (81% of AUM in U.S. ETFs).

The conjectured channel of noise propagation is arbitrage trading. The demand shocks in the ETF market put pressure on ETF prices. To the extent that the ETF price deviates from the net asset value (NAV) of the portfolio holdings, arbitrageurs trade the underlying securities in the same direction as the initial price pressure. Thus, arbitrage can transfer price pressure from the ETF market to the portfolio holdings. This effect is similar to that of mutual fund flows on the prices of the portfolio holdings (Coval and Stafford, 2007; Lou, 2012). The main difference relative to mutual funds is that transactions in ETFs, as well as arbitrage activity, take place continuously throughout the day. This fact makes ETFs a more rapid conduit for the propagation of demand shocks than other managed portfolios.

Our empirical analysis starts by showing that ETFs attract short-term investors. ETFs are, on average, significantly more liquid than the basket of underlying securities in terms of bid-ask spread, price impact, and turnover. For example, the value-weighted portfolio of all equity-based ETFs in the U.S. trades at a bid-ask spread that is 20 basis points (bps) lower than the spread for the equivalent portfolio of underlying stocks. Theories positing that short-horizon clienteles self-select into assets with lower trading costs (Amihud and Mendelson, 1986) suggest that ETFs should be the preferred habitat of high-turnover investors. Indeed, using 13-F institutional holdings data, we find that the institutions holding ETFs have a significantly shorter horizon than those holding the underlying securities. We take this evidence as satisfying a necessary condition for the argument that ETFs are more appealing than stocks for noise traders who wish to express their views at a low cost and high frequency.

In the core of our analysis, we test whether there is a positive causal link between ETF ownership and noise in stock prices. ETF ownership is the total fraction of a stock's capitalization

---

leveraged and inverse ETFs increases stock volatility; they find mixed evidence. Bradley and Litan (2010) voice concerns that ETFs may drain the liquidity of already illiquid stocks and commodities. Madhavan (2012) relates market fragmentation in ETF trading to the Flash Crash of 2010. In a recent study, Da and Shive (2014) find that ETF ownership increases the comovement of stocks in the same basket.

that is held by ETFs. We find a positive relation between ETF ownership and stock volatility. We can argue that the relation is causal thanks to the natural experiment provided by the Russell index reconstitution. In addition, prices of stocks with higher ETF ownership display stronger deviations from a random walk at the intraday and daily frequencies, which is consistent with the increase in volatility being due to noise

We use two different empirical strategies to generate these results. In our first strategy, we use OLS regressions of daily volatility on ETF ownership at the stock level and at a monthly frequency. In this analysis, a one-standard-deviation increase in ETF ownership is associated with a statistically significant increase in daily volatility that ranges between 9% and 15% of a standard deviation, for S&P 500 stocks. The effect is, therefore, economically significant. The magnitude is smaller by a factor of four, but still statistically significant, when we extend the sample to a universe that includes smaller firms (Russell 3000). The effect is weaker for these stocks, probably because ETF arbitrageurs focus on the largest stocks in each basket when trading the replicating portfolios, in order to minimize transaction costs and to achieve larger profits.

The observed increase in volatility is consistent with greater noise in stock prices. However, it could also reflect higher investor attention, which makes prices react more strongly to fundamental information, as shown by Andrei and Hasler (2015). To investigate whether the increased volatility reflects an increase in noise, we measure the impact of ETFs on the mean-reverting component of prices. First, we construct the absolute difference from one of intraday and daily variance ratios of stock returns (when the variance ratio equals one, prices follow a random walk (Lo and MacKinlay, 1988; O'Hara and Ye, 2011). We find that the deviation in the variance ratios of stock returns from unity increases with ETF ownership, suggesting a link between the presence of ETFs and lower price efficiency of the underlying securities. We also estimate predictive regressions of stock returns as a function of ETF flows at the stock level and daily frequency. We find that almost half of the contemporaneous positive impact of flows reverts over the next twenty days, confirming that the presence of ETFs is significantly related to the mean-reverting component of prices.

Our second empirical strategy aims at identifying truly exogenous variation in ETF ownership. Although the OLS regressions control for observable stock characteristics and include stock fixed effects, there is a legitimate concern that ETF ownership is an endogenous variable.[3] To address this concern, we rely on the natural experiment provided by the annual reconstitution of the Russell indexes. We draw inspiration from Chang, Hong, and Liskovich (2015) who implement a regression discontinuity design (RDD) exploiting the mechanical rule allocating stocks between the Russell 1000 (top 1000 stocks by market capitalization) and the Russell 2000 (next 2000 stocks by market capitalization) indexes in June of each year. Due to the large difference in index weights, the top stocks in the Russell 2000 receive significantly larger amounts of passive money than do the bottom stocks in the Russell 1000. For our purpose, a switch to either index generates a large amount of exogenous variation in ETF ownership, which we use to identify the effect of interest in a close neighborhood of the cutoff.

This empirical methodology confirms that the impact of ETF ownership on volatility is positive and strongly statistically significant. The RDD estimates exceed those from the OLS regressions, averaging around 55% of a standard deviation, which suggests a negative omitted variable bias in the OLS specifications. The replication of the variance ratio exercise within the RDD context also confirms the sign and significance of the OLS results with a larger magnitude. To make sense of the larger RDD coefficients, we also note that the RDD slopes measure the weighted average effect across the units in the sample, giving more weight to units that are more likely to receive treatment (the 'index switchers', in our context). Hence, for stocks far away from the cutoff, the effect is likely to be closer to the smaller OLS estimates.

We provide additional evidence on the channel that drives the effect of ETFs on volatility. According to the main hypothesis of the paper, the impact of noise traders on ETF prices propagates to the prices of the underlying securities because arbitrageurs take hedging positions in portfolios replicating the ETF basket. These trades occur whenever the ETF price diverges from the NAV. To test this channel, we ask whether the impact of ETF arbitrage activity on stock prices is weaker for securities that display higher arbitrage costs. Indeed, we find that a proxy for arbitrage activity (the difference between the ETF price and the NAV, labeled 'mispricing') has a smaller effect on volatility and noise for stocks in the top half the distribution of the bid-ask spread and of

[3] For example, new ETFs might track investment themes that have gained popularity among investors. The stocks in these segments of the market might be more volatile because of the attention they already receive, not because ETFs attract noise trading. This mechanism would generate a positive bias in the OLS estimates. Alternatively, index members, which end up in ETF portfolios, might systematically be less volatile than non-index members because they are more established companies. This fact would generate a negative omitted variable bias.

share-lending fees, i.e., for stocks with higher limits to arbitrage. Moreover, strongly supporting the arbitrage channel, the coefficient on share-lending fees is significant only in the subsample for which the arbitrage trades involve shorting the stock (that is, when mispricing is negative).

Thanks to the possibility of identifying exogenous variation in ETF ownership, we can conclude that ETFs increase noise in the prices of the stocks that they own. This result is sufficient to establish a new dimension along which institutional trading can destabilize prices, and it runs contrary to the belief that the rise in passive investing is unambiguously related to increased pricing efficiency.

This cross-sectional evidence, however, could follow from a migration of noise traders from securities with low ETF ownership to those with high ownership. Therefore, in the last part of our analysis, we address the question of whether ETFs induce an increase in the overall noise in the stock market, as opposed to just a redistribution of noise across securities. ETFs can attract new noise traders because of their innovative characteristics. Specifically, relative to standard mutual funds, which clear trades once a day, ETFs permit continuous intraday trading at low cost. In addition, ETFs provide a variety of investment themes previously not offered by mutual funds. Finally, ETFs allow the expression of negative views through short selling, which is not possible with mutual funds. In general, it is plausible that ETFs enable investors to take long and short positions at higher frequency and lower cost, on a wider range of asset classes. Consequently, noise trading has the potential to leave a bigger footprint in the market thanks to the introduction of ETFs.

In support of this conjecture, we provide suggestive time-series evidence that the average share of ETF ownership in the market is positively associated with average stock volatility. The effects persists when we control for aggregate ownership by other mutual funds and include a time trend, which helps to absorb omitted factors. Although the identification strategy does not allow us to make unambiguous causal inference from these times-series estimates, the finding is fully consistent with the view that ETFs attract non-fundamental demand that would not otherwise reach the stock market if ETFs did not exist. In other words, the overall noise appears to increase because of ETFs.

Our study relates to different strands of the literature. There is mounting evidence on the role of institutions in impounding non-fundamental shocks into asset prices as a result of flows from their investors (Brunnermeier and Nagel, 2004; Coval and Stafford, 2007; Ben-David, Franzoni, and Moussawi 2012; Cella, Ellul, and Giannetti, 2013; Lou, 2012; Vayanos and Woolley, 2013). We highlight a previously unexplored channel that is typical of a new class of institutional portfolios: arbitrage activity between ETFs and the underlying baskets. Our paper indirectly relates to the rich literature on the effect of indexing (Shleifer, 1986; Barberis, Shleifer, and Wurgler, 2005; Greenwood, 2005; Wurgler, 2011; Chang, Hong, and Liskovich, 2015). The trigger for the effect that we measure is trading in ETFs, as opposed to index reconstitution. Index membership matters only in defining the stocks that are affected by ETFs. Moreover, by documenting an effect of institutional ownership on volatility, we join a body of work that focuses on the impact of institutions on the second moments of returns (Greenwood and Thesmar, 2011; Anton and Polk, 2014; Lou and Polk, 2014). Closely related, Basak and Pavlova (2013a, 2013b) argue theoretically that the inclusion of an asset in an index tracked by institutional investors increases the non-fundamental volatility in that asset's prices.

Another theme in the literature that our study relates to is the long-running debate on the effect of derivatives on the quality of the underlying securities' prices. On one side of the debate is the concern that liquidity shocks in derivatives markets can trickle down to the cash market, adding noise to prices. For example, Stein (1987) shows that imperfectly informed speculators in futures markets can destabilize spot prices. Among the supporters of the alternative view, Grossman (1989) 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; Turnovsky, 1983).[4] We contribute to this literature by providing systematic evidence from a new type of derivative, ETFs. In December 2014, the assets under management in ETFs tracking the S&P 500 surpassed the open interest in futures on the same index, suggesting that ETFs are becoming the derivative of choice to achieve exposure to the stock market (Amery, 2015).

[4] Earlier studies that examine the impact of derivatives on volatility focused on futures. The proposed economic channel in this literature is the same as the one that we test in this paper. 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.

The paper proceeds as follows. Section 2 provides institutional details on ETFs and develops the testable hypotheses, while Section 3 describes the data. Section 4 presents the main evidence of the effect of ETF ownership on stock volatility and noise. In Section 5, we provide evidence on the role of arbitrage in driving the main effect on volatility and noise. Section 6 addresses the question of whether ETFs attract a "new layer" of volatility to the stock market. Section 7 concludes.

# 2 Institutional Details and Hypotheses Development

## 2.1 Mechanics of Arbitrage

Exchange traded funds (ETFs) are investment companies that typically focus on a single 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 (ticker: SPY). SPY is currently the largest ETF in the world, with about $181 billion of assets (December 2014). In 1995, another SPDR, the S&P MidCap 400 Index (ticker: MDY) was introduced, and the number of ETFs subsequently 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 the S&P 500 and Russell 3000 universes 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 0.22% in 2000 to 3.90% in 2012. The table shows that the number of ETFs holding the average stock in the S&P500 universe grew from about 2 to about 49 during the same period. The average assets under management (AUM) for ETFs holding S&P 500 stocks was, in 2012, about $5bn. The statistics for the Russell 3000 universe paint a similar picture.

Unlike futures, ETFs do not involve a rollover of the expiring contract. Rollover can erode performance for investors with horizons spanning beyond the short maturity of a futures contract. According to BlackRock, the annualized rollover cost of a futures position in large cap stocks

(S&P 500, Euro Stoxx 50, FTSE 100) ranges from 0.9% to 1.4%. The total expense ratio for an ETF on the same indexes can be as low as 0.05% (e.g., the Vanguard S&P 500 ETF, ticker: VOO). This lower cost can explain the fact that in December 2014 the assets in ETFs tracking the S&P 500 surpassed the open interest in futures contracts for the first time (see Amery, 2015).

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% of the assets in the sector according to BlackRock (December 2014). These more complex products are studied by Cheng and Madhavan (2009), among others. We also omit active ETFs that are still below 1% of AUM in the sector.

ETFs are traded in the secondary market by retail and institutional investors, in a similar fashion to closed-end funds. 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).[5] Some institutional investors (called "authorized participants," or APs) who 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 the creation/redemption of ETF shares, we distinguish the two cases of (i) an ETF premium (the price of the ETF exceeds the NAV) and (ii) an ETF discount (the ETF price is below the NAV). In the case of a premium, APs have an incentive to buy the underlying securities, submit them to the ETF sponsor, and ask for newly

[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 orders of magnitude 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.

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 a 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 number of units involved. This fee would amount to, at most, 2.9 bps for a single creation unit in SPY (that is, 50,000 shares worth about $10.2 million as of December 2014), or 0.6 bps for five creation units. During our sample period (2000–2012), share creation/redemption occurs on 9.2% of the trading days for the average ETF. However, a share creation event occurs on 72% of the trading days in our sample, across all ETFs. For the largest ETF, the S&P 500 SPDR, flows into and out of the fund occurred almost every day in 2012 (99.2% of the trading days).

While the occurrence of share creations/redemptions for the average ETF is not frequent enough to justify an impact on price volatility, ETF arbitrage also takes place continuously throughout the day as a result of the activity of hedge funds and high-frequency traders. These investors do not need to engage in primary market trades. On the secondary market, they can buy the inexpensive asset and short sell the more expensive one between the ETF and the basket of underlying securities. They hold the positions until prices converge, at which point they close down the positions to realize the profit. ETF sponsors facilitate arbitrage 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 a very popular trading strategy in recent years. According to some industry participants, statistical arbitrage accounts for 50% of the volume in the S&P 500 SPDR, which is the most traded security in the U.S. with $26 billion average daily volume (last 3 months of 2014).[7] This intraday arbitrage is the most relevant channel of propagation of noise from the ETF market to the underlying securities, according to the main testable hypothesis of this paper.[8]

These institutional details, with some modifications, also apply to synthetic ETFs, which replicate the performance of the index using total return swaps and other derivatives, and for which creation and redemption are handled in cash. The secondary market arbitrage still involves transactions in the underlying securities. Thus, the potential for the propagation of demand shocks from the ETF market to the underlying securities via arbitrage is also present among synthetic ETFs. Similarly, the arbitrage process is an inherent characteristic of all types of ETFs, beyond the equity-based ones that are studied here. Hence, one should expect the effects that we describe in this paper to play out for all types of underlying assets.

## 2.2 ETFs vs. Stocks: Liquidity, Investor Types, and Trading Horizon

The main testable hypothesis of the paper, discussed in detail below, posits that ETFs are appealing to noise traders because they are more liquid than the underlying securities. In this subsection, we study how ETF liquidity contrasts to that of their portfolio constituents. Moreover, in order to provide a description of the users of ETFs, we compare the clienteles of ETFs and common stocks in terms of their trading horizon and institutional type.

The bid-ask spreads on ETFs are particularly low, potentially due a lack of information asymmetry. For a few representative ETFs, Madhavan and Sobczyk (2014) provide evidence that the bid-ask spread is lower than the average spread in the corresponding basket. These authors put forward a convincing argument for the higher liquidity of ETFs. Informed investors are more likely to trade individual securities and market makers impose higher bid-ask spreads to overcome

[7] http://ftalphaville.ft.com/blog/2009/07/30/64451/statistical–arbitrage–and–the–big–retail–etf–con/. Also see http://www.indexuniverse.com/publications/journalofindexes/joi–articles/4036–the–etf–index–pricing–relationship.html for a detailed example of ETF arbitrage. See http://ftalphaville.ft.com/2011/05/18/572086/how-profitable-is-etf-arbitrage/ and http://ftalphaville.ft.com/blog/2011/06/06/584876/manufacturing–arbitrage–with–etfs/ for the profitability of ETF arbitrage. ETF prices can also be arbitraged against other ETFs (Marshall, Nguyen, and Visaltanachoti, 2013) or against futures contracts (Richie, Daigler, and Gleason, 2008).

[8] 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 (e.g., 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 referring to arbitrage, we imply the broader definition of "risky arbitrage."

adverse selection. In contrast, investors who place uninformed directional bets or trade for hedging purposes are more likely to trade entire baskets, such as ETFs. As a result, ETF spreads do not contain an adverse selection premium.

We carry out a similar analysis in our sample covering all U.S.-equity-based ETFs listed on U.S. exchanges (660 different products; see Section 3 for details on sample construction). In Table 2, Panel A, we present systematic evidence on the difference in liquidity between ETFs and the underlying portfolios along three dimensions: the percentage bid-ask spread, the Amihud (2002) measure of price impact, and turnover. For all the ETFs in our sample, we compute the average of each liquidity measure across all the stocks in the basket in a given quarter. Then, to replicate the strategy of an investor that allocates funds to all ETFs according to their market capitalization, we take the value-weighted mean of these measures across all ETFs in a given quarter. The table reports the time-series average of these means in the 52 quarters of the sample (2000:Q1-2012:Q12), along with the results of tests for the statistical significance of their difference. Along all three dimensions, the average ETF is significantly more liquid than its basket stocks. The bid-ask spread is lower by about 20 bps. Price impact, as measured by the Amihud ratio, is also significantly lower for ETFs. Finally, ETFs' turnover is higher by about 8.3%.

A corollary of the conjecture that ETFs are more liquid than the underlying baskets is that ETF investors should display higher turnover. This prediction stems, for example, from Amihud and Mendelson's (1986) clientele effect, whereby short-horizon investors choose to trade in more liquid securities. The bottom of Table 2, Panel A supports this conjecture. We compare ETFs to their underlying baskets in terms of two measures of the investor churn ratio. The first measure comes from Cella, Ellul, and Giannetti (2013), who compute an institutional-investor-level churn ratio as the sum of quarterly absolute changes in dollar holdings over average assets under management, using institutional holdings in the 13-F filings. This measure is then averaged across institutions at the stock level using the fraction of a company held by each institution as weight. The second measure differs only in that the investor-level churn ratio is computed as the minimum between the absolute value of buys and sells, divided by prior quarter holdings.[9] In Table 2, Panel A, we note that the average ETF has a significantly higher investor churn ratio than its underling

[9] Buys (sells) are the sum of the dollar value of the quarterly positive (negative) changes in stock holdings for a given institutional portfolio, as reported in the SEC 13-F form. Values are computed using beginning-of-quarter prices.

basket by about 6.7% per quarter, for the first measure, and 2.9%, for the second measure. These differences are economically significant as the average churn ratio for the basket of stocks is 24% and 12.5%, respectively, for the two measures. The evidence confirms that ETFs, rather than stocks, are the preferred habitat of investors with a short trading horizon.

Next, we compare classes of ownership for all ETFs in our sample and all common stocks in CRSP. Panel B of Table 2 uses Thomson-Reuters' classification of institutional owners filing the 13-F form. We provide variable definitions in Appendix Table A1 and the detailed definition of various investor classes in Appendix Table A2. The panel reports shares held by each group as a fraction of total shares outstanding. The first striking fact is that the institutional ownership of ETFs is by far smaller (at 47.4% on average) than the institutional ownership of stocks (at 62.1% on average) throughout the entire sample period.[10] One can roughly infer retail ownership as the complement to one of institutional ownership.[11] Based on Stambaugh's (2014) argument that noise traders are mostly present among retail investors, this evidence suggests a higher density of uninformed investors among ETF clients.

In analyzing Table 2, Panel B, two additional patterns emerge. First, investment companies, which are mostly comprised of mutual funds, have minimal investments (1.7%) in ETFs, compared to stocks (16.3%). Mutual funds only use ETFs to temporarily park their cash and avoid accumulating tracking error with respect to their benchmark. Second, research firms, which include broker-dealers, have greater ETF ownership (5.8%) than ownership of stocks (0.6%). This class of owners, along with hedge funds, corresponds to ETF arbitrageurs and market makers (including APs). In sum, Panel B of Table 2 paints a picture in which ETFs are mostly traded by retail investors, who are more likely to act as noise traders. Arbitrageurs are also overrepresented by virtue of the peculiar arbitrage mechanism that keeps ETF prices in line with the NAV.

[10] It is worth noting that part of the direct institutional ownership in stocks is through ETFs.

[11] This way of computing retail ownership is an approximation due to two elements. First, small institutions managing less than $100 million and professional investors managing solely their own proprietary accounts are not required to file a 13-F form. Second, reported shares include shares that are short-sold. Because we compute ownership as a fraction of shares outstanding, total institutional ownership for a firm could exceed one. This issue is especially relevant for ETFs, as short interest for some ETFs can be very large (even exceeding the total shares outstanding). However, expressing ownership as a fraction of the shares outstanding plus shares short sold would give an even higher estimate of retail ownership: 1 – Institutional Shares / (Shares Outstanding + Shares Sold Short) > 1 – Institutional Shares / Shares Outstanding.

From Panel A of Table 2, we learn that ETF investors have a significantly shorter investment horizon than investors in the underlying baskets. A related question is whether *all* investor classes turn over their ETF portfolio more often than they turn over their stock portfolio. Panel C of Table 2 addresses this issue by computing separately the quarterly churn ratio of the ETF and stock portfolios, for each institution filing a 13-F form, using Cella, Ellul, and Giannetti's (2013) churn ratio as defined above. The churn ratio is then averaged within each investor class. The striking evidence is that all groups of institutions trade their ETF portfolios faster than their stock portfolios, except for Venture Capital, which nevertheless holds a negligible fraction of ETFs (Panel B).

The institutional class with the fastest turnover in ETFs is hedge funds (85.9% quarterly). Besides being arbitrageurs in the ETF market, hedge funds use ETFs to take directional bets on specific market segments or asset classes. Also, ETFs are part of statistical arbitrage strategies to hedge market or industry risk when taking positions in mispriced securities. In the next subsection, we argue that the arbitrage trades employing ETFs have the potential to propagate mispricing to the underlying securities.

### 2.3 Hypotheses Development

Our empirical analysis draws inspiration from the literature on the destabilizing impact of institutional flows (e.g., Coval and Stafford, 2007; Lou, 2012; Vayanos and Woolley, 2013). The ETF market that we study differs from the typical framework of this prior work by the fact that investors can trade ETF shares in the secondary market continuously throughout the day. This high-frequency arbitrage activity can transfer the price pressure from the ETF market to the prices of the underlying securities. As a result, the demand for ETF shares translates into demand for the underlying securities, similarly to the effect of mutual fund flows. What makes ETFs special, relative to standard mutual funds, is that ETFs allow investors to access the market continuously and at a low trading cost. Hence, ETFs attract potentially more noise trading than standard mutual funds do.

The main testable hypothesis of the paper is that ETFs are a catalyst for noise traders and that noise propagates to the underlying securities via arbitrage. According to this hypothesis, stocks with higher ETF ownership should display higher non-fundamental volatility, everything else being equal.

To illustrate the arbitrage channel for noise propagation, we imagine a situation in which the ETF price and the net asset value (NAV) of its portfolio are aligned at the level of the fundamental value, as in Figure 1a. Then, a noise trading shock, i.e., one that is unrelated to fundamentals, hits the ETF market. Arbitrageurs absorb the liquidity demand by shorting the ETF. Because they are risk averse, 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 that they take, the prices of the basket securities have to rise, as in Figure 1c. Eventually, when other sources of liquidity materialize, prices revert to fundamentals (Figure 1d).[12]

To make a concrete example of this channel, consider hedge funds' trading practices in ETFs. This group of institutions has the highest turnover (see Table 2, Panel C) and, therefore, has a high likelihood of being the marginal investor in the ETF market. Some hedge funds that specialize in high-frequency strategies carry out arbitrage trades of ETFs against the underlying baskets. These trades conform to the mechanism described in Figure 1. In addition, hedge funds can impound mispricing indirectly through their use of ETFs in statistical arbitrage. Suppose hedge funds buy an underpriced stock and hedge the industry risk by shorting the corresponding sector ETF. This trade puts downward pressure on the ETF price (as in Figure 1b). Then, cross-market arbitrageurs transfer the price pressure to the securities in the ETF basket (as in Figure 1c). This argument suggests that ETFs can propagate mispricing to the underlying securities not only because they are traded directly by uninformed investors, but also because they are traded indirectly through their participation in long-short strategies that involve other mispriced securities.

We note that the sequence of events in Figure 1 generates predictions that partly overlap with those from an alternative scenario positing gradual price discovery after a *fundamental* shock, as opposed to noise trading. If price discovery occurs first in the ETF market, ETF prices adjust

[12] The maintained assumption is that arbitrageurs have limited risk-bearing capacity. A similar effect arises in models with risk averse market makers, such as Grossman and Miller (1988) and Greenwood (2005).

immediately to the new information, while the underlying securities' prices remain temporarily fixed ("stale pricing"). We illustrate this scenario in Figure 2. The initial equilibrium (Figure 2a) is perturbed by a shock to the fundamental value of the ETF components (Figure 2b). The ETF price moves first because of price discovery (Figure 2c), and the prices of the underlying securities move with a delay because of stale pricing (Figure 2d). In this alternative situation, ETFs improve price discovery and the arbitrage activity facilitates the adjustment of prices to fundamentals. As a result, there could be a positive link between ETF ownership and "good" volatility (i.e., fundamental volatility). To disentangle the two scenarios, it is not sufficient to show that stocks with higher ETF ownership display higher volatility. We also need to show that ETFs are associated with increased mean reversion in prices, which follows from the propagation of noise (as per Figure 1).

The testable hypothesis spelled out above actually posits an *increase* in noise trading in the underlying securities because of ETF ownership. If the same amount of noise traders merely shifted from trading a given stock to trading the ETFs holding that security, the noise in the stock's price would not increase. To observe an increase in noise, one also needs the additional assumption that noise traders prefer ETFs to stocks as their preferred habitat. Under this assumption, the creation of ETFs entails a migration of noise traders from stocks with low ETF ownership to stocks with high ETF ownership. Noise traders, especially those that affect price volatility at high frequency, are likely to be short-horizon investors. Table 2, Panel A reveals that ETFs are more liquid than the underlying securities and, as a result, they attract investors with higher turnover. Therefore, Table 2 provides background evidence that supports the main hypothesis.

Rather than simply redistributing existing noise from securities with low ETF ownership to those with high ownership, this new asset class can cause a *new layer* of noise to materialize in the stock market. The effect could follow from the enhanced trading opportunities that come with ETFs. For example, relative to standard mutual funds (including index funds), ETFs allow intraday trading and shorting at a low cost in a wide variety of market segments.[13] Hence, noise traders can

[13] As an example, the Vanguard 500 Index Fund is a passive mutual fund tracking the S&P 500 with AUM equal to $198.7 billion as of January 2015. It has the same portfolio as the Vanguard S&P 500 ETF, which has AUM of $28.12 billion for the same date. The index mutual fund has total expense ratio of 0.17% for Investor Shares (minimum investment of $3,000) and 0.05% for Admiral Shares (minimum investment of $10,000). The ETF's expense ratio is 0.05% (with no minimum investment). As per the index fund prospectus, Vanguard discourages frequent trading in its funds, with the exception of ETFs. Therefore, the company reserves the right to reject any purchase request of index fund shares without notice and regardless of size. Moreover, Vanguard prohibits investors' purchases into the

gain access to previously unavailable opportunities to express their views. The possibility that ETFs attract a new layer of noise qualifies as a second testable hypothesis of the paper.

To test this conjecture, we shall look for a significant positive relation between the average stock volatility in the market and the average ETF ownership of stocks, in the time series. To measure the effect of interest more closely, we will need to control for other contemporaneous developments in the market. Admittedly, the time-series tests can never completely rule out omitted factors, so that the evidence in favor of this second testable hypothesis will remain suggestive.

## 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. We first draw information from CRSP for all securities that have a historical share code of 73, which exclusively defines ETFs in this data set. 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.[14] Our initial sample consists of 1,673 ETFs between 1993 and 2012.

Because very few ETFs traded during the 1990s, we restrict the sample to the 2000–2012 period. We further 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 affecting the ease of replication (short bias, volatility, and futures-based ETFs, commodities, etc.). Therefore, we exclude leveraged ETFs, 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 the CRSP Style Codes and Lipper prospectus objective

index fund for 60 days after an investor has redeemed out of that fund. Given the higher costs and restrictions of index funds, it seems reasonable to conclude that noise traders with short investment horizons will prefer ETFs to index funds.

[14] Note that in 2011, the time of the first draft of this paper, 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.

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).[15] We end up with 660 distinct equity ETF securities.

We obtain quarterly holdings information using the 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.[16] 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 ownership measure. We also use the Thomson-Reuters Mutual Fund holdings database to compute the ownership by mutual funds other than ETFs, that is, index funds and active funds. To do that, we use the index fund flag in the CRSP Mutual Fund Database, and merge it with Thomson-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 total shares outstanding at day-end to compute the daily market capitalization of each ETF and to measure the net share creations/redemptions (i.e., flows) for each ETF daily. Because CRSP shares outstanding figures are stale during the month, we assess the accuracy of three databases that provide data on shares outstanding 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

[15] The Lipper Asset Code is not sufficient to accurately filter for U.S. domestic equity funds, because the Equity Funds code comprises a wide array of U.S. and global funds that implement various direct investment or alternative/inverse strategies. Instead, we use the Lipper Objective Code classifications that are assigned by Lipper to a specific population of equity funds and that 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: Broad 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, and 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).

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

three and sometimes as many as 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.

As a dependent variable of our main tests, we compute daily stock volatility at the monthly frequency as the standard deviation of daily returns within a month. For the tests that are reported in the appendix, we compute volatility at a daily frequency using second-by-second data from the Trade and Quote database (TAQ). 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.[17]

We extract stock lending fees from the Markit Securities Finance (formerly Data Explorers) database.[18] We use the variable that reports the average lending fee over the prior seven days. Table 3 reports summary statistics for the variables that we use in the analysis. Panel A presents summary statistics for the monthly-stock-level sample of our main regressions; Panel B reports the correlations for the same variables. Panel C presents summary statistics for the variables that are used in the return regressions at the daily frequency. Panel D presents statistics for the stock-day-level sample. We further describe these variables in later sections and provide definitions in Appendix Table A1.

# 4 The Effect of ETF Ownership on Volatility

## 4.1 ETF Ownership and Volatility: OLS Regressions

We start by asking whether ETF ownership leads to an increase in the volatility of the underlying securities. In our first set of tests, we exploit variation in ETF ownership across stocks and over time in a simple OLS framework.

---

[17] We also compute intraday volatility using intraday returns based on National Best Bid and Offer (NBBO) midpoints; the results are similar.

[18] The database contains about 85% of the over-the-counter (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 computes a fee value that is the volume-weighted average of each contract-level security loan fee.

ETF ownership of stock *i* in month *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 at the end of the month:

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

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 $j$, which is extracted from the most recent quarterly report; and $AUM_{j,t}$ is the assets under management of ETF *j* at the end of the month.

Based on Equation (1), variation in ETF ownership 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 a sector index). Second, there is variation in ETFs' assets under management over time and across products. Third, there is variation in weighting schemes. For example, the S&P 500 and the Russell 2000 are capitalization-weighted, but the Dow Jones is price-weighted; also, our sample contains 17 equal-weighted products.

The three sources of variation in ETF ownership present different degrees of exogeneity with respect to the dependent variable of interest, stock volatility. The portfolio weights follow the weighting scheme of the index mechanically. Hence, they are the most exogenous component in Equation (1). One caveat is that, if the weights do not grow at the same rate as the market capitalization at the denominator (e.g., for equal-weighted indexes), there could be a spurious link between ETF ownership and volatility resulting from the correlation between stock size and volatility. To avoid this issue, we include market capitalization (in logarithm) as a control in our regressions. Instead, ETF's AUM as well as the number of ETFs covering a stock are admittedly less exogenous. For example, investors' demand for existing or new ETFs may relate to how popular a given sector or asset class is at a given point in time. This popularity also affects the amount of trading intensity and the volatility of the underlying securities. This argument can generate a positive relation between ETF ownership and volatility that confounds the causal effect that we are trying to identify. On the other hand, the number of ETFs tracking a given stock depends on the number of indexes in which a stock appears. If established, less volatile firms are more likely to be members of an index, then there can be a negative bias in the relation between ETF ownership and volatility.

In our tests, we take several steps to guard against potentially omitted variables. First, we include stock and month fixed effects. In addition, we control for stock size and liquidity as observable characteristics that relate to volatility. Yet, we cannot entirely avoid the concern that ETF ownership is an endogenous variable within this framework. For this reason, in the next subsection, we provide additional analysis that derives exogenous variation of ETF ownership from the annual reconstitution of the Russell indexes.

With this caveat in mind, we start by reporting the results of OLS regressions of daily volatility in a given month on ETF ownership at the end of the prior month. In Table 4, we present separate regressions for S&P 500 stocks and for the broader sample of Russell 3000 stocks. The goal is to assess how the effect of interest varies with firm size. Besides the log of market capitalization, we include controls for liquidity: the inverse of the stock price, the Amihud (2002) illiquidity measure of price impact, and the bid-ask spread. All the controls date from the end of the prior month. We also include stock and month fixed effects in all regressions. Standard errors are clustered at the stock level.

The results of the analysis are presented in Table 4. To ease interpretation, we standardize volatility and ETF ownership by subtracting the sample mean and dividing by the sample standard deviation. From Column (1) of Table 4, we infer that the relationship between ETF ownership and volatility is positive and strongly statistically significant. The economic magnitude is also large, as a one-standard-deviation move in ownership is associated with 15% of a standard deviation change in daily volatility. This result provides full support for the hypothesis that ETFs impound noise in the underlying securities' prices.

Next, we test whether ETF ownership captures a different effect from the ownership of other institutional investors. Among these, open-end mutual funds are the most similar to ETFs because they also receive daily flows. ETFs are, however, different from other open-end funds in that they allow intraday trading. In Column (2) of Table 4, we include lagged ownership by active and index mutual funds, measured in the same way as ETF ownership (and standardized). The coefficients on both mutual fund ownership variables are positive and significant. However, the point estimates of both mutual fund ownership variables are less than a quarter in magnitude than the slope on ETF ownership, which remains intact. Thus, it appears that ETF ownership has an independent and stronger tie to volatility, which, according to the main hypothesis, depends on the fact that ETFs attract high-turnover investors.

In Column (3) of Table 4, we include the lagged dependent variable to address the concern that the persistence in volatility could introduce reverse causality. The coefficient on ETF ownership remains large and significant at 9.3% of a standard deviation.

Extending the universe to smaller stocks (Columns (4) to (6)), the relationship between ETF ownership on volatility is weaker, amounting to about 3.7% of a standard deviation. Moreover, the slope is no longer statistically distinguishable from the coefficients for other mutual funds (Column (5)).

The lower sensitivity of volatility to ETF ownership in a sample that is dominated by small stocks is consistent with the main hypothesis. The arbitrage activity that occurs at high frequency throughout the day does not require the creation or redemption of ETF shares. Hence, arbitrageurs can choose to concentrate on the larger stocks in the ETF baskets when constructing the replicating portfolio, in order to minimize transaction costs. Such behavior, called 'optimized replication' or 'representative sampling,' can explain why smaller stocks inherit less of the noise coming from the ETF market.

Given that ETF trading and the arbitrage activity involving the underlying securities occur intraday, one should expect the effect that we identify to also be visible at higher frequencies. In Appendix Table A3, we replicate the analysis of Table 4 using intraday volatility as the dependent variable, computed from second-by-second returns within a day. We also update ETF ownership on a daily, as opposed to a monthly, basis, and lag it by one day. The results from these daily stock-level regressions confirm the sign and significance from the monthly sample. The economic magnitude is also in the same ballpark (the variables of interest are standardized). This evidence confirms that the effect that we identify relates to the intraday frequency at which ETFs trade. We give more emphasis to the results using daily volatility (Table 4) to stress the fact that we are not merely identifying a microstructure effect that washes out at lower frequencies.

## 4.2 Identification Using a Regression Discontinuity Design

An identification based on cross-sectional and time-series variation in ETF ownership, which underlies the OLS results in Table 4, can raise doubts if the stock-level controls fail to capture characteristics that co-determine ETF ownership and volatility. For this reason, in this subsection we corroborate our main results with a more robust identification approach.

Chang, Hong, and Liskovich (2015) devise an identification strategy that exploits the exogenous variation in the membership of the Russell 1000 and the Russell 2000 indexes; they cast it within a regression discontinuity design (RDD). The identifying assumption in an RDD is that the treated units (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).

The reconstitution of the Russell indexes provides a natural experiment that lends itself nicely to an RDD. The Russell 1000 index is comprised of the top 1000 stocks by market capitalization, while the Russell 2000 includes the next 2000 stocks. Russell Inc. reconstitutes the indexes in June of each year based only on end-of-May stock capitalization; hence, no discretion is involved in index assignment. Index composition remains constant for the rest of the year. For stocks in a close neighborhood of the cutoff, changes in index membership are exogenous events, as they result from random variation in stock prices at the end of May.

Chang, Hong, and Liskovich (2015) corroborate the validity of the RDD in the context of the Russell 1000/Russell 2000 experiment.[19] The authors show that, although the amount of passive assets benchmarked to the Russell 1000 is 2 to 3.5 times 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. Consequently, a significantly larger amount of passive money tracks the top Russell 2000 stocks.

Figure 3 provides evidence that is consistent with the latter finding in the context of ETFs. The figure plots average ETF ownership as a function of market capitalization rankings for the

[19] In particular, they show that firms close to the cutoff are similar in terms of the distribution of some baseline covariates (returns in the month before reconstitution, assets, and earnings per share), a condition necessary for a valid RDD. In addition, they test for manipulation of the ranking variable using McCrary's (2008) methodology and reject this hypothesis.

Russell 3000 universe, in bins of 10 stocks. We note that around the 1000th position, there is a discontinuity in ownership. Stocks immediately after the cutoff appear to display higher ownership than stocks immediately to the left.[20] Spurred by this evidence, we focus on stocks that switch between indexes, and we use the event of a switch as an instrument for ETF ownership. Then, we regress our outcome variable, daily stock volatility, on instrumented ETF ownership. To identify the effect of interest, we rely on the insight from the RDD that variation in ETF ownership around the cutoff is exogenous, once we control for the ranking variable (i.e., market capitalization).

An additional identifying assumption that needs to be satisfied is excludability (or "the exclusion restriction"), that is, the requirement that the event 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. Later in this section, we discuss reasons why this assumption may fail and we propose solutions.

Following Chang, Hong, and Liskovich (2015), we cast our analysis at the monthly frequency. The first index reconstitution in our sample occurs in May 2000. Thus, we include all months between June 2000 and December 2012. We need to restrict the sample to stocks in the neighborhood of the cutoff in order to compare ex-ante identical units. Rules for defining the cutoff changed during the sample period. Until 2006, the cutoff was simply the 1000th position in terms of market capitalization. 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 (2015) 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 squares 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 remains constant for all the

[20] For stocks in the Russell 1000 (to the left of the cutoff), Figure 3 highlights a decrease in ETF ownership as market capitalization increases (moving towards the origin in the graph). This fact is consistent with the negative unconditional relationship between ETF ownership and size (Table 3, Panel B). The negative sign is likely due to the fact that not all ETFs use value weighting, so that portfolio weights do not grow as fast as market capitalization.

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 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 RDD does not require the inclusion of covariates for identification, Lee and Lemieux (2010) suggest that covariates help improve estimation efficiency. In addition to stock characteristics, we include controls for ownership by other institutions that can also increase their holdings of a stock after it becomes part of an index (index and active mutual funds). While time fixed effects are part of the regression, we do not include stock fixed effects because identification in an RDD is inherently cross sectional. Standard errors are clustered at the monthly level. We standardize the ownership variables and volatility in the relevant samples to ease interpretation. Finally, following the requirements of an RDD, we include first (Panels A and B of Table 5) and second degree (Panel C of Table 5) polynomials of the ranking variable (i.e., market capitalization), whose estimates we do not report to save space.

Table 5, Panel A shows 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 instrument 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 the Russell 1000 index 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 remain 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 slope on the switch indicator in Column (1) suggests that ETF ownership in the twelve months after reconstitution increases for the stocks switching to the Russell 2000 by about 12.6% of a standard deviation. Column (2) focuses on stocks that start out in the Russell 2000 in May prior to reconstitution, with the same bandwidth

(±50 stocks around the cutoff). For the stocks that switch to the Russell 1000 after reconstitution, ETF ownership decreases by about 66.3% of a standard deviation.

Across bandwidths, the estimates unambiguously reveal 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, on average, by about 41% of a standard deviation than bottom Russell 1000 stocks. (Re-running these regressions without standardization gives an average estimate of about 45 bps of ETF ownership.) The strong statistical significance of the first stage regressions, except in Column (1), reassures us about the validity of the instrument.

Table 5, Panel B reports the second stage estimates of the effect of ETF ownership on volatility in the next month. Mirroring the layout in Panel A, the instruments are indicators for a switch to either index, and the sample is 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 they are for switchers to the Russell 1000. To explain these different magnitudes, one could speculate that once they appear as top members of the Russell 2000 on arbitrageurs' radar screens, stocks do not immediately leave arbitrageurs' portfolios when they move into the Russell 1000 because they can still prove to be valid hedging instruments.

The magnitudes in Table 5, Panel B are considerably larger than the OLS estimates in Table 4. Leaving aside Column (1), for which even the first stage is barely significant, the average effect across the other columns of Panel B is 55% of a one-standard-deviation increase in volatility for a one-standard-deviation rise in ETF ownership. This finding may suggest that the OLS estimates suffer from a negative omitted variable bias.

Another reason why the RDD coefficients are larger than the OLS estimates is that the two procedures weight the units in the sample differently. In particular, the RDD estimates measure the weighted average treatment effect, where the weights are the ex-ante probability that a unit receives treatment (Lee and Lemieux, 2010). In other words, the RDD estimates over-weight stocks that are highly likely to switch indexes. These stocks can experience a drastic change from not being included in arbitrageurs' strategies to having top weights in these strategies, and vice versa. Arguably, we should expect that changes in ETF ownership have a bigger effect on these stocks. Given these considerations, we are inclined to conclude that the RDD estimates represent an upper bound, while the OLS coefficients are the lower bound, for the effect of ETF ownership on volatility.

A requirement for an RDD is to include controls for the polynomials of the ranking variable, which in our case is market capitalization. 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 of Table 5, we control for a linear specification of the ranking variable. Panel C replicates the instrumental variable estimation with a quadratic polynomial (the first stage is adjusted accordingly). 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. A potential violation occurs if, after appearing among the top stocks in the Russell 2000, a firm becomes more visible to investors. It is then possible that prices react more quickly to fundamental information and returns become more volatile, as shown by Andrei and Hasler (2015). In this case, price efficiency increases. In the next subsection, however, we show that price efficiency decreases with ETF ownership, also within the RDD. Therefore, the evidence seems to rule out this type of violation of the exclusion restriction.

More generally, the exclusion restriction is not satisfied when there is a correlated omitted variable, which varies with index switches and affects volatility. ETF ownership could merely be a proxy for this omitted factor. We cannot conduct a proper test of the exclusion restriction because we have only one instrument (i.e., exact identification). However, we provide further suggestive evidence on the validity of the RDD by combining the cross-sectional identification from the index switching experiment with time-series variation in ETF ownership. In particular, if the RDD is truly measuring the causal effect of ETF ownership, we should observe a stronger impact of index switching on volatility at times when aggregate ETF ownership is larger. The underlying logic is that a larger presence of ETFs in the market should leave a bigger footprint on stocks that switch indexes. Following this argument, we regress stock-level volatility on the interaction between the index switching indicator and the (equally-weighted) average of ETF ownership in the entire market. If the exclusion restriction is satisfied, we expect that switching to the Russell 2000

(Russell 1000) has a more positive (negative) effect on volatility at times when ETF ownership is overall larger.[21] The regressions include additional interactions to control for aggregate ownership by other mutual funds (passive and active) and for a time trend, given that ETF ownership increases over time. The uninteracted variables are also present, as well as the usual stock-level controls, and month fixed effects. Standard errors are clustered by month. The estimates in Panel D of Table 5 are mostly consistent with the causal interpretation for the effect of ETF ownership. In three out of four specifications, the addition to the Russell 2000 has a larger impact on volatility at times of higher average ETF ownership (second row). In two cases, the effect is statistically significant. In all four specifications focusing on the switch to the Russell 1000, the decrease in volatility is significantly larger in months when ETF ownership is higher in the market (seventh row). Although they are not a direct test of the exclusion restriction, these results are reassuring on the validity of the RDD.

Given the outcome of the RDD, we feel more confident in imputing a causal interpretation to the positive relation between ETF ownership and stock level volatility. This evidence is consistent with the main testable hypothesis. We next study whether the observed increase in volatility corresponds to an increase in noise in stock prices.

## 4.3 Identifying the Impact on Non-Fundamental Volatility

### 4.3.1 Variance Ratios

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. In addition, Andrei and Hasler (2015) prove theoretically and empirically that investor attention increases the sensitivity of prices to fundamentals and, therefore, volatility. If ETF arbitrage makes prices adjust more promptly to fundamentals, or if stocks in ETFs are exposed to higher investor attention, it could be the case that the *fundamental* volatility of the underlying securities goes up. This increase in volatility differs from the prediction of the hypothesis that is

[21] We remind the reader that we expect a switch to the Russell 2000 (Russell 1000) to increase (decrease) volatility because that switch increases (decreases) ETF ownership, based on the evidence in Panel A of Table 5.

tested in this paper, which instead focuses on *non-fundamental* volatility, or noise as defined by Black (1986).

O'Hara and Ye (2011) use variance ratios to measure price efficiency. At time *t*, stock *i*'s variance ratio is defined 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 returns in the estimation window corresponding to time *t*, and the denominator is *k* times the variance of the single-period log returns in the same window *t* (also see Lo and MacKinlay, 1988).[22] As argued by these authors, in an efficient market, the ratio of variances should be closer to one as prices follow a random walk, and the quantity in Equation (2) approaches zero. This simple device provides a non-parametric test of the impact of ETFs on non-fundamental volatility. If ETFs add noise to prices, *VR* should increase with ETF ownership.

In our application, we construct the variance ratio using two different horizons. First, we measure single-period returns from transaction prices at five-second intervals and choose $k = 3$, so that multi-period returns are measured over 15-second intervals. To estimate both variances, we use all the returns within a day. Then, we average the daily estimates over the month to obtain monthly observations. The choice of 15-second time-intervals follows from the observation that ETF sponsors disseminate information about the portfolio NAV at 15-second intervals to facilitate high-frequency arbitrage. This frequency is therefore relevant to capture the intraday effect of arbitrageurs on the underlying stock prices. Second, to show that the effect of interest persists at lower frequencies, we also compute the ratio of the five-day return variance to five times the one-day return variance. To have sufficient observations to estimate these variances, we use all the returns within a quarter.

Table 6 reports estimates from regressions of the standardized values of the stock-level variance ratio on standardized ETF ownership in the prior period. For the 15-second frequency

[22] Strictly speaking, only the first element in the absolute value in Equation (2) is a variance ratio. We label the whole expression as a "variance ratio" for convenience.

variance ratio (VR 15), the sample is monthly, while for the 5-day frequency (VR 5), it is quarterly. We include the same set of controls as in the previous tables.

Panel A shows OLS regressions. The results point unambiguously to a positive and significant relation between ETF ownership and variance ratios. At both frequencies, the evidence suggests that the prices of stocks with higher ETF ownership are farther away from a random walk and therefore contain more noise. The effect is twice as large at the intraday frequency as it is at the five-day frequency (10% of a standard deviation of VR 15 for a one-standard-deviation change in ETF ownership, for S&P 500 stocks). It is, however, statistically and economically significant for the 5-day frequency as well. Consistent with the pattern in Table 4, the effect is reduced, but still significant, when the universe is extended to smaller stocks (Russell 3000).

Given the concerns about the potential endogeneity of ETF ownership in the OLS regressions, we implement the RDD based on the Russell indexes reconstitution, using the variance ratio as dependent variable in the second-stage regressions. Panel B of Table 6 reports the results for VR 15, from the monthly sample, while Panel C has the results for VR 5, from the quarterly sample. In all specifications, the RDD confirms the positive slope on ETF ownership. Statistical significance is present in the majority of cases. Finally, the larger magnitude of the RDD slopes than the OLS slopes mirrors the previous evidence regarding the effect on total volatility and can be explained in the same way. The results from the RDD give us more confidence on the causal interpretation of the positive link between ETF ownership and noise in stock prices.

#### 4.3.2 Price Reversals

An alternative way to test whether ETFs add noise to the underlying securities is to look for direct evidence of the sequence of events predicted by the conjectured channel for noise propagation (summarized in Figure 1). Following a demand shock in the ETF market, the prices of the underlying securities should move in the same direction as the initial shock. Then, because the fundamentals have not changed, prices should revert to the initial level. Finding evidence of mean reversion in prices would also contribute to ruling out the alternative story that ETFs merely improve price discovery (as in Figure 2), which could also explain the increase in volatility.

We use stock-level ETF flows at the daily frequency as a conditioning variable to identify reversals in the prices of the underlying stocks. As explained above, ETF flows (redemptions and creations) are the result of APs' arbitrage activity, which responds to ETF prices' deviations from the NAV. 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. Dollar daily ETF flows are then expressed as a fraction of prior-day stock capitalization.

On the day when flows occur, we expect a price move in the same direction as the flows, irrespective of whether the motive for trade is fundamental or non-fundamental (i.e., noise trading). 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 (five, ten, and twenty days) on stock-level flows, using overlapping daily observations. We include the usual stock-level controls and time fixed effects. The standard errors are clustered at the day level, and we correct for the autocorrelation of residuals induced by overlapping observations for multiday returns using the Newey and West (1987) estimate of variance.

In Table 7, returns are in percent, while net flows are standardized. From Column (1), we note that, on the same day, ETF flows and returns move in the same direction. The contemporaneous price move is about 17.5 bps for a one-standard-deviation change in net flows for S&P 500 stocks. The high significance is not surprising, as flows and returns are measured on the same day (hence, this is not a predictive regression). In addition, we note that the magnitude of the change in prices exceeds the half-spread, which is about 8.5 bps for the sample of large stocks. This magnitude rules out the possibility that flows cause a simple bid-ask bounce.

More relevant to identifying the transmission of noise, ETF flows predict a reversal of the underlying stocks' prices in the next twenty days (Columns (2)-(4)). The evidence is consistent with the conjecture that the demand shocks in the ETF market add a mean-reverting component to stock prices. From Column (4), we can infer that almost half of the initial price impact is reversed (8.3/17.5 = 0.47). Extending the horizon farther out to 40 days does not increase the magnitude of reversals (not reported). As in the prior tables, the absolute effects are smaller in the extended universe of Russell 3000 stocks

In sum, the evidence in this subsection suggests that the positive link between ETF ownership and volatility, which we report in Tables 4 and 5, is consistent with an increase in noise in stock prices. Specifically, ETFs appear to add a mean-reverting component to stock prices both intraday and at the daily frequency. The finding of a 50% reversal of the initial price impact of ETF flows leads us to believe that at least half of the impact of ETFs on return volatility is due to noise propagation.

## 5 Exploring the Arbitrage Channel

Having established a causal link between ETF ownership and noise in the prices of the underlying securities, we next look for evidence that noise propagates through the arbitrage channel. To this end, we first define a proxy for arbitrage activity. The difference between the ETF price and the net asset value of the underlying basket (NAV), labeled ETF mispricing, is a signal for the profitability of ETF arbitrage. Hence, we expect a stock's involvement in arbitrage trades to be a positive function of the mispricing of the ETFs that hold the stock. Using this proxy, we study whether arbitrage activity has an incremental impact on volatility and noise for a given level of ETF ownership.[23]

Then, we conjecture that the proxy for expected arbitrage activity should have a weaker effect on stock prices for stocks that are harder to arbitrage. In other words, we seek evidence that limits to arbitrage play a role in attenuating the propagation of noise to the underlying securities. This evidence would indirectly testify to the importance of the arbitrage channel.

We use two proxies for limits of arbitrage: the stock-level bid-ask spread and share-lending fees. First, because ETF arbitrage involves a roundtrip transaction in the stock, a large stock-level bid-ask spread reduces the profitability of arbitrage trades and therefore the incidence of arbitrage trading in a given stock. Second, when the arbitrage transaction involves shorting the stock (i.e., the NAV is above the ETF price), higher stock-lending fees discourage arbitrageurs. In addition,

[23] It could actually be the case that ETF mispricing signals a *lack of* arbitrage activity. That is, more mispricing is present when arbitrageurs refrain from entering the market. This could be an issue for our tests if the reason why arbitrageurs abstain from their trades is volatility in the underlying securities, which is the dependent variable in our tests. In such a case, the endogeneity of mispricing could bring a positive spurious correlation with volatility. To address this concern, we control for the lagged value of the dependent variable, so that we study the impact of mispricing on innovations in volatility. This contributes to attenuate the endogeneity concern because arbitrage trades are not likely to condition on innovations in volatility in the next period. Further, in the tests in which we interact mispricing with measures of the limits of arbitrage, this potential endogeneity would lead to the opposite sign of the coefficient relative to what we find. See the discussion below.

a high share-lending fee can reflect a shortage of shares for lending, meaning that some arbitrageurs may simply not be able to carry out the trade (Cohen, Diether, and Malloy, 2007).

Given the high-frequency fluctuations in arbitrage activity, we carry out our tests at the daily frequency, which allows us to measure the variables of interest in a more timely way. Thus, the dependent variables for these tests are intraday volatility, which is estimated from second-by-second returns within a day, and the daily variance ratio resulting from the comparison of fifteen-second returns to three times five-second returns within a day (VR 15, see Section 4). The main explanatory variable is the stock-level measure of absolute ETF mispricing in the prior day. This variable is calculated by summing the absolute dollar mispricing (i.e., the difference between the ETF price and NAV, as a fraction of the ETF price, multiplied by the dollar holdings in the stock) across all ETFs holding stock $i$, and expressing this quantity as a fraction of a stock's capitalization:

$$Abs\left(mispricing_{i,t}\right) = \frac{\sum_{j=1}^{J} w_{i,j,t} * AUM_{j,t} * \left|mispricing_{j,t}\right|}{Mkt\ Cap_{i,t}}. \tag{3}$$

This variable interacts the effect of the ETF mispricing, which is a signal for the attractiveness of the stock for arbitrage trades, with the ownership of each ETF in the stock's capital base, which measures the relative importance of each ETF for the given stock. We take the absolute value of mispricing because arbitrage activity is triggered by both positive and negative discrepancies between ETF prices and the NAV. It is therefore important to avoid netting out these deviations across ETFs. For a second set of tests, in which we condition on the direction of the arbitrage trades, we refer to net mispricing, which differs from the definition in Equation (3) for the omission of the absolute value.

In Panel A of Table 8, the dependent variable is intraday volatility. The sample consists of S&P 500 stocks, where, according to our prior results, most of the effect of ETFs occurs. In Column (1), we test whether absolute mispricing at the close of day $t$ – 1, which is a proxy for arbitrage activity on day $t$, has an incremental effect on volatility for a given level of ETF ownership. Besides the usual controls, we include the mispricing on day $t$ – 2 and the lagged dependent variable. The goal is to capture the effect of the innovation in mispricing on the innovation in volatility, given that mispricing on day $t - 1$ could itself depend on volatility (i.e., ETFs holding stocks that are more volatile are more likely to be mispriced, as discussed in footnote 23). We also include the return on the stock on day $t - 1$ to capture variation in mispricing that is exogenous to movements in the stock price itself. That is, we identify variation in mispricing resulting from movements in the ETF price or in the prices of the other stocks in the basket, but not in the stock's own price. We note that the effect of absolute mispricing is positive and significant, amounting to about 2.3 of a standard deviation of the dependent variable for a one-standard-deviation change in mispricing (both variables are standardized). Further, the effect of ETF ownership drops in magnitude relative to the specification without mispricing (see Appendix Table A3). This evidence supports the view that arbitrage activity, as proxied by mispricing, is the transmission channel for the effect of ETF ownership on volatility.

Next, we report specifications that include interactions of absolute mispricing with the proxies for arbitrage costs. For each measure of limits of arbitrage, we define a dummy variable for stocks that are in the top of half of the distribution of the variable in the prior period.[24] We leave out stock fixed effects, because we wish to achieve identification from the cross-sectional variation in the proxies. (Including stock fixed effects has no material impact on the results.) From Table 8, Panel A, Column (2), we infer that the effect of arbitrage on volatility, as proxied by absolute mispricing, is significantly weaker for stocks with a high bid-ask spread. This evidence suggests that limits of arbitrage are playing a role in the transmission of noise to the underlying securities.

Next, we break up the sample by the sign of net mispricing. A priori, we do not expect the sign of mispricing to matter for the interaction with the bid-ask spread, because the arbitrage trade involves a roundtrip transaction in the underlying stock in any case. The results in Columns (3) and (4) confirm this conjecture.

In Column (5), Panel A, Table 8, share-lending fees have a marginally significant impact in attenuating the effect of arbitrage on volatility. More importantly, we now expect this effect to differ based on the sign of net mispricing. Only when mispricing is negative (i.e., the ETF price is below the NAV) does the arbitrage trade involve a short sale of the underlying stocks. The

[24] Information on share-lending fees is sparse, especially in the initial part of the sample. Therefore, we use the average fee in the month.

estimates in Columns (6) and (7) square nicely with this prediction and provide strong evidence for the role of arbitrage activity in generating the effect of interest.

It is worth noting that the sign of the interactions with the proxies for arbitrage costs tends to rule out concerns about the endogeneity of mispricing (see footnote 23). Indeed, if mispricing was capturing the fact that arbitrageurs abstain from trading because volatility discourages them, we would expect this effect to be even stronger for illiquid stocks or for stocks that are hard to locate, given that these characteristics correlate positively with volatility. That is, the sign on the interactions with arbitrage costs should be positive. Instead, contrary to this view, the interactions have negative and significant coefficients.

Panel B of Table 8 replicates the analysis using the variance ratio from intraday returns as dependent variable. Mostly, the results mirror the evidence in Panel A. Hence, the evidence further supports the role of arbitrage in transmitting the noise to the prices of the securities in the ETF baskets.

For completeness, in Appendix Table A4, we report the analysis for the Russell 3000 universe. As expected, in this sample, the effects are weaker or non-existent. These results confirm our prior belief that arbitrageurs tend to focus on the larger stocks in the ETF baskets when doing optimized replication. Therefore, the effect of interest is located mostly among large stocks.

## 6 Are ETFs Attracting a "New Layer" of Noise?

In the previous sections, we show a positive link between ETF ownership and stock volatility. The identification provided by the regression discontinuity design allows us to attach a causal interpretation to these estimates. We also find that ETF ownership increases the mean-reverting component of stock prices. These results are consistent with the argument that stocks with higher ETF ownership are more attractive to noise traders. Therefore, the evidence supports the first testable hypothesis in Section 2.

The second testable hypothesis is that the noise hitting ETF-owned stocks represents a *new layer* of demand, which would not be present in the market if ETFs did not exist. The argument is that ETFs provide previously unavailable trading opportunities, at low cost and high frequency, which cause new traders and/or new trading strategies to materialize. For example, ETFs might make the hedging of industry risk in statistical arbitrage strategies that involve mispriced securities significantly cheaper, so that the volume of these trades might increase. In this sense, the introduction of ETFs is analogous to a decrease in trading costs that enables traders to operate at higher frequency.[25]

The alternative view to this hypothesis is that ETFs merely provide a convenient conduit for existing investors who wish to trade the underlying securities. According to this view, noise is reshuffled from stocks with low ETF ownership to stocks with high ETF ownership. We label this argument the "reshuffling hypothesis." To stay with the previous example, the same statistical arbitrageurs that currently employ ETFs for hedging purposes were previously constructing a hedging portfolio using stocks in the same industry as the mispriced security.

As argued in Section 2, the evidence in Table 2 suggests that ETFs attract high-turnover investors. However, it does not rule out the reshuffling hypothesis, that is, the possibility that these investors would directly trade in stocks had the ETFs not been in existence. Therefore, we need to produce evidence that allows us to more convincingly separate the new-layer hypothesis from the reshuffling hypothesis. This evidence can only come from comparing the time-series evolutions of ETF ownership and volatility. The ultimate prediction of the new-layer hypothesis is that the growth in the ETF market attracts more noise trading to the stock market. Hence, we should observe higher volatility at times of higher ETF stock ownership. The reshuffling hypothesis, instead, predicts that aggregate volatility should not change because of ETFs.

In Panel A of Table 9, we report the estimates from a regression of average daily volatility on lagged average ETF ownership across all stocks in CRSP. The frequency is monthly and volatility is computed using the daily returns in a month. We include lagged volatility to set up the regression as a test for Granger causality. To mitigate the concern that ETF ownership proxies for omitted factors relating to institutional ownership, we include lagged average ownership by index and active funds. Importantly, we add a time trend as a catchall control for developments in aggregate conditions (e.g., a protracted reduction in trading costs). We find that ETF ownership significantly predicts volatility, with a positive sign. The economic magnitude is closer to that in

[25] Of course, a similar argument applies to futures and other derivatives. ETFs, however, allow a higher degree of specialization in terms of the segments of the market they cover.

Table 4 than to that in Table 5. In Column (2), we replicate the analysis in first differences. The results are robust to this modification, and the magnitude is even larger.

This evidence supports the hypothesis that ETF ownership adds a new layer of volatility to the stock market. The caveat is that the time-series identification does not allow us to rule out the possibility that time-varying omitted factors could be driving our results. However, we believe that including controls for the ownership of other mutual funds, as well as a time trend, attenuates this concern.

While this analysis does not support the extreme version of the reshuffling hypothesis for the market-wide effect of ETF ownership on volatility, it is still interesting to ask whether some stocks experience a decrease in volatility at the expense of others as ETF ownership increases. In other words, we ask whether some partial reshuffling of noise trading is taking place. To this purpose, we test whether, as aggregate ETF ownership increases, volatility declines for some groups of stocks and rises for others. We sort the universe of stocks into five quintiles by ETF ownership. The average ETF ownership in the bottom quintile is 0.70% of a stock's capitalization, while in the top quintile it is about 4%. For each quintile, at the monthly frequency, we regress the average volatility for that group of stocks on the (lagged) average ETF ownership across the entire market and include controls. The explanatory variables are the same across quintiles, because the goal is to test whether the same aggregate developments are related to different changes in the volatility of different groups of stocks. From Panel B of Table 9, we note that all groups of stocks experience a significant increase in volatility as aggregate ETF ownership increases. This evidence does not support a partial reshuffling of noise across stocks. Quite relevantly, the effect of interest is strongest in the quintile with top ETF ownership, which strengthens the case for a causal interpretation of the time-series association between ETF ownership and volatility. Finally, running the same regressions in first differences strongly confirms our results (Table 9, Panel C).

To conclude, while the time-series setting of this analysis prevents us from drawing an unambiguous causal inference, the evidence in this section is consistent with the view that ETFs attract a new layer of demand shocks to the market as opposed to causing a reshuffling of existing demand across stocks.

## 7 Conclusion

With $2.8 trillion of assets under management globally (December 2014), ETFs are rising steadily among the big players in the asset management industry. This asset class is also capturing an increasing share of transactions in financial markets. For example, in August 2010, ETFs and other exchange traded products accounted for about 40% of all trading volume in U.S. markets. This explosive growth has attracted the attention of regulators. In particular, the Securities and Exchange Commission (SEC) has begun to review the potential role of ETFs in inflating the volatility of the underlying securities.[26]

The success of ETFs is justified by the fact that these investment vehicles provide an unprecedented source of diversification at low cost and high liquidity. However, the evidence in this paper seems to point to an unintended effect of this relatively new asset class, which can stoke regulators' concerns.

We present results showing that the stocks in ETFs' baskets display higher volatility than otherwise similar securities. Through a regression discontinuity design, we are able to attach a causal interpretation to this finding. The presence of ETFs also causes the underlying securities' prices to diverge from random walks, both intraday and daily. These effects are significantly related to proxies for the intensity of arbitrage activity between the ETFs and their baskets.

This evidence paints a picture in which noise trading in the ETF market is passed down to the prices of the underlying securities by the transmission chain of arbitrage trades. Moreover, because of their ease of trade and cost effectiveness, ETFs attract higher turnover investors than the average stock in their baskets. Consequently, noise in stock prices increases with ETF ownership.

In addition, we find that aggregate volatility varies significantly over time with aggregate ETF ownership in the stock market, controlling for ownership by other mutual funds and for a time

[26] Regulators have investigated 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 lent out). Concerns have been expressed that a run on ETFs might 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*, October 19, 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.

trend. With the caveat that the time-series identification of this effect does not allow for a conclusive causal inference, the evidence suggests that ETFs bring a new layer of noise to the market, as opposed to just causing a migration of existing noise traders across securities. We explain this finding with the new trading opportunities, at low cost and high frequency, made possible by ETFs.

A new theoretical framework seems necessary to gauge the tradeoff between the decreased transaction costs and the improved access to diversification that ETFs bring about and the deterioration in price efficiency revealed by our empirical analysis. The general equilibrium and welfare implications of this important wave of financial innovation therefore remain unclear.

To conclude, the effects that we describe resonate with the literature showing that flows into institutional portfolios impound noise into asset prices (e.g., Coval and Stafford, 2007; Lou, 2012). Along with this prior evidence, our results suggest that the recent rise in institutional stock ownership is not by itself a guarantee that stock prices are more efficient. Noise traders can still cause mispricing through their allocations to institutional portfolios.

## References

Amery, Paul, 2015, ETFs giants bet on futures for flows, *Financial Times*, February 1, 2015.

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.

Andrei, Daniel, and Michael Hasler, 2015, Investor attention and stock market volatility, *Review of Financial Studies* 28(1), 33-72.

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

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, 2013b, A model of financialization of commodities, Working paper, London Business School.

Ben-David, Itzhak, Francesco Franzoni, and Rabih Moussawi, 2012, Hedge fund stock trading during the financial crisis of 2007-2009, *Review of Financial Studies* 25(1), 1–54.

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, 2014, ETP landscape, *Industry Highlights*, December.

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.

Brunnermeier, Markus K., and Stefan Nagel, 2004, Hedge funds and the technology bubble, *Journal of Finance* 59(5), 2013–2040.

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

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, 2015, Regression discontinuity and the price effects of stock market indexing, *Review of Financial Studies*, forthcoming.

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.

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, 2014, 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.

French, Kenneth R., 2008, The cost of active investing, *Journal of Finance* 63(4), 1537–1573.

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.

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.

Grossman, Sanford J., and Merton H. Miller, 1988, Liquidity and market structure, *Journal of Finance*, 43(3), 617-633.

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, 2012, A flow–based explanation of returns predictability, *Review of Financial Studies* 25(12), 3457–3489.

Lou, Dong, and Christopher Polk, 2014, Comomentum: Inferring arbitrage activity from return correlations, Working paper, London School of Economics.

Madhavan, Ananth, 2012, Exchange-traded funds, market structure and the Flash Crash, Working paper, BlackRock Inc.

Madhavan, Ananth and Aleksander Sobczyk, 2014, Price Dynamics and Liquidity of Exchange-Traded Funds, SSRN working paper.

Marshall, Ben R., Nhut H. Nguyen, and Nuttawat Visaltanachoti, 2013, ETF arbitrage: Intraday evidence, *Journal of Banking and Finance* 37(9), 3486–3498.

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

Newey, Whitney K. and Kenneth D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55(3), 703–708.

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.

Stambaugh, Robert F., 2014, Investment noise and trends, *Journal of Finance* 69 (4), 1415–1453.

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

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.

Vayanos, Dimitri, and Paul Woolley, 2013, An institutional theory of momentum and reversal *Review of Financial Studies* 26 (5), 1087–1145.

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).

## 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 stock 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 | 5577.69 | 0.64 | 0.22 | 2.41 | 5138.81 | 0.53 | 0.25 |
| 2001 | 13.45 | 2173.41 | 0.42 | 0.48 | 8.91 | 1062.08 | 0.16 | 0.36 |
| 2002 | 15.47 | 2798.87 | 0.45 | 0.90 | 10.18 | 1185.39 | 0.14 | 0.83 |
| 2003 | 15.95 | 3542.45 | 0.45 | 1.05 | 10.42 | 1465.49 | 0.14 | 0.95 |
| 2004 | 21.40 | 3451.84 | 0.47 | 1.22 | 14.30 | 1702.26 | 0.14 | 1.26 |
| 2005 | 24.75 | 3758.30 | 0.49 | 1.51 | 15.73 | 2040.02 | 0.16 | 1.55 |
| 2006 | 25.80 | 4337.34 | 0.51 | 1.67 | 16.81 | 2447.86 | 0.18 | 1.84 |
| 2007 | 36.04 | 4082.81 | 0.64 | 2.00 | 22.60 | 2439.07 | 0.24 | 2.21 |
| 2008 | 50.61 | 2980.85 | 0.69 | 2.76 | 30.26 | 1789.18 | 0.28 | 2.87 |
| 2009 | 53.19 | 2733.88 | 0.67 | 3.27 | 31.30 | 1710.54 | 0.26 | 3.53 |
| 2010 | 52.08 | 3260.33 | 0.68 | 3.30 | 30.08 | 2311.04 | 0.27 | 3.74 |
| 2011 | 52.77 | 3977.15 | 0.67 | 3.61 | 28.87 | 2937.45 | 0.27 | 3.81 |
| 2012 | 49.25 | 5033.17 | 0.67 | 3.90 | 27.24 | 3429.71 | 0.26 | 3.91 |
| Average | 30.69 | 3563.73 | 0.57 | 2.09 | 20.13 | 2064.87 | 0.21 | 2.37 |

## Table 2. ETFs vs. Stocks: Liquidity, Institutional Ownership, Churn Ratio

The table reports statistics for ETF- and stock-level liquidity, investor turnover, and ownership. Panel A shows the security-level liquidity measures (bid-ask spread, Amihud (2002) ratio, and daily turnover) as well as the churn ratio measures of the investors in the securities (churn ratios 1 and 2). For all ETFs in our sample, we compute the average measure of liquidity or churn ratio across the stocks in the basket in a given quarter. Then, we value-weight the ETF-level and basket-level measures across all ETFs at the quarter level using ETF market capitalization (thus having 52 quarters in our sample). Churn ratio 1 is from Cella, Ellul, and Giannetti (2013), who compute an institutional-investor-level churn ratio as the sum of quarterly absolute changes in dollar holdings over average assets under management (the data are from SEC 13-F filings). This measure is then averaged across institutions at the stock level using the fraction of a company held by each institution as weight. Churn ratio 2 differs only in that the investor-level churn ratio is computed as the minimum between the absolute value of buys and sells, divided by prior quarter holdings. Panel B presents information about institutional ownership: averaged across all 52 quarters, in the first quarter of the sample, and in the last quarter of the sample. Ownership averages are presented for the ETFs in our sample and all the stocks in CRSP. Panel C presents the institutional classes' turnover (churn ratio 1) separately for the ETF portfolio and the stock portfolio. Variable descriptions are provided in Appendix Table A1. *t*-statistics for the test of the null hypothesis that the difference is equal to zero are presented in parentheses. $^{***}$, $^{**}$, and $^{*}$ represent statistical significance at the 1%, 5%, and 10% levels, respectively. The sample ranges between 2000:Q1 and 2012:Q4.

### Panel A: Liquidity and Investors' Churn Ratio Measures

| Liquidity measures | Variable | Quarters | ETFs | Stocks | Difference | t-stat |
|---|---|---|---|---|---|---|
| Security-level | Bid-Ask Spread | 52 | 0.003 | 0.005 | -0.002*** | (-3.518) |
| | Amihud ratio | 52 | 0.002 | 0.008 | -0.006*** | (-9.702) |
| | Daily turnover | 52 | 0.093 | 0.011 | 0.083*** | (13.462) |
| Investor-level | Churn Ratio 1 | 52 | 0.307 | 0.240 | 0.067*** | (10.195) |
| | Churn Ratio 2 | 52 | 0.154 | 0.125 | 0.029*** | (7.493) |

### Panel B: Types of Institutional Ownership

| | Ownership averaged across the sample | | | Ownership at 2000:Q1 | | Ownership at 2012:Q4 | |
|---|---|---|---|---|---|---|---|
| Type of Institution | Quarters | ETFs | Stocks | ETFs | Stocks | ETFs | Stocks |
| All Institutions | 52 | 0.474 | 0.621 | 0.280 | 0.511 | 0.492 | 0.651 |
| Banks | 52 | 0.131 | 0.137 | 0.052 | 0.114 | 0.202 | 0.116 |
| Endowments | 52 | 0.006 | 0.001 | 0.000 | 0.003 | 0.001 | 0.001 |
| Hedge Funds | 52 | 0.033 | 0.030 | 0.022 | 0.019 | 0.028 | 0.036 |
| Insurance | 52 | 0.014 | 0.033 | 0.007 | 0.034 | 0.011 | 0.026 |
| Investment Advisors | 52 | 0.198 | 0.211 | 0.125 | 0.166 | 0.167 | 0.231 |
| Investment Companies | 52 | 0.017 | 0.163 | 0.010 | 0.139 | 0.023 | 0.196 |
| Pension Funds | 52 | 0.009 | 0.035 | 0.001 | 0.031 | 0.008 | 0.026 |
| Individual Investor (in 13F) | 52 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.001 |
| Research Firms | 52 | 0.058 | 0.006 | 0.061 | 0.003 | 0.028 | 0.008 |
| Corporations | 52 | 0.003 | 0.001 | 0.000 | 0.001 | 0.017 | 0.003 |
| Venture Capital | 52 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.001 |
| Private Equity | 52 | 0.000 | 0.001 | 0.000 | 0.000 | 0.000 | 0.001 |
| Sovereign Funds | 19 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.001 |

**Table 2. ETFs vs. Stocks: Liquidity, Institutional Ownership, Investor Horizon (Cont.)**

**Panel C: Institutional Turnover, by Type of Institution**

| Type of Institution | Observations | Turnover in ETFs | Turnover in Stocks | Difference | t-stat |
|---|---|---|---|---|---|
| All Institutions | 52 | 0.671 | 0.247 | 0.424*** | (31.363) |
| Banks | 52 | 0.551 | 0.170 | 0.381*** | (15.282) |
| Endowments | 52 | 0.499 | 0.183 | 0.317*** | (9.185) |
| Hedge Funds | 52 | 0.859 | 0.662 | 0.197*** | (17.064) |
| Insurance | 52 | 0.549 | 0.205 | 0.344*** | (17.129) |
| Investment Advisors | 52 | 0.737 | 0.288 | 0.450*** | (59.553) |
| Investment Companies | 52 | 0.670 | 0.208 | 0.462*** | (18.961) |
| Pension Funds | 52 | 0.660 | 0.145 | 0.515*** | (27.298) |
| Individual Investors (in 13F) | 47 | 0.531 | 0.169 | 0.362*** | (3.544) |
| Research Firm | 52 | 0.661 | 0.456 | 0.205*** | (25.997) |
| Corporations | 40 | 0.300 | 0.257 | 0.043** | (2.288) |
| Venture Capital | 47 | 0.158 | 0.226 | -0.069*** | (-4.865) |
| Private Equity | 27 | 0.455 | 0.221 | 0.234*** | (4.266) |
| Sovereign Funds | 18 | 0.550 | 0.353 | 0.197 | (1.422) |

## Table 3. Summary Statistics

The table presents summary statistics for the variables used in the study. Panels A shows summary statistics for the stock-month sample. Panel B reports correlations for the same sample, and Panel C shows summary statistics for the variables used in the return regressions (stock-day sample). Panel D shows summary statistics for the stock-day sample. Panels A, C, and D present separate statistics for the S&P 500 and the Russell 3000 universes. The samples range between January 2000 and December 2012.

### Panel A: Monthly Sample

S&P 500

| | N | Mean | Std Dev | Min | Median | Max |
|---|---|---|---|---|---|---|
| Daily stock volatility (%) | 46,984 | 2.03 | 1.22 | 0.61 | 1.70 | 10.80 |
| ETF ownership (%) | 46,984 | 2.09 | 1.41 | 0.05 | 1.73 | 9.36 |
| Index Fund ownership (%) | 46,984 | 6.08 | 1.90 | 0.44 | 5.92 | 11.80 |
| Active Fund ownership (%) | 46,984 | 18.10 | 6.19 | 0.80 | 18.10 | 34.20 |
| log(Mktcap ($m)) | 46,984 | 9.04 | 0.86 | 5.44 | 9.05 | 11.30 |
| 1/Price | 46,984 | 0.04 | 0.04 | 0.01 | 0.03 | 0.45 |
| Amihud | 46,984 | 0.00 | 0.00 | 0.00 | 0.00 | 0.03 |
| Bid-ask spread (%) | 46,984 | 0.17 | 0.24 | 0.02 | 0.08 | 1.86 |

Russell 3000

| | N | Mean | Std Dev | Min | Median | Max |
|---|---|---|---|---|---|---|
| Daily stock volatility (%) | 275,334 | 2.55 | 1.42 | 0.61 | 2.20 | 10.80 |
| ETF ownership (%) | 275,334 | 2.37 | 1.71 | 0.02 | 1.95 | 9.38 |
| Index Fund ownership (%) | 275,334 | 4.45 | 2.30 | 0.35 | 4.26 | 11.80 |
| Active Fund ownership (%) | 275,334 | 15.60 | 7.82 | 0.80 | 15.60 | 34.20 |
| log(Mktcap ($m)) | 275,334 | 6.97 | 1.34 | 4.32 | 6.77 | 11.30 |
| 1/Price | 275,334 | 0.07 | 0.06 | 0.01 | 0.05 | 0.45 |
| Amihud | 275,334 | 0.01 | 0.03 | 0.00 | 0.00 | 0.24 |
| Bid-ask spread (%) | 275,334 | 0.25 | 0.26 | 0.02 | 0.16 | 1.87 |

### Panel B: Correlations

| | | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|---|
| Daily stock volatility | (1) | 1.00 | | | | | | |
| ETF ownership | (2) | 0.01 | 1.00 | | | | | |
| Index Fund ownership | (3) | -0.06 | 0.31 | 1.00 | | | | |
| Active Fund ownership | (4) | -0.08 | 0.19 | 0.39 | 1.00 | | | |
| log(Mktcap) | (5) | -0.31 | -0.03 | 0.31 | 0.38 | 1.00 | | |
| 1/Price | (6) | 0.34 | -0.02 | -0.13 | -0.28 | -0.45 | 1.00 | |
| Amihud | (7) | 0.18 | -0.15 | -0.25 | -0.41 | -0.52 | 0.33 | 1.00 |
| Bid-ask spread | (8) | 0.23 | -0.36 | -0.26 | -0.32 | -0.39 | 0.29 | 0.54 |

## Table 3. Summary Statistics (Cont.)

### Panel C: Variables Used in Return Regressions (Daily Frequency)

S&P 500

| | N | Mean | Std Dev | Min | Median | Max |
|---|---|---|---|---|---|---|
| Ret(t) (%) | 1,242,412 | 0.06 | 2.14 | -9.46 | 0.02 | 10.40 |
| Ret(t+1,t+5) (%) | 1,242,412 | 0.24 | 4.57 | -19.92 | 0.24 | 21.33 |
| Ret(t+1,t+10) (%) | 1,242,412 | 0.45 | 6.20 | -23.82 | 0.50 | 25.24 |
| Ret(t+1,t+20) (%) | 1,242,412 | 0.87 | 8.70 | -31.35 | 1.05 | 33.64 |
| net(ETF Flows) (%) | 1,242,412 | 0.00 | 0.02 | -0.12 | 0.00 | 0.12 |

Russell 3000

| | N | Mean | Std Dev | Min | Median | Max |
|---|---|---|---|---|---|---|
| Ret(t) (%) | 6,643,402 | 0.07 | 2.47 | -9.46 | 0.00 | 10.41 |
| Ret(t+1,t+5) (%) | 6,643,402 | 0.19 | 5.23 | -19.92 | 0.16 | 21.33 |
| Ret(t+1,t+10) (%) | 6,643,402 | 0.38 | 7.05 | -23.83 | 0.38 | 25.25 |
| Ret(t+1,t+20) (%) | 6,643,402 | 0.74 | 9.95 | -31.35 | 0.80 | 33.64 |
| net(ETF Flows) (%) | 6,643,402 | 0.00 | 0.03 | -0.12 | 0.00 | 0.12 |

### Panel D: Daily Sample

S&P 500

| | N | Mean | Std Dev | Min | Median | Max |
|---|---|---|---|---|---|---|
| ETF ownership (%) | 1,097,681 | 2.28 | 1.36 | 0.03 | 2.13 | 9.77 |
| Abs(mispricing) (bps) | 1,092,266 | 0.25 | 0.28 | 0.03 | 0.16 | 2.39 |
| Net(mispricing) (bps) | 1,069,860 | -0.05 | 0.24 | -1.16 | 0.00 | 0.47 |
| Intraday volatility (%) | 1,097,681 | 0.02 | 0.01 | 0.00 | 0.01 | 0.12 |
| Variance Ratio (VR 15) | 1,067,110 | -2.04 | 1.04 | -8.23 | -1.82 | -0.55 |
| Share lending fee (%) | 1,097,681 | 0.22 | 1.21 | 0.00 | 0.10 | 81.23 |

Russell 3000

| | N | Mean | Std Dev | Min | Median | Max |
|---|---|---|---|---|---|---|
| ETF ownership (%) | 6,015,994 | 2.50 | 1.75 | 0.03 | 2.15 | 9.77 |
| Abs(mispricing) (%) | 5,777,219 | 0.30 | 0.34 | 0.03 | 0.20 | 2.39 |
| Net(mispricing) (%) | 5,652,821 | -0.10 | 0.32 | -1.16 | -0.02 | 0.47 |
| Intraday volatility (%) | 6,015,994 | 0.02 | 0.02 | 0.00 | 0.02 | 0.12 |
| Variance Ratio (VR 15) | 5,913,865 | -2.78 | 1.38 | -8.23 | -2.51 | -0.55 |
| Share lending fee (%) | 6,015,994 | 0.64 | 3.05 | 0.00 | 0.14 | 131.54 |

## Table 4. ETF Ownership and Stock Volatility

The table reports estimates from OLS regressions of daily volatility on ETF ownership 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 and volatility is computed using all daily returns within the month. The dependent variable as well as the ownership variables have been standardized by subtracting the mean and dividing by the standard deviation. Variable descriptions are provided in Appendix Table A1. 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. The sample ranges between January 2000 and December 2012.

| Dependent variable: | Daily stock volatility | | | | | |
|---|---|---|---|---|---|---|
| Sample: | S&P 500 | | | Russell 3000 | | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| ETF ownership | 0.150*** | 0.142*** | 0.093*** | 0.037*** | 0.027*** | 0.020*** |
| | (8.144) | (7.769) | (7.634) | (6.155) | (4.445) | (4.373) |
| log(Mktcap (t-1)) | -0.119*** | -0.127*** | -0.100*** | -0.151*** | -0.161*** | -0.128*** |
| | (-3.793) | (-4.143) | (-4.824) | (-12.313) | (-13.013) | (-13.432) |
| 1/Price (t-1) | 3.178*** | 3.093*** | 1.725*** | 1.453*** | 1.503*** | 1.064*** |
| | (6.016) | (5.933) | (4.878) | (10.227) | (10.629) | (9.775) |
| Amihud (t-1) | 53.319*** | 57.042*** | 25.716*** | 0.502** | 0.560*** | 0.118 |
| | (4.186) | (4.513) | (2.940) | (2.539) | (2.837) | (0.795) |
| Bid-ask spread (t-1) | 17.438*** | 18.421*** | 9.258** | 3.609 | 5.002** | 2.186 |
| | (2.779) | (2.969) | (2.276) | (1.604) | (2.247) | (1.302) |
| Index Fund Ownership | | 0.033*** | 0.021*** | | 0.028*** | 0.020*** |
| | | (2.854) | (2.807) | | (5.896) | (5.686) |
| Active Fund Ownership | | 0.033** | 0.021** | | 0.033*** | 0.024*** |
| | | (2.333) | (2.322) | | (5.787) | (5.736) |
| Volatility (t-1) | | | 0.360*** | | | 0.261*** |
| | | | (26.998) | | | (58.203) |
| Stock fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Month fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 46,984 | 46,984 | 46,984 | 275,334 | 275,334 | 275,334 |
| Adjusted $R^2$ | 0.615 | 0.615 | 0.664 | 0.549 | 0.550 | 0.580 |

## Table 5. Regression Discontinuity Design around the Russell Index Reconstitution

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 at the stock level. In Panel A, the dependent variable is ETF ownership. The explanatory variables are: a dummy for inclusion in the Russell 2000, for stocks in the Russell 1000 before index reconstitution (Columns (1), (3), (5), and (7)), and 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. 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. In Panel B, the dependent variable is daily stock volatility (computed using all daily returns within a month). The main explanatory variable is instrumented ETF ownership. The instruments are either a dummy for inclusion in the Russell 2000 for stocks in the Russell 1000 before reconstitution (Columns (1), (3), (5), and (7)) or a dummy for 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. For the two-stage estimation whose results are reported in Panels B and C, ETF ownership, as well as ownership by index and active funds, is standardized by subtracting the mean and dividing by the standard deviation in the estimation sample. In Panel D, the dependent variable is daily volatility. The main explanatory variable is an interaction between the dummy variables for index inclusion and average ETF ownership in the market. Other explanatory variables include interactions between the index inclusion dummies and ownership by index and active funds, a time trend. The dependent variable as well as the ownership variables have been standardized by subtracting the mean and dividing by the standard deviation. Variable descriptions are provided in Appendix Table A1. Standard errors are clustered at the month level. *t*-statistics are presented in parentheses. $^{***}$, $^{**}$, and $^{*}$ represent statistical significance at the 1%, 5%, and 10% levels, respectively. The sample ranges between January 2000 and December 2012.

## Panel A: First-Stage Regressions

| Dependent variable: | ETF Ownership | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Sample: | ± 50 stocks around cutoff | | ± 100 stocks around cutoff | | ± 150 stocks around cutoff | | ± 200 stocks around cutoff | |
| | R1000 | R2000 | R1000 | R2000 | R1000 | R2000 | R1000 | R2000 |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| In Russell 2000 | 0.126* | | 0.304*** | | 0.337*** | | 0.403*** | |
| | (1.811) | | (6.710) | | (7.214) | | (7.969) | |
| In Russell 1000 | | -0.663*** | | -0.379*** | | -0.495*** | | -0.546*** |
| | | (-9.454) | | (-6.474) | | (-10.391) | | (-11.320) |
| log(Mktcap (t-1)) | -0.639*** | -0.485*** | -0.599*** | -0.307*** | -0.585*** | -0.204*** | -0.562*** | -0.233*** |
| | (-6.167) | (-7.121) | (-7.419) | (-6.812) | (-8.066) | (-4.648) | (-8.249) | (-5.326) |
| 1/Price (t-1) | -1.912*** | -1.960*** | -1.724*** | -2.009*** | -2.018*** | -1.716*** | -1.700*** | -1.620*** |
| | (-7.908) | (-5.537) | (-6.651) | (-6.877) | (-8.304) | (-7.104) | (-7.168) | (-8.333) |
| Amihud (t-1) | -0.242 | -17.580*** | -1.956 | -0.954* | -2.027 | -1.627** | -3.971** | -3.017*** |
| | (-0.280) | (-6.737) | (-1.519) | (-1.854) | (-1.580) | (-2.390) | (-2.500) | (-3.258) |
| Bid-ask spread (t-1) | -13.830*** | 15.155*** | -9.355*** | 6.818*** | -9.238*** | 0.931 | -7.839*** | 1.646 |
| | (-3.468) | (5.568) | (-3.313) | (3.293) | (-3.693) | (0.441) | (-3.874) | (0.779) |
| Index fund ownership | 0.228*** | 0.288*** | 0.293*** | 0.300*** | 0.335*** | 0.371*** | 0.347*** | 0.371*** |
| | (8.426) | (7.839) | (13.329) | (10.108) | (15.866) | (11.357) | (17.395) | (11.174) |
| Active fund ownership | -0.110*** | 0.010 | -0.090*** | 0.040** | -0.060*** | 0.017 | -0.026* | -0.000 |
| | (-3.596) | (0.445) | (-3.579) | (2.353) | (-4.283) | (1.647) | (-1.859) | (-0.003) |
| 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 |
| Adjusted $R^2$ | 0.477 | 0.697 | 0.496 | 0.656 | 0.513 | 0.643 | 0.517 | 0.629 |

## Table 5. Regression Discontinuity Design around the Russell Index Reconstitution (Cont.)

### Panel B: Second-Stage Regressions, First Degree Polynomial

| Polynomial: | Linear specification | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Dependent variable: | Daily stock volatility | | | | | | | |
| Sample: | ± 50 stocks around cutoff | | ± 100 stocks around cutoff | | ± 150 stocks around cutoff | | ± 200 stocks around cutoff | |
| Instrument: | R1000 | R2000 | R1000 | R2000 | R1000 | R2000 | R1000 | R2000 |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| ETF ownership (instrumented) | 1.746 | 0.356*** | 0.630*** | 0.601*** | 0.812*** | 0.408*** | 0.722*** | 0.322*** |
| | (1.580) | (3.605) | (2.934) | (3.770) | (3.744) | (4.091) | (3.834) | (4.127) |
| log(Mktcap (t-1)) | -0.045 | -0.614*** | -0.418*** | -0.696*** | -0.426*** | -0.694*** | -0.518*** | -0.770*** |
| | (-0.110) | (-7.795) | (-4.054) | (-8.670) | (-4.368) | (-9.039) | (-6.166) | (-10.483) |
| 1/Price (t-1) | 4.039*** | 2.589*** | 2.850*** | 2.648*** | 3.541*** | 2.935*** | 3.170*** | 2.661*** |
| | (2.924) | (8.076) | (9.325) | (9.708) | (10.595) | (11.868) | (12.886) | (14.892) |
| Amihud (t-1) | -2.901*** | 11.792*** | -1.831** | 2.170*** | -2.476*** | 1.704** | -1.164 | 2.106*** |
| | (-2.898) | (3.312) | (-2.228) | (3.204) | (-3.074) | (2.538) | (-1.313) | (3.249) |
| Bid-ask spread (t-1) | 10.323 | -15.770*** | -0.181 | -13.486*** | 1.214 | -11.209*** | 2.053 | -12.625*** |
| | (1.067) | (-5.278) | (-0.047) | (-5.245) | (0.345) | (-3.344) | (0.782) | (-4.306) |
| Index fund ownership | -0.169 | 0.017 | -0.014 | -0.031 | -0.067* | -0.013 | -0.067* | 0.002 |
| | (-1.159) | (0.774) | (-0.381) | (-1.165) | (-1.656) | (-0.657) | (-1.912) | (0.121) |
| Active fund ownership | 0.152* | 0.132*** | 0.048** | 0.099*** | 0.054*** | 0.103*** | 0.048*** | 0.111*** |
| | (1.787) | (7.460) | (2.417) | (6.958) | (3.631) | (8.965) | (4.381) | (10.283) |
| Month fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 3,705 | 4,801 | 7,249 | 9,851 | 11,069 | 14,933 | 15,188 | 19,603 |

## Table 5. Regression Discontinuity Design around the Russell Index Reconstitution (Cont.)

### Panel C: Second-Stage Regressions, Second Degree Polynomial

| Polynomial: | Quadratic specification | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Dependent variable: | Daily stock volatility | | | | | | | |
| Sample: | ± 50 stocks around cutoff | | ± 100 stocks around cutoff | | ± 150 stocks around cutoff | | ± 200 stocks around cutoff | |
| Instrument: | R1000 | R2000 | R1000 | R2000 | R1000 | R2000 | R1000 | R2000 |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| ETF ownership (instrumented) | 1.495 | 0.463*** | 0.652 | 0.433*** | 0.636*** | 0.479*** | 0.660*** | 0.366*** |
| | (0.831) | (3.906) | (1.275) | (3.926) | (2.976) | (4.607) | (3.314) | (4.115) |
| log(Mktcap (t-1)) | -0.108 | -0.581*** | -0.410** | -0.720*** | -0.484*** | -0.687*** | -0.538*** | -0.764*** |
| | (-0.158) | (-6.785) | (-2.260) | (-9.257) | (-4.857) | (-8.829) | (-6.025) | (-10.371) |
| 1/Price (t-1) | 3.762 | 2.658*** | 2.872*** | 2.465*** | 3.305*** | 2.997*** | 3.078*** | 2.697*** |
| | (1.625) | (8.097) | (4.903) | (9.789) | (9.935) | (11.835) | (12.207) | (14.823) |
| Amihud (t-1) | -2.927*** | 12.883*** | -1.832** | 2.119*** | -2.640*** | 1.811*** | -1.267 | 2.209*** |
| | (-3.108) | (3.627) | (-2.031) | (3.226) | (-3.545) | (2.628) | (-1.451) | (3.275) |
| Bid-ask spread (t-1) | 11.595 | -16.302*** | -0.301 | -12.969*** | 0.570 | -11.131*** | 2.267 | -12.700*** |
| | (0.745) | (-5.385) | (-0.073) | (-5.391) | (0.171) | (-3.265) | (0.875) | (-4.284) |
| Index fund ownership | -0.135 | -0.002 | -0.018 | -0.002 | -0.033 | -0.028 | -0.053 | -0.006 |
| | (-0.558) | (-0.083) | (-0.216) | (-0.099) | (-0.809) | (-1.337) | (-1.377) | (-0.310) |
| Active fund ownership | 0.120 | 0.131*** | 0.050* | 0.103*** | 0.048*** | 0.103*** | 0.046*** | 0.111*** |
| | (1.102) | (7.087) | (1.666) | (7.355) | (3.402) | (8.848) | (4.234) | (10.257) |
| Month fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 3,705 | 4,801 | 7,249 | 9,851 | 11,069 | 14,933 | 15,188 | 19,603 |

## Table 5. Regression Discontinuity Design around the Russell Index Reconstitution (Cont.)

## Panel D: Interaction with Average ETF Ownership

| Dependent variable: | Daily stock volatility | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Sample: | ± 50 stocks around cutoff | | ± 100 stocks around cutoff | | ± 150 stocks around cutoff | | ± 200 stocks around cutoff | |
| | R1000 | R2000 | R1000 | R2000 | R1000 | R2000 | R1000 | R2000 |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| In Russell 2000 | -1.449 | | 1.326 | | 4.425*** | | 6.535*** | |
| | (-0.490) | | (0.870) | | (2.877) | | (4.136) | |
| × Average ETF ownership in the market | -0.235 | | 0.090 | | 0.265** | | 0.396*** | |
| | (-0.985) | | (0.717) | | (2.038) | | (2.983) | |
| × Average Index Fund ownership in the market | -0.058** | | -0.051*** | | -0.036*** | | -0.036** | |
| | (-2.251) | | (-3.381) | | (-2.672) | | (-2.490) | |
| × Average Active Fund ownership in the market | 0.026 | | 0.011 | | -0.059*** | | -0.087*** | |
| | (0.444) | | (0.439) | | (-2.672) | | (-3.216) | |
| × Trend | 0.003 | | -0.002 | | -0.006* | | -0.009*** | |
| | (0.556) | | (-0.692) | | (-1.923) | | (-2.872) | |
| In Russell 1000 | | -8.479*** | | -3.261** | | -3.825*** | | -4.277*** |
| | | (-2.883) | | (-2.529) | | (-3.662) | | (-4.495) |
| × Average ETF ownership in the market | | -0.526** | | -0.198* | | -0.217** | | -0.209** |
| | | (-2.548) | | (-1.938) | | (-2.417) | | (-2.501) |
| × Average Index Fund ownership in the market | | -0.009 | | 0.023 | | 0.004 | | -0.018 |
| | | (-0.255) | | (1.347) | | (0.320) | | (-1.514) |
| × Average Active Fund ownership in the market | | 0.188*** | | 0.038 | | 0.050** | | 0.074*** |
| | | (2.634) | | (1.187) | | (2.475) | | (3.842) |
| × Trend | | 0.010** | | 0.005* | | 0.006** | | 0.006*** |
| | | (2.144) | | (1.851) | | (2.572) | | (2.897) |
| Average ETF ownership in the market | 1.696*** | 0.764*** | 1.063*** | 0.955*** | 0.873*** | 1.084*** | 0.882*** | 0.969*** |
| | (12.835) | (11.661) | (9.621) | (14.541) | (14.598) | (16.825) | (10.015) | (15.376) |
| Average Index Fund ownership in the market | -0.058 | -0.231*** | -0.146*** | -0.118*** | -0.147*** | -0.077*** | -0.134*** | -0.100*** |
| | (-1.487) | (-25.007) | (-8.652) | (-11.943) | (-12.930) | (-5.270) | (-15.098) | (-8.486) |
| Average Active Fund ownership in the market | -0.385*** | -0.188*** | -0.259*** | -0.260*** | -0.166*** | -0.316*** | -0.173*** | -0.261*** |
| | (-5.676) | (-15.323) | (-10.922) | (-19.952) | (-10.382) | (-18.742) | (-11.738) | (-15.949) |
| Trend | -0.040*** | -0.014*** | -0.026*** | -0.021*** | -0.021*** | -0.024*** | -0.021*** | -0.022*** |
| | (-13.707) | (-7.167) | (-10.549) | (-11.357) | (-16.351) | (-13.378) | (-10.766) | (-12.307) |
| log(Mktcap (t-1)) | -0.742*** | -0.807*** | -0.661*** | -0.869*** | -0.722*** | -0.805*** | -0.741*** | -0.859*** |
| | (-8.955) | (-9.139) | (-8.447) | (-12.331) | (-9.810) | (-11.498) | (-10.531) | (-12.774) |
| 1/Price (t-1) | 1.850*** | 2.372*** | 2.198*** | 2.155*** | 2.567*** | 2.802*** | 2.401*** | 2.562*** |
| | (7.162) | (7.056) | (9.499) | (8.906) | (12.514) | (11.740) | (14.537) | (12.457) |
| Amihud (t-1) | -2.409*** | 5.778** | -2.106*** | 0.773 | -3.017*** | 0.207 | -2.408*** | 0.491 |
| | (-3.045) | (2.520) | (-3.130) | (1.322) | (-4.371) | (0.349) | (-3.782) | (0.873) |
| Bid-ask spread (t-1) | -5.685 | -11.804*** | -4.191 | -11.372*** | -3.225 | -12.642*** | -2.022 | -13.501*** |
| | (-1.324) | (-5.791) | (-1.100) | (-4.908) | (-1.052) | (-4.142) | (-0.858) | (-4.959) |
| Month fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 3,725 | 4,838 | 7,293 | 9,907 | 11,138 | 15,018 | 15,282 | 19,715 |
| Adjusted $R^2$ | 0.617 | 0.457 | 0.579 | 0.469 | 0.577 | 0.440 | 0.556 | 0.456 |

## Table 6. 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. VR 15 seconds is the absolute value of the ratio of the variance of 15-second log returns on day *t* and 3 times the variance of 5-second log returns on day *t*, minus 1, using data from the TAQ database and averaging the numerator and denominator within a month. VR 5 days is the absolute value of the ratio of the variance of 5-day returns in a given quarter on and 5 times the variance of one-day returns in the same quarter, minus 1. Panel A presents OLS regressions. Panels B and C show RDD regressions for 15-seconds variance ratio (Panel B) and for 5-days variance ratio (Panel C) based on the Russell 1000/Russell 2000 inclusion experiment. The dependent variable as well as the ownership variables have been standardized by subtracting the mean and dividing by the standard deviation. Variable descriptions are provided in Appendix Table A1. Standard errors are clustered at the stock level (Panel A) and the time level (Panels C and D). *t*-statistics are presented in parentheses. ***, **, and * represent statistical significance at the 1%, 5%, and 10% levels, respectively. The sample ranges between January 2000 and December 2012.

### Panel A: Variance Ratios (OLS)

| Sample: | S&P 500 | | Russell 3000 | |
|---|---|---|---|---|
| Dependent Variable: | VR 15 seconds | VR 5 days | VR 15 seconds | VR 5 days |
| | (1) | (2) | (3) | (4) |
| ETF ownership | 0.101*** | 0.047** | 0.040*** | 0.013* |
| | (4.563) | (2.140) | (5.966) | (1.889) |
| log(Mktcap (t-1)) | 0.135*** | 0.068** | 0.160*** | -0.011 |
| | (4.421) | (2.271) | (13.151) | (-0.976) |
| 1/Price (t-1) | 2.999*** | 0.342 | 0.985*** | -0.030 |
| | (5.631) | (0.733) | (8.259) | (-0.248) |
| Amihud (t-1) | -105.619*** | 47.411*** | -1.569*** | 1.708*** |
| | (-5.601) | (3.360) | (-10.330) | (9.466) |
| Bid-ask spread (t-1) | -1.521 | -7.226** | -24.620*** | 2.144 |
| | (-0.409) | (-2.181) | (-14.580) | (1.311) |
| Index Fund Ownership | 0.030** | -0.006 | 0.018*** | -0.003 |
| | (2.105) | (-0.441) | (3.416) | (-0.466) |
| Active Fund Ownership | 0.035** | -0.009 | 0.062*** | -0.041*** |
| | (2.422) | (-0.538) | (10.234) | (-6.524) |
| Stock fixed effects | Yes | Yes | Yes | Yes |
| Month fixed effects | Yes | Yes | Yes | Yes |
| Observations | 48,567 | 19,991 | 298,190 | 112,532 |
| Adjusted $R^2$ | 0.466 | 0.031 | 0.455 | 0.042 |

## Table 6. ETF Ownership and Price Efficiency: Variance Ratios (Cont.)

### Panel B: Variance Ratios – 15 Seconds (RDD)

| Dependent variable: | Variance Ratio 5 days | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Sample: | ± 50 stocks around cutoff | | ± 100 stocks around cutoff | | ± 150 stocks around cutoff | | ± 200 stocks around cutoff | |
| | R1000 | R2000 | R1000 | R2000 | R1000 | R2000 | R1000 | R2000 |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| In Russell 2000 | 0.970 | | 0.231 | | 0.343* | | 0.341** | |
| | (0.975) | | (1.149) | | (1.917) | | (2.014) | |
| In Russell 1000 | | 0.305 | | 0.576** | | 0.479*** | | 0.595*** |
| | | (1.658) | | (2.356) | | (2.803) | | (3.549) |
| log(Mktcap (t-1)) | 0.345 | 0.176 | 0.115 | 0.265*** | 0.074 | 0.159** | 0.055 | 0.111* |
| | (1.122) | (1.276) | (1.110) | (2.874) | (0.934) | (2.531) | (0.783) | (1.840) |
| 1/Price (t-1) | 1.111 | -0.163 | 0.275 | 0.453 | 0.291 | 0.134 | 0.099 | 0.210 |
| | (0.923) | (-0.209) | (0.743) | (1.061) | (0.747) | (0.385) | (0.354) | (0.708) |
| Amihud (t-1) | -1.203 | 4.257 | -0.885 | 0.411 | -0.555 | 0.993 | -0.424 | 1.969 |
| | (-0.632) | (1.111) | (-0.592) | (0.358) | (-0.368) | (0.797) | (-0.277) | (1.326) |
| Bid-ask spread (t-1) | 4.163 | 6.028 | -3.688 | 3.513 | -1.570 | 8.246*** | 4.421 | 6.525*** |
| | (0.354) | (1.268) | (-0.701) | (1.250) | (-0.373) | (3.928) | (1.524) | (2.789) |
| Index fund ownership | -0.140 | -0.079 | -0.051 | -0.148*** | -0.095** | -0.140*** | -0.101** | -0.161*** |
| | (-1.040) | (-1.446) | (-1.274) | (-3.224) | (-2.420) | (-3.726) | (-2.594) | (-4.752) |
| Active fund ownership | -0.018 | -0.059** | -0.070*** | -0.049* | -0.037 | -0.036* | -0.059*** | -0.035* |
| | (-0.214) | (-2.175) | (-2.835) | (-1.974) | (-1.582) | (-1.824) | (-3.224) | (-1.999) |
| Month fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 1,126 | 1,458 | 2,204 | 2,982 | 3,364 | 4,529 | 4,626 | 5,938 |

## Table 6. ETF Ownership and Price Efficiency: Variance Ratios (Cont.)

### Panel C: Variance Ratios – 5 Days (RDD)

| Dependent variable: | Variance Ratio 5 days | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 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 | 0.970 | | 0.231 | | 0.343* | | 0.341** | |
| | (0.975) | | (1.149) | | (1.917) | | (2.014) | |
| In Russell 1000 | | 0.305 | | 0.576** | | 0.479*** | | 0.595*** |
| | | (1.658) | | (2.356) | | (2.803) | | (3.549) |
| log(Mktcap (t-1)) | 0.345 | 0.176 | 0.115 | 0.265*** | 0.074 | 0.159** | 0.055 | 0.111* |
| | (1.122) | (1.276) | (1.110) | (2.874) | (0.934) | (2.531) | (0.783) | (1.840) |
| 1/Price (t-1) | 1.111 | -0.163 | 0.275 | 0.453 | 0.291 | 0.134 | 0.099 | 0.210 |
| | (0.923) | (-0.209) | (0.743) | (1.061) | (0.747) | (0.385) | (0.354) | (0.708) |
| Amihud (t-1) | -1.203 | 4.257 | -0.885 | 0.411 | -0.555 | 0.993 | -0.424 | 1.969 |
| | (-0.632) | (1.111) | (-0.592) | (0.358) | (-0.368) | (0.797) | (-0.277) | (1.326) |
| Bid-ask spread (t-1) | 4.163 | 6.028 | -3.688 | 3.513 | -1.570 | 8.246*** | 4.421 | 6.525*** |
| | (0.354) | (1.268) | (-0.701) | (1.250) | (-0.373) | (3.928) | (1.524) | (2.789) |
| Index fund ownership | -0.140 | -0.079 | -0.051 | -0.148*** | -0.095** | -0.140*** | -0.101** | -0.161*** |
| | (-1.040) | (-1.446) | (-1.274) | (-3.224) | (-2.420) | (-3.726) | (-2.594) | (-4.752) |
| Active fund ownership | -0.018 | -0.059** | -0.070*** | -0.049* | -0.037 | -0.036* | -0.059*** | -0.035* |
| | (-0.214) | (-2.175) | (-2.835) | (-1.974) | (-1.582) | (-1.824) | (-3.224) | (-1.999) |
| Month fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 1,126 | 1,458 | 2,204 | 2,982 | 3,364 | 4,529 | 4,626 | 5,938 |

## Table 7. Price Reversals

The table reports estimates from OLS regressions of one- and multiday returns on ETF flows and controls. The specifications also include the *k*-period lagged dependent variable, where *k* is set to have the return-measurement horizon end in *t*-1. 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 percentages. Flows have been standardized by subtracting the mean and dividing by the standard deviation. Variable descriptions are provided in Appendix Table A1. Standard errors are clustered at the day level and are computed using the Newey and West (1987) estimator. *t*-statistics are presented in parentheses. ***, **, and * represent statistical significance at the 1%, 5%, and 10% levels, respectively. The sample ranges between January 2000 and December 2012.

| 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) |
| net(ETF Flows) | 0.175*** | -0.028 | -0.060** | -0.083*** | 0.066*** | -0.007 | 0.002 | -0.031* |
| | (17.763) | (-1.579) | (-2.517) | (-2.642) | (12.458) | (-0.715) | (0.138) | (-1.954) |
| log(Mktcap (t-1)) | 0.027*** | -0.032*** | -0.062*** | -0.118*** | 0.016*** | 0.013** | 0.030*** | 0.064*** |
| | (7.378) | (-3.870) | (-5.476) | (-7.421) | (5.042) | (2.007) | (3.498) | (5.734) |
| 1/Price (t-1) | -0.890*** | 0.901*** | 1.809*** | 4.751*** | -0.670*** | -0.152 | -0.286** | 0.364* |
| | (-7.887) | (3.689) | (5.640) | (10.116) | (-14.942) | (-1.546) | (-2.181) | (1.825) |
| Amihud (t-1) | 35.841*** | -20.719* | -47.955*** | -53.526** | 0.308*** | -1.359*** | -2.498*** | -4.255*** |
| | (6.472) | (-1.858) | (-3.112) | (-2.406) | (3.004) | (-6.432) | (-9.186) | (-11.275) |
| Bid-ask spread (t-1) | 4.526*** | 8.695*** | 21.027*** | 40.012*** | 4.954*** | 1.089 | 4.127 | 9.572** |
| | (3.098) | (3.257) | (5.914) | (8.174) | (4.489) | (0.467) | (1.401) | (2.085) |
| Lagged dep. variable | -2.084*** | -1.753*** | -1.884*** | -1.972*** | -2.290*** | -1.455*** | -1.605*** | -1.431*** |
| | (-6.442) | (-6.192) | (-7.174) | (-7.705) | (-13.324) | (-9.973) | (-11.993) | (-10.796) |
| Day fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 1,242,568 | 1,242,568 | 1,242,568 | 1,242,568 | 6,643,402 | 6,643,402 | 6,643,402 | 6,643,402 |
| Adjusted $R^2$ | 0.327 | 0.302 | 0.281 | 0.279 | 0.283 | 0.252 | 0.226 | 0.223 |

## Table 8. Evidence on the Arbitrage Channel

The table reports estimates from OLS regressions of intraday volatility (Panel A) and intraday variance ratio (Panel B) on absolute stock-level mispricing in the prior period interacted with measures of arbitrage costs. The frequency is daily and the observations are at the stock level. The sample includes S&P 500 stocks. In Columns (2)-(4), arbitrage cost is captured by the bid-ask spread in the prior day, and in Columns (5)-(7), by the average share-lending fee in the month. For both measures of arbitrage costs, we construct dummy variables denoting whether the stock is in the top half of the distribution of that measure in the relevant period. In Columns (3) and (6), we restrict the sample to observations for which the stock-level mispricing is positive. In Columns (4) and (7), we restrict the sample to observations for which the stock-level mispricing is negative. Variable descriptions are provided in Appendix Table A1. 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. The sample ranges between January 2000 and December 2012.

### Panel A: Intraday Volatility

| Dependent variable: | Intraday stock volatility | | | | | | |
|---|---|---|---|---|---|---|---|
| | All | All | Misp > 0 | Misp < 0 | All | Misp > 0 | Misp < 0 |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Abs(Mispricing) (t-1) | 0.023*** | 0.042*** | 0.072*** | 0.033*** | 0.020*** | 0.047*** | 0.017*** |
| | (7.077) | (8.640) | (10.290) | (7.086) | (4.743) | (7.545) | (4.375) |
| × I(High bid-ask spread) | | -0.036*** | -0.035*** | -0.029*** | | | |
| | | (-5.507) | (-3.372) | (-5.051) | | | |
| × I(High lending fee) | | | | | -0.008* | 0.006 | -0.010*** |
| | | | | | (-1.798) | (0.823) | (-2.619) |
| High bid-ask spread | | 0.032*** | 0.034*** | 0.031*** | | | |
| | | (5.805) | (5.802) | (5.452) | | | |
| High lending fee | | | | | -0.002 | 0.001 | -0.002 |
| | | | | | (-0.557) | (0.167) | (-0.669) |
| ETF ownership (t-1) | 0.022*** | 0.023*** | 0.009 | 0.024*** | 0.021*** | -0.006 | 0.030*** |
| | (3.616) | (3.833) | (1.386) | (3.748) | (4.189) | (-1.287) | (5.594) |
| log(Mktcap (t-1)) | 0.054*** | 0.066*** | 0.076*** | 0.059*** | 0.033*** | 0.034*** | 0.032*** |
| | (2.708) | (3.254) | (3.702) | (2.787) | (5.146) | (5.356) | (4.880) |
| 1/Price (t-1) | 6.630*** | 6.687*** | 6.461*** | 6.902*** | 4.986*** | 4.740*** | 5.211*** |
| | (10.488) | (10.507) | (10.057) | (10.656) | (13.320) | (12.515) | (13.917) |
| Amihud (t-1) | -18.192 | -13.541 | -3.129 | -21.976 | -17.591 | -12.130 | -22.122 |
| | (-0.878) | (-0.662) | (-0.148) | (-1.093) | (-0.880) | (-0.597) | (-1.111) |
| Bid-ask spread (t-1) | 12.231** | 8.724 | 7.600 | 11.951* | 16.098*** | 15.318*** | 19.825*** |
| | (2.066) | (1.480) | (1.326) | (1.932) | (2.724) | (2.748) | (3.119) |
| Ret (t-1) | -0.158*** | -0.166*** | -0.630*** | 0.314*** | -0.239*** | -0.709*** | 0.263*** |
| | (-3.422) | (-3.598) | (-9.619) | (4.179) | (-4.870) | (-9.876) | (3.303) |
| Dependent variable (t-1) | 0.417*** | 0.416*** | 0.426*** | 0.400*** | 0.571*** | 0.583*** | 0.553*** |
| | (21.935) | (21.765) | (21.694) | (21.259) | (27.003) | (26.132) | (27.148) |
| Abs(Mispricing) (t-2) | 0.024*** | 0.026*** | 0.055*** | 0.015*** | 0.017*** | 0.057*** | 0.000 |
| | (7.186) | (7.763) | (10.299) | (4.652) | (4.591) | (9.834) | (0.080) |
| Day fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Stock fixed effects | Yes | No | No | No | No | No | No |
| Observations | 1,090,370 | 1,090,370 | 543,895 | 546,475 | 1,090,370 | 543,895 | 546,475 |
| Adjusted $R^2$ | 0.576 | 0.577 | 0.578 | 0.577 | 0.530 | 0.534 | 0.529 |

## Table 8. Evidence on the Arbitrage Channel (Cont.)

### Panel B: Intraday Variance Ratio

| Dependent variable: | Intraday variance ratio (VR 15) | | | | | | |
|---|---|---|---|---|---|---|---|
| | All | All | Misp > 0 | Misp < 0 | All | Misp > 0 | Misp < 0 |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Abs(Mispricing) (t-1) | 0.002 | 0.017*** | 0.035*** | 0.012** | 0.005 | 0.025*** | 0.005 |
| | (0.767) | (3.566) | (5.680) | (2.474) | (1.068) | (3.875) | (0.946) |
| × I(High bid-ask spread) | | -0.028*** | -0.035*** | -0.022*** | | | |
| | | (-4.323) | (-4.206) | (-3.541) | | | |
| × I(High lending fee) | | | | | -0.010* | 0.001 | -0.011** |
| | | | | | (-1.894) | (0.167) | (-2.030) |
| High bid-ask spread | | 0.035*** | 0.035*** | 0.034*** | | | |
| | | (5.142) | (4.901) | (4.910) | | | |
| High lending fee | | | | | -0.010** | -0.007 | -0.012*** |
| | | | | | (-2.486) | (-1.647) | (-2.680) |
| ETF ownership (t-1) | 0.039*** | 0.041*** | 0.032*** | 0.041*** | 0.028*** | 0.009 | 0.033*** |
| | (5.338) | (5.485) | (3.995) | (5.379) | (4.092) | (1.257) | (4.597) |
| log(Mktcap (t-1)) | 0.009 | 0.022 | 0.029 | 0.017 | 0.147*** | 0.147*** | 0.146*** |
| | (0.504) | (1.187) | (1.548) | (0.867) | (18.171) | (17.927) | (18.072) |
| 1/Price (t-1) | 1.869*** | 1.915*** | 1.964*** | 1.852*** | 4.694*** | 4.647*** | 4.689*** |
| | (6.874) | (7.072) | (7.328) | (6.643) | (13.077) | (12.071) | (13.865) |
| Amihud (t-1) | -75.300*** | -70.687*** | -68.454*** | -71.723*** | -94.458*** | -93.626*** | -94.543*** |
| | (-4.937) | (-4.721) | (-4.349) | (-4.965) | (-3.634) | (-3.577) | (-3.678) |
| Bid-ask spread (t-1) | 1.692 | -1.900 | -2.050 | -0.888 | 13.193** | 14.489** | 13.441** |
| | (0.504) | (-0.573) | (-0.611) | (-0.254) | (2.030) | (2.341) | (1.965) |
| Ret (t-1) | -0.027 | -0.036 | 0.015 | -0.068 | -0.030 | 0.040 | -0.067 |
| | (-0.821) | (-1.103) | (0.278) | (-1.442) | (-0.842) | (0.652) | (-1.176) |
| Dependent variable (t-1) | 0.139*** | 0.139*** | 0.139*** | 0.137*** | 0.240*** | 0.242*** | 0.236*** |
| | (42.952) | (43.139) | (39.920) | (40.767) | (30.071) | (29.473) | (29.824) |
| Abs(Mispricing) (t-2) | -0.002 | -0.000 | 0.015*** | -0.006* | -0.004 | 0.030*** | -0.019*** |
| | (-0.631) | (-0.114) | (3.439) | (-1.833) | (-1.108) | (5.590) | (-4.577) |
| Day fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Stock fixed effects | Yes | No | No | No | No | No | No |
| Observations | 1,050,652 | 1,050,652 | 524,652 | 526,000 | 1,050,652 | 524,652 | 526,000 |
| Adjusted $R^2$ | 0.246 | 0.246 | 0.249 | 0.245 | 0.180 | 0.182 | 0.179 |

## Table 9. Volatility and ETF Ownership in the Time-Series

Panel A reports estimates from a time-series regression at the monthly frequency of average daily volatility in a given month across the stocks in the CRSP universe on lagged average ETF ownership for the same universe. The controls include lagged average volatility, lagged average index and active fund ownership (IF and AF variables respectively), and a time trend. The same regression is performed in first differences, excluding the time trend. Panel B reports estimates from time-series regressions of average volatility in each quintile of ETF ownership on lagged average ETF ownership across all stocks in the CRSP universe. Panel C reports estimates from time-series regressions of the changes in the average volatility in each quintile of ETF ownership on lagged changes in the average ETF ownership across all stocks in the CRSP universe. The other ownership variables are also computed as averages across all stocks. The dependent as well as the ownership variables have been standardized by subtracting the mean and dividing by the standard deviation. Variable descriptions are provided in Appendix Table A1. *t*-statistics are presented in parentheses. ***, **, and * represent statistical significance at the 1%, 5%, and 10% levels, respectively. The sample ranges between January 2000 and December 2012.

### Panel A: Regression of Volatility on ETF Ownership

| Dependent Variable: | Volatility (t+1) | ΔVolatility (t+1) |
|---|---|---|
| | (1) | (2) |
| ETF own (t) | 0.216*** | |
| | (3.938) | |
| IF own (t) | -0.066 | |
| | (-1.106) | |
| AF own (t) | 0.082 | |
| | (1.629) | |
| Volatility (t) | 0.701*** | |
| | (12.939) | |
| Trend | -0.001 | |
| | (-0.483) | |
| ΔETF own (t) | | 0.344*** |
| | | (4.388) |
| ΔIF own (t) | | -0.052 |
| | | (-0.632) |
| ΔAF own (t) | | 0.030 |
| | | (0.358) |
| ΔVolatility (t) | | -0.150* |
| | | (-1.934) |
| Observations | 149 | 148 |
| $R^2$ | 0.721 | 0.151 |

**Table 9. Volatility and ETF Ownership in the Time-Series (Cont.)**

**Panel B: Regression of Volatility and ETF Ownership, by Quintiles of ETF Ownership**

| Dependent Variable: | Volatility (t+1) | | | | |
|---|---|---|---|---|---|
| Quintile of ETF ownership: | Smallest | 2 | 3 | 4 | Largest |
| ETF ownership (t) | 0.189*** | 0.213*** | 0.206*** | 0.216*** | 0.247*** |
| | (3.532) | (4.075) | (3.890) | (4.154) | (3.856) |
| Index mutual funds ownership (t) | -0.063 | -0.068 | -0.049 | -0.049 | -0.093 |
| | (-1.062) | (-1.144) | (-0.823) | (-0.863) | (-1.381) |
| Active mutual funds ownership (t) | 0.025 | 0.060 | 0.091* | 0.089* | 0.150** |
| | (0.497) | (1.193) | (1.799) | (1.823) | (2.560) |
| Volatility (t) | 0.760*** | 0.688*** | 0.656*** | 0.638*** | 0.667*** |
| | (15.050) | (12.508) | (11.261) | (10.779) | (11.548) |
| Trend | -0.000 | -0.000 | -0.001 | -0.000 | -0.002 |
| | (-0.027) | (-0.238) | (-0.606) | (-0.183) | (-1.216) |
| Observations | 149 | 149 | 149 | 149 | 149 |
| $R^2$ | 0.759 | 0.691 | 0.648 | 0.662 | 0.726 |

**Panel C: Regression of Changes in Volatility and ETF Ownership, by Quintiles of ETF Ownership**

| Dependent Variable: | ΔVolatility (t+1) | | | | |
|---|---|---|---|---|---|
| Quintile of ETF ownership: | Smallest | 2 | 3 | 4 | Largest |
| ΔETF ownership (t) | 0.297*** | 0.289*** | 0.326*** | 0.312*** | 0.384*** |
| | (3.886) | (3.654) | (4.161) | (4.189) | (4.482) |
| ΔIndex mutual funds ownership (t) | -0.049 | -0.064 | -0.040 | -0.061 | -0.030 |
| | (-0.607) | (-0.777) | (-0.479) | (-0.764) | (-0.326) |
| ΔActive mutual funds ownership (t) | -0.016 | 0.009 | 0.053 | 0.044 | 0.051 |
| | (-0.195) | (0.114) | (0.638) | (0.549) | (0.551) |
| ΔVolatility (t) | -0.098 | -0.149* | -0.184** | -0.200** | -0.197** |
| | (-1.219) | (-1.859) | (-2.356) | (-2.596) | (-2.540) |
| Observations | 148 | 148 | 148 | 148 | 148 |
| $R^2$ | 0.120 | 0.128 | 0.157 | 0.159 | 0.150 |

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

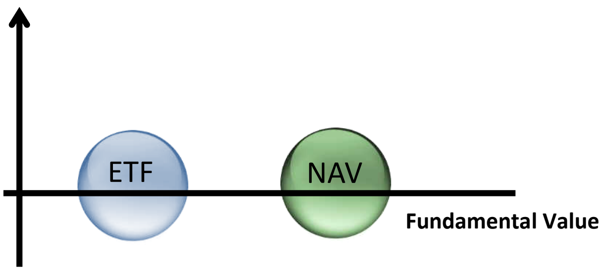



**Figure 1a.** Initial equilibrium

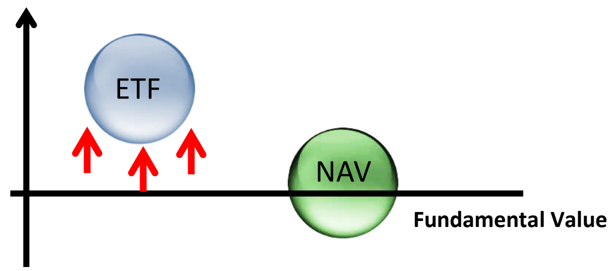



**Figure 1b.** Non-fundamental shock to ETF

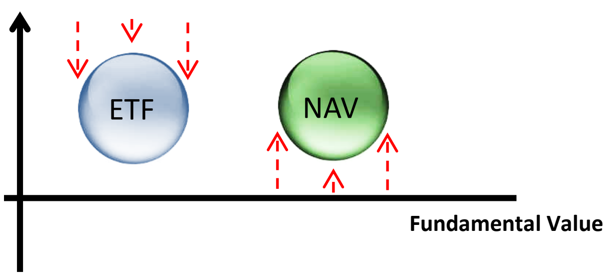



**Figure 1c.** Initial outcome of arbitrage: the non-fundamental shock is propagated to the NAV, and the ETF price starts reverting to the fundamental value.

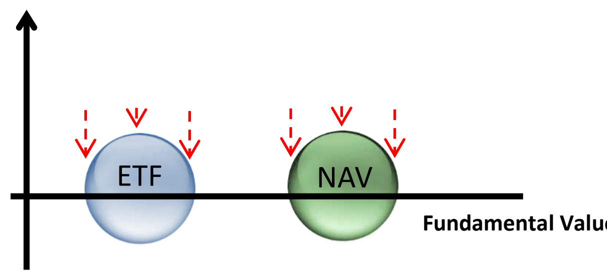



**Figure 1d.** Re-establishment of equilibrium: after some time, both the ETF price and the NAV revert to the fundamental value.

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

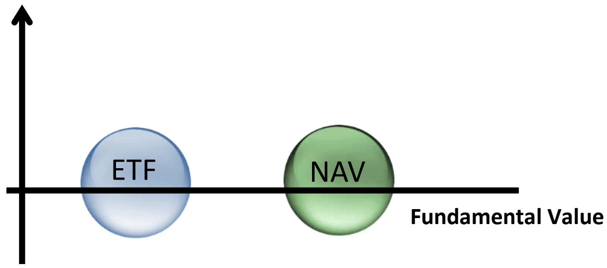



**Figure 2a.** Initial equilibrium

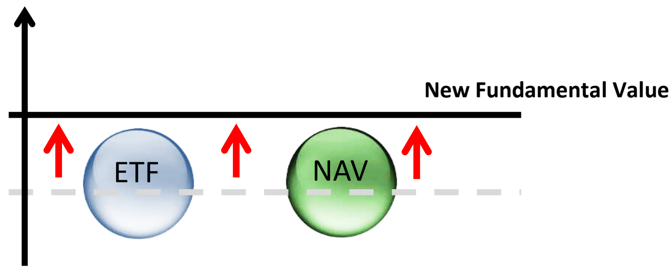



**Figure 2b.** Shock to fundamental value

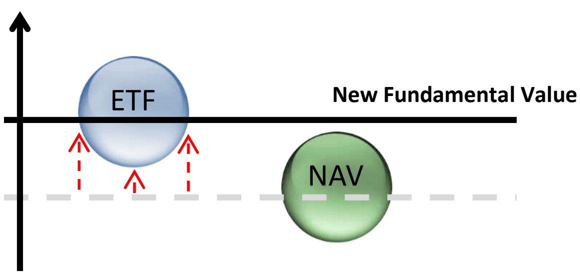



**Figure 2c.** Price discovery takes place in the ETF market. The ETF price moves to the new fundamental value.

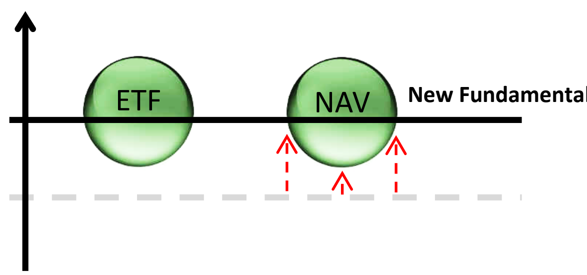
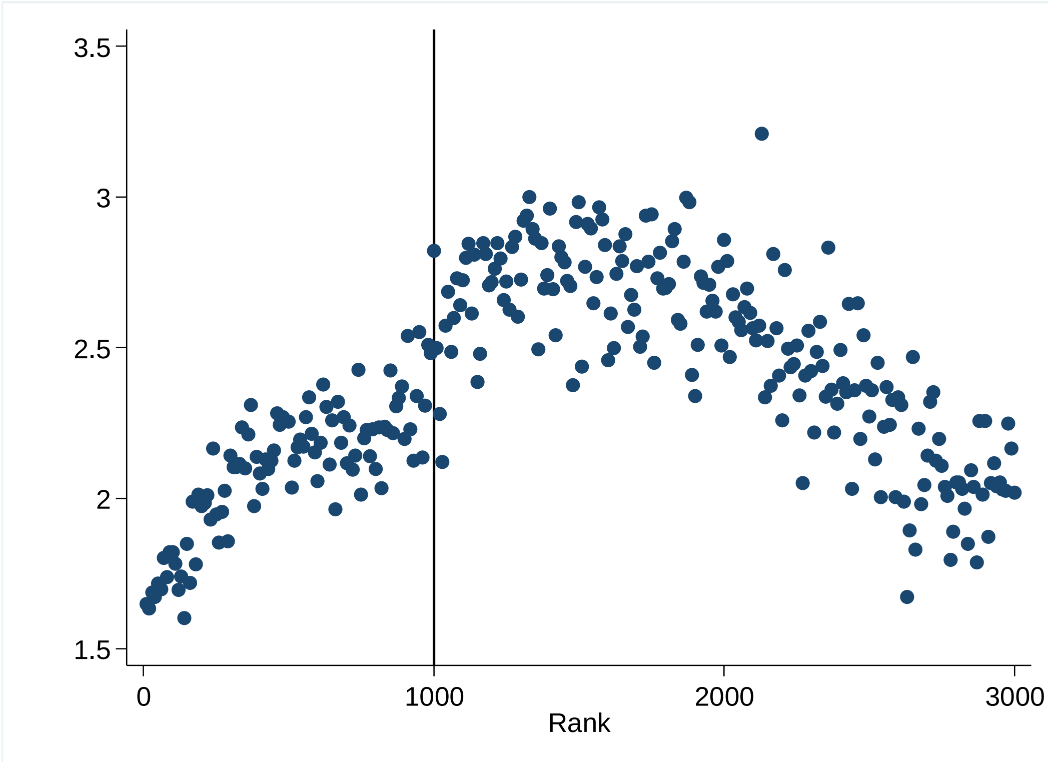



**Figure 2d.** After a delay, the NAV catches up with the new fundamental.

**Figure 3: ETF Ownership around the Russell Cutoff**




The figure reports average ETF ownership (in %) for stocks ranked by market capitalization and included in the Russell 3000. The average is computed first by ranking over time, then across the ranking in bins of 10 stocks. The vertical line denotes the 1000th rank. The sample ranges between January 2000 and December 2012.

## Appendix Table A1. Variable Definitions

| Variable | Description | Source |
|---|---|---|
| ETF ownership | The sum of the ownership of all ETFs holding the stock, using the most recent quarterly investment company reports for equity ETFs. The lagged quarterly portfolio weights are interacted with daily ETF AUM and daily stock capitalization, to compute daily ownership. The monthly variable is defined accordingly. | Thomson-Reuters, CRSP, Bloomberg |
| Index (or active) mutual fund ownership | The sum of the ownership by all index (or active) mutual funds holding the stock, using the most recent quarterly investment company reports. | Thomson-Reuters, CRSP Mutual Fund, and MFLinks |
| Daily volatility | Standard deviation of daily stock returns within a month. | CRSP |
| Intraday volatility | Standard deviation of second-by-second intraday returns. | TAQ |
| Variance ratio 15 seconds | The ratio of 15-second log return variance divided by 3 times the 5-second log return variance minus 1. The numerator and denominator are computed using returns within a day and averaged over a month. The dependent variable in the regressions is the logarithm of the absolute value of this difference. | TAQ |
| Variance ratio 5 days | The ratio of 5-day return variance divided by 5 times the 1-day return variance minus 1. The numerator and denominator are computed using daily and 5-day returns within a quarter. The dependent variable in the regressions is the logarithm of the absolute value of this difference. | CRSP |
| Net(ETF flows) | Stock-day-level measure. Weighted average of the 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 |
| Abs(mispricing) | Sum of absolute dollar mispricing across all the ETFs holding the stock divided by stock capitalization (Equation (3)). Dollar mispricing is the product of ETF mispricing (i.e., the difference between the ETF price and its NAV, as a fraction of the ETF price) times dollar holdings of an ETF in the stock. | Thomson-Reuters, CRSP, Bloomberg |
| Net(mispricing) | Similar construction to abs(mispricing). The only difference is that the ETF-level mispricing is not in absolute value. | Thomson-Reuters, CRSP, Bloomberg |
| Lending fees | Share-lending fee at the security level, 7-day average. Average within the month. | Markit |
| 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 ratio | Absolute return scaled by dollar volume in $million, average within the month. Based on Amihud (2002). | CRSP |
| Bid-ask spread | The quoted spread divided by the bid-ask midpoint. End-of-month value. | CRSP |
| Churn ratio 1 | This measure follows Cella, Ellul, and Giannetti (2013) in computing the investor-level churn ratio, which is then aggregated at the stock level using ownership weights. | CRSP, 13-F |
| Churn ratio 2 | This measure uses an investor-level churn ratio that is computed as the minimum between the absolute value of buys and sells divided by prior quarter holdings. Buys and sells use prior quarter prices. | CRSP, 13-F |

## Appendix Table A2. Institution Type Definitions

Source: Thomson Reuters Owner Types - Global Equity Ownership Feed

| Institution Type | Definition |
|---|---|
| Banks | These firms perform all of the functions of a retail bank. As a retail bank, a portfolio of investments are put together by an investment adviser and sold in units to investors by brokers. They may also handle Trust Accounts, which are outside companies or individuals that have a bank manage their money for their own pensions or for various other reasons. They invest the money their customers hold in their accounts in order to make interest payments and their own profits. |
| Endowments | Endowment Funds are permanent gifts, often to universities or colleges, which are re-invested to ensure continuing profit. |
| Hedge Funds | A hedge fund management firm who, through its hedge fund products, is permitted to use aggressive strategies that are unavailable to mutual funds, including selling short, leverage, program trading, swaps, arbitrage, and derivatives. Many times they are highly secretive because they use risky investment styles and also involve high net investors. Since they are restricted by law to less than 100 investors, the minimum investment is typically $1 million. |
| Insurance | Insurance Companies invest in a similar fashion as Investment Advisors. They re-invest the money they take in in order to make coverage payouts as well as their own profits. |
| Investment Advisors | This is the most common institution type found in the database. These are buy-side institutions who invest in stocks (equities) or bonds (fixed income). They have discretionary power over assets under management (AUM) and actually make buy/sell decisions. |
| Investment Companies | An investment vehicle operated by an investment company which raises money from shareholders and invests in a group of assets, in accordance with a stated set of objectives. Includes mutual funds. |
| Pension Funds | A qualified retirement plan set up by a corporation, labor union, government, or other organization for its employees. In order to be included in the TF database, the pension fund must manage a portion of its assets internally. |
| Individual Investors (in 13-F) | Individual investors that file the 13-F because they exercises investment discretion over the account of any other natural person or entity. |
| Research Firm | A firm that writes research intended for the buy-side community. The firm does not have an underwriting business or investment banking business. The firm does not have a proprietary trading operation. These firms typically charge for their individual research reports. |
| Corporations | Typically a business organization that is given many legal rights as an entity separate from its owners. For ownership purposes, these entities will typically be set up to represent its strategic investments. |
| Venture Capital | A firm that specializes in providing money to startup firms and small businesses with exceptional growth potential. |
| Private Equity | Firm that invests solely in private equity investments (i.e., privately held companies). They provide equity financing to small and middle market companies engaged in a variety of industries. They often focus on management buyouts, industry consolidations, recapitalization of existing business and other private equity opportunities. |
| Sovereign Funds | State-owned institutions, which invest public resources to reduce the unpredictability of government revenues, offset the boom-bust cycles' adverse effect on government spending and the national economy or foster savings for future generations. As such, SWFs aim to diversify and boost risk-adjusted returns by holding baskets of currencies, credit, and equities. |

## Appendix Table A3. ETF Ownership and Stock Intraday Volatility

The table reports estimates from OLS regressions of intraday volatility and variance ratio 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. Intraday stock volatility is computed using second-by-second data from the TAQ database. VR 15 seconds is the absolute value of the ratio of the variance of 15-second log returns on day *t* and 3 times the variance of 5-second log returns on day *t*, minus 1, using data from the TAQ database. Variable descriptions are provided in Appendix Table A1. 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. The sample ranges between January 2000 and December 2012.

| Sample: | S&P 500 | | Russell 3000 | |
|---|---|---|---|---|
| Dependent variable: | Intraday volatility | VR 15 seconds | Intraday volatility | VR 15 seconds |
| | (1) | (2) | (3) | (4) |
| ETF ownership (t-1) | 0.097*** | 0.077*** | 0.027*** | 0.029*** |
| | (7.354) | (7.028) | (8.969) | (8.822) |
| log(Mktcap (t-1)) | 0.056*** | 0.013 | -0.133*** | 0.017*** |
| | (2.707) | (0.677) | (-18.537) | (2.878) |
| 1/Price (t-1) | 6.519*** | 1.814*** | 0.869*** | 0.073** |
| | (10.161) | (6.782) | (16.066) | (2.355) |
| Amihud (t-1) | -8.315 | -65.158*** | -0.346*** | -1.473*** |
| | (-0.418) | (-4.735) | (-4.762) | (-21.103) |
| Bid-ask spread (t-1) | 12.147** | 2.461 | 25.945*** | -0.048 |
| | (2.067) | (0.763) | (17.562) | (-0.042) |
| Ret (t-1) | -0.068 | 0.011 | -0.077*** | 0.099*** |
| | (-1.507) | (0.333) | (-4.805) | (7.587) |
| Volatility (t-1) | 0.420*** | 0.139*** | 0.443*** | 0.101*** |
| | (23.098) | (43.627) | (95.021) | (87.165) |
| Stock fixed effects | Yes | Yes | Yes | Yes |
| Day fixed effects | Yes | Yes | Yes | Yes |
| Observations | 1,100,709 | 1,062,094 | 6,033,380 | 5,976,055 |
| Adjusted $R^2$ | 0.576 | 0.246 | 0.554 | 0.239 |

## Appendix Table A4. Evidence on the Arbitrage Channel (Russell 3000 Sample)

The table reports estimates from OLS regressions of intraday volatility (Panel A) and intraday variance ratio (Panel B) on absolute stock-level mispricing in the prior period interacted with measures of arbitrage costs. The frequency is daily and the observations are at the stock level. The sample includes Russell 3000 stocks. In Columns (2)-(4), arbitrage cost is captured by the bid-ask spread in the prior day, and in Columns (5)-(7), by the average share-lending fee in the month. For both measures of arbitrage costs, we construct dummy variables denoting whether the stock is in the top half of the distribution of that measure in the relevant period. In Columns (3) and (6), we restrict the sample to observations for which the stock-level mispricing is positive. In Columns (4) and (7), we restrict the sample to observations for which the stock-level mispricing is negative. Variable descriptions are provided in Appendix Table A1. 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. The sample ranges between January 2000 and December 2012.

### Panel A: Intraday Volatility

| Dependent variable: | Intraday stock volatility | | | | | | |
|---|---|---|---|---|---|---|---|
| | All | All | Misp > 0 | Misp < 0 | All | Misp > 0 | Misp < 0 |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Abs(Mispricing) (t-1) | 0.006*** | 0.008*** | 0.015*** | 0.006*** | 0.002 | 0.007*** | -0.000 |
| | (6.511) | (6.248) | (7.713) | (4.457) | (1.267) | (4.000) | (-0.189) |
| × I(High bid-ask spread) | | -0.005* | -0.013*** | -0.001 | | | |
| | | (-1.841) | (-4.128) | (-0.535) | | | |
| × I(High lending fee) | | | | | 0.005** | 0.006** | 0.003 |
| | | | | | (2.241) | (2.407) | (1.459) |
| High bid-ask spread | | 0.049*** | 0.050*** | 0.046*** | | | |
| | | (11.731) | (11.975) | (10.690) | | | |
| High lending fee | | | | | 0.046*** | 0.047*** | 0.045*** |
| | | | | | (19.600) | (19.288) | (18.723) |
| ETF ownership (t-1) | 0.014*** | 0.015*** | 0.016*** | 0.013*** | -0.003* | -0.006*** | -0.002 |
| | (5.305) | (5.815) | (6.164) | (4.823) | (-1.763) | (-3.378) | (-1.228) |
| log(Mktcap (t-1)) | -0.143*** | -0.130*** | -0.122*** | -0.137*** | -0.017*** | -0.015*** | -0.019*** |
| | (-19.509) | (-17.651) | (-16.616) | (-17.818) | (-8.067) | (-6.812) | (-8.806) |
| 1/Price (t-1) | 0.855*** | 0.868*** | 0.903*** | 0.837*** | 1.392*** | 1.391*** | 1.393*** |
| | (15.498) | (15.665) | (17.022) | (13.792) | (30.624) | (30.180) | (29.583) |
| Amihud (t-1) | -0.395*** | -0.303*** | -0.289*** | -0.323*** | -0.833*** | -0.797*** | -0.870*** |
| | (-5.386) | (-4.208) | (-3.990) | (-4.167) | (-11.564) | (-11.035) | (-11.376) |
| Bid-ask spread (t-1) | 27.481*** | 24.867*** | 23.181*** | 26.511*** | 26.369*** | 25.340*** | 27.450*** |
| | (17.473) | (15.937) | (15.126) | (15.667) | (14.489) | (14.727) | (13.814) |
| Ret (t-1) | -0.104*** | -0.112*** | -0.485*** | 0.201*** | -0.245*** | -0.614*** | 0.073*** |
| | (-6.335) | (-6.855) | (-20.810) | (9.232) | (-13.996) | (-24.655) | (3.064) |
| Dependent variable (t-1) | 0.439*** | 0.438*** | 0.445*** | 0.431*** | 0.578*** | 0.587*** | 0.570*** |
| | (90.625) | (91.451) | (89.141) | (90.748) | (94.325) | (89.992) | (96.232) |
| Abs(Mispricing) (t-2) | 0.002** | 0.002** | 0.006*** | 0.001 | 0.000 | 0.007*** | -0.004*** |
| | (2.382) | (2.413) | (4.369) | (0.883) | (0.215) | (4.885) | (-3.077) |
| Day fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Stock fixed effects | Yes | No | No | No | No | No | No |
| Observations | 5,743,866 | 5,743,866 | 2,692,645 | 3,051,221 | 5,743,866 | 2,692,645 | 3,051,221 |
| Adjusted $R^2$ | 0.556 | 0.556 | 0.556 | 0.557 | 0.513 | 0.513 | 0.513 |

## Appendix Table A4. Evidence on the Arbitrage Channel (Russell 3000 Sample) (Cont.)

### Panel B: Intraday Variance Ratio

| Dependent variable: | Intraday variance ratio (VR 15) | | | | | | |
|---|---|---|---|---|---|---|---|
| | All | All | Misp > 0 | Misp < 0 | All | Misp > 0 | Misp < 0 |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Abs(Mispricing) (t-1) | 0.001 | -0.002 | 0.003* | -0.004** | -0.003 | 0.003 | -0.005** |
| | (0.558) | (-1.557) | (1.682) | (-2.361) | (-1.476) | (1.282) | (-2.486) |
| × I(High bid-ask spread) | | 0.008*** | 0.003 | 0.009*** | | | |
| | | (3.158) | (0.988) | (3.313) | | | |
| × I(High lending fee) | | | | | 0.019*** | 0.024*** | 0.016*** |
| | | | | | (6.244) | (6.445) | (4.940) |
| High bid-ask spread | | 0.016*** | 0.017*** | 0.014*** | | | |
| | | (4.022) | (4.116) | (3.433) | | | |
| High lending fee | | | | | 0.040*** | 0.042*** | 0.038*** |
| | | | | | (9.801) | (10.022) | (9.165) |
| ETF ownership (t-1) | 0.027*** | 0.027*** | 0.028*** | 0.025*** | 0.005* | 0.001 | 0.007** |
| | (9.750) | (9.625) | (9.369) | (8.861) | (1.920) | (0.379) | (2.413) |
| log(Mktcap (t-1)) | 0.008 | 0.013** | 0.012* | 0.015** | 0.145*** | 0.146*** | 0.146*** |
| | (1.372) | (2.043) | (1.861) | (2.240) | (38.347) | (40.599) | (36.103) |
| 1/Price (t-1) | 0.083** | 0.085** | 0.075** | 0.099*** | 0.601*** | 0.519*** | 0.696*** |
| | (2.543) | (2.561) | (2.492) | (2.589) | (3.781) | (3.451) | (4.156) |
| Amihud (t-1) | -1.493*** | -1.461*** | -1.413*** | -1.494*** | -2.097*** | -2.027*** | -2.154*** |
| | (-21.505) | (-21.224) | (-19.508) | (-20.888) | (-18.543) | (-17.569) | (-18.747) |
| Bid-ask spread (t-1) | -0.364 | -1.034 | -1.229 | -0.911 | -7.941* | -6.249 | -9.718** |
| | (-0.311) | (-0.865) | (-1.029) | (-0.699) | (-1.956) | (-1.557) | (-2.338) |
| Ret (t-1) | 0.103*** | 0.101*** | 0.079*** | 0.113*** | 0.064*** | -0.092** | 0.180*** |
| | (7.631) | (7.529) | (3.866) | (6.442) | (2.806) | (-2.495) | (8.099) |
| Dependent variable (t-1) | 0.101*** | 0.101*** | 0.102*** | 0.099*** | 0.216*** | 0.220*** | 0.212*** |
| | (86.566) | (86.683) | (78.422) | (80.111) | (43.266) | (44.285) | (41.500) |
| Abs(Mispricing) (t-2) | -0.001 | -0.001 | 0.002 | -0.001 | 0.005*** | 0.012*** | 0.001 |
| | (-0.938) | (-0.530) | (1.579) | (-0.910) | (2.929) | (6.493) | (0.522) |
| Day fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Stock fixed effects | Yes | No | No | No | No | No | No |
| Observations | 5,679,729 | 5,679,729 | 2,666,323 | 3,013,406 | 5,679,729 | 2,666,323 | 3,013,406 |
| Adjusted $R^2$ | 0.245 | 0.245 | 0.247 | 0.243 | 0.166 | 0.168 | 0.166 |