ETF Sampling and Index Arbitrage

Jonathan Brogaard, **Davidson Heath**, and **Da Huang***

April 23, 2025

Abstract

This paper shows that exchange-traded funds (ETFs) "sample" their indexes, systematically underweighting or omitting illiquid index stocks. As a result, arbitrage activity between the ETF and its index has heterogeneous effects on underlying asset markets. Using an instrumental variables approach, we find that the trading activity of ETFs reduces liquidity and price efficiency and increases volatility and co-movement for liquid stocks, but has no effect on illiquid stocks. Our results demonstrate that the effects of passive investing on asset markets depend on how passive funds replicate their target index.

JEL: G11, G12, G20

Keywords: Passive investing, exchange traded funds, market quality

---

*Brogaard, jonathan.brogaard@eccles.utah.edu, University of Utah David Eccles School of Business; Heath, davidson.heath@eccles.utah.edu, University of Utah David Eccles School of Business; Huang (corresponding author), da.huang@northeastern.edu, Northeastern University D'Amore-McKim School of Business. We thank Rajesh Aggarwal, Jennie Bai, Zahi Ben-David, Markus Broman, Ben Connault, Shaun Davies (discussant), Larry Glosten (discussant), Konstantin Milbradt, Pamela Moulton, Joonki Noh (discussant), Matt Ringgenberg, Sophie Shive, Elaine Wah, Pradeep Yadav, Feng Zhang, Zhuo Zhong (discussant), and participants at Oklahoma University, the Plato MI3 Conference, the IEX Market Structure Conference, the Finance in the Cloud Conference, and the FMA and AFA for comments. We especially thank Shaun Davies for sharing his insights, code and data.

I. Introduction

Over the last two decades there has been a vast rise in passive investing, and in particular the use of exchange traded funds (ETFs). In 2002 there was \$102 billion of assets under management in U.S. ETFs; by 2020 there was \$5.4 trillion (Statista, 2020). At the same time, trading in ETFs grew from 3% of U.S. equity trading volume to over 30%. There is a growing literature that examines the effects of ETFs on the activity and quality of the underlying asset markets that they track. These studies, both theoretical and empirical, commonly assume that ETFs replicate their target index *pro rata*. Yet in practice, ETFs often specify a basket that deviates from their target index, a method called sampling.[1]

This paper shows that ETFs have a common incentive to tilt their basket weights away from illiquid assets and toward liquid assets. As a result, ETF-index arbitrage by authorized participants ("AP arbitrage") is channeled toward liquid stocks and is smaller or absent for illiquid stocks. This finding means that the effects of AP arbitrage on asset markets are channeled toward liquid stocks. Models and estimates that miss this institutional detail will understate the impact on liquid stocks and overstate the impact on illiquid stocks. In our sample, the treatment effects averaged across all stocks understate the true treatment effects on liquid stocks by up to 58%.

A fundamental tradeoff faced by all passive funds is to minimize tracking error while controlling transaction costs. This tradeoff is faced by an ETF provider who sets the creation and redemption basket for authorized participants. The fund could minimize tracking error by perfectly replicating the target index, but this approach would lead to high transaction costs.

[1] ETFs track their index through creation and redemption baskets, posted daily, which specify the set of assets that authorized participants – large market-making firms – can exchange for ETF shares.

Alternatively the fund could minimize transaction costs by only picking the largest and most liquid index assets, but this approach would lead to high tracking error. Funds therefore choose a basket that balances transaction costs and tracking error. Our first prediction is that ETFs systematically underweight or omit illiquid index assets from their basket.

In practice, ETFs have considerable discretion in defining their daily creation/redemption baskets (Lettau and Madhavan, 2018) and many ETFs' basket weights diverge from the weights of their target index. For example, six of the ten largest ETFs in 2019 state in their prospectus that they statistically replicate their target index via a basket of representative securities.[2] In our sample of ETFs, sampling is more popular with funds that track broader indexes constituents, but is also used by ETFs tracking narrower and more liquid large-cap indexes.

To date the empirical literature investigates the effects of ETFs on asset markets *on average*, finding that increased ETF holdings cause changes in a stock's correlation with the market (Da and Shive, 2018), volatility (Ben-David, Franzoni, and Moussawi, 2018), liquidity (Sağlam, Tuzun, and Wermers, 2019) and price efficiency (Israeli, Lee, and Sridharan, 2017; Glosten, Nallareddy, and Zou, 2021). We reexamine these outcomes using a novel research design and focusing on heterogeneity in the effects of AP arbitrage. In the most liquid stocks we find that increased AP arbitrage leads to higher market correlation and lower market quality. However, in illiquid stocks the effects are significantly smaller or even of the opposite sign. The average treatment effects on market quality documented in the prior literature miss this variation, which is important for interpreting the overall impact of ETFs on asset markets.

[2]One example is the Vanguard Total Stock Market ETF (VTI). The fund's prospectus states, "The Fund invests by sampling the Index, meaning that it holds a broadly diversified collection of securities that, in the aggregate, approximates the full Index in terms of key characteristics."

A growing literature examines how ETFs interface with underlying asset markets. Easley, Michayluk, O'Hara, and Putniņš (2021) show that many ETFs diverge from the value-weighted market portfolio. Evans, Moussawi, Pagano, and Sedunov (2019) show that through shorting ETFs arbitrageurs limit potential adverse effects on the liquidity of the underlying securities. Shim and Todorov (2022) show that bond ETFs use strategic sampling to limit the potential adverse effects of ETF arbitrage on the underlying securities. In contemporaneous work, Koont, Ma, Pastor, and Zeng (2023) construct a model of bond ETFs' basket choice. The basic tradeoff between transaction costs and tracking error is the same in both models. Koont et al. (2023) focus on how funds steer their basket toward the index, relative to their holdings. By contrast, we examine equity ETFs and focus on how funds tilt the weighting within their basket.

We first document that the inclusion or exclusion of an index stock in an ETF basket is strongly predicted by its liquidity. An index member stock with an effective spread that is one standard deviation higher is nearly three times more likely to be omitted from the ETF's daily basket. This pattern holds across different fund types – it is strongest in funds that track a total market index, but is economically and statistically significant in small-cap and large-cap funds as well.

Next we examine the effects of ETFs and authorized participants creation/redemption activity on the underlying assets. We focus on the creation and redemption of ETF shares in exchange for the posted basket of assets – "primary flow" – and the AP arbitrage activity that ETF primary flow induces in the underlying assets. Looking at stocks' intraday trading patterns we find that on days with more AP arbitrage activity, liquid stocks see higher turnover at the end of the trading day, consistent with AP arbitrage shifting trading volume toward the end of the trading day. This pattern is much smaller and statistically insignificant among illiquid stocks.

We proceed in two steps to isolate the causal effect of ETF index arbitrage from potential confounding variables, such as information arrival and market conditions. First, we instrument the daily primary flow of each ETF with its stale lagged returns. Box, Davis, Evans, and Lynch (2021) find that lagged ETF returns and trading activity do not predict future returns, which suggests that they do not contain market-relevant information. Dannhauser and Pontiff (2021) and Broman (2022) find that uninformed retail investors exhibit return-chasing behavior and this pattern is especially strong in ETFs. We build on this insight and show that lagged ETF returns predict the net primary flows in and out of each ETF. Second, we multiply the instrumented primary flow by the ETF's posted basket, which is pre-announced each day and thus induces non-discretionary trading (Greenwood, 2007; Lou, 2012).

We show that the cross-sectional effects of ETF index arbitrage differ systematically from the average effects. AP arbitrage causes liquid stocks to have worse liquidity and price efficiency, and higher volatility and comovement with the market. For illiquid stocks, these effects are smaller or even go the opposite way. This heterogeneity means that unconditional estimates understate the effects on asset market quality. For example, a one standard deviation increase in ETF index arbitrage causes the pricing error of the most liquid tercile of stocks to increase by 1.91%, but has a slightly negative and insignificant effect on the least liquid tercile of stocks. Pooling all stocks together yields an estimated effect of 0.56%, which understates the effect by nearly 70%. These results are robust to a variety of controls and high dimensional fixed effects structures.

We check three alternative explanations for our findings. First, we examine market fragmentation and algorithmic trading. Regulation National Market System (Reg NMS) was established in 2005 and had significant effects on quoted spreads, market fragmentation, and market quality that differ by stock market capitalization (Haslag and Ringgenberg, 2022). Second, algorithmic and high-frequency trading have risen significantly during our sample period, and their effects interact with asset liquidity in complex ways (Li, Wang, and Ye, 2021). Either of these factors could potentially explain our findings via an alternative channel. We control for these factors directly and find that stock-by-day measures of market fragmentation and algorithmic trading activity do not explain the differential effects of AP arbitrage on underlying assets. We also examine the role of market-moving news in two ways, by directly controlling for daily market returns and by restricting the sample to "no-news" days on which the market return is less than $\pm 1\%$. Our results are again qualitatively the same, inconsistent with market-moving news explaining our results.

This paper contributes to the empirical and theoretical literature on ETFs and passive investing. Empirically, Greenwood (2007) and Da and Shive (2018) find that a higher index weight leads stocks to co-move more with the index and less with stocks that are not in the index. Glosten et al. (2021) find a positive relation between ETF ownership and stock information efficiency. Ben-David et al. (2018) find that increased ETF ownership leads to higher stock volatility; by contrast Box et al. (2021) find little evidence that ETF trading impacts underlying constituent security prices. Israeli et al. (2017) find that increased ETF ownership leads to lower price efficiency and higher return synchronicity. Sağlam et al. (2019) and Marta (2019) find that ETF ownership improves the liquidity of index stocks and bonds respectively, and Agarwal, Hanouna, Moussawi, and Stahel (2018) find that higher ETF ownership increases the commonality in liquidity of the underlying stocks. Our findings add nuance to the ongoing investigation of the effects of ETFs on volatility and market quality (although ETFs also affect underlying stocks via other channels). Specifically, we find that increased AP arbitrage activity leads to higher volatility and lower liquidity and market quality.

This paper also relates to the theoretical literature on the impact of ETFs and passive investing on asset markets. Carpenter (2000), Basak, Pavlova, and Shapiro (2007), and Basak and Pavlova (2013) show that funds tilt their portfolio toward stocks that belong to their benchmark. Our empirical findings confirm the predictions of Malamud (2016) and Cespa and Foucault (2014), who construct equilibrium models in which ETF creation/redemption serves as an information propagation mechanism. Similarly, Bhattacharya and O'Hara (2018,0) show that ETFs can increase market fragility and lead to market runs. Pan and Zeng (2023) construct a model in which a liquid ETF tracks a single illiquid asset, and they analyze the effects of authorized participants' market making activity on the asset's liquidity. Many theoretical models simply assume that the fund replicates its benchmark pro rata. By contrast, we show that ETFs replicate their index strategically. As a result, their portfolio is tilted relative to their target index, in a common direction across ETFs. Models of the impact of ETFs can be made richer and more realistic by incorporating this feature.

II. Data

Daily market data for both stocks and ETFs are from the CRSP daily file. The stock-level data covers all U.S. listed common stocks in CRSP from January 2015 through December 2019. Intraday transaction-level data comes from TAQ. The sample period is from 2015 to 2019.[3] We

[3]The sample starts from 2015 because our institutional subscription of millisecond TAQ starts in 2015. Transaction-level data needed to construct the variables of interest before 2015 is only available on the

exclude any stocks with a prior-month closing market capitalization of less than $300 million ("micro-caps") and exclude any stocks with a prior-month closing price per share less than $5, following Asparouhova, Bessembinder, and Kalcheva (2013). After these filters, we are left with 3309 stocks in the sample from January 2015 to December 2019.

Table 1 Panel A reports the summary statistics of stocks in the sample. The sample contains 3,404 unique stocks and covers effectively all listed U.S. common stocks during our sample period. The definition of all variables can be found in Appendix A.

Insert Table 1 About Here

The ETF data starts from U.S. equity ETFs in the CRSP mutual funds database with a prior-month closing assets under management (AUM) of at least $100 million. We obtain daily ETF shares outstanding from Bloomberg as Ben-David et al. (2018) documents that shares outstanding data from CRSP is less accurate. Daily shares owned of ETFs by Robinhood users are obtained from the website Robintrack.net. We obtain data on ETF basket weights from Markit. Historical data on index constituents and weights are from Russell Investments for Russell indexes, from S&P for S&P indexes, and are calculated from CRSP data for CRSP indexes.

We use the benchmark index disclosed in the ETF's prospectus for the indices whose weights are obtainable from index providers such as S&P, Russell, and CRSP. For ETFs whose benchmark index weights are not obtainable from the index providers, we impute the benchmark following Cremers and Petajisto (2009). When we compare the imputed and true index for those

second-by-second level, which can lead to severe measurement error (Holden and Jacobsen, 2014). Untabulated analyses show our results are robust to earlier sample with second-level TAQ data.

ETFs where we know the true index, in every case, the imputed index is the true index. This observation bolsters confidence in our imputation.

Table 1 Panel B shows summary statistics for the ETFs in the sample. Consistent with other studies (Dannhauser and Pontiff, 2021) ETF expense ratios are low industry-wide with a mean (median) of 19 (15) basis points per year. The tracking error, which we calculate each month as the standard deviation of the daily difference in returns between the fund and its benchmark index, is also quite low with a mean (median) of 7.8 (6.1) basis points.

Figure 1(a) shows how the total trading volume and assets under management (AUM) of U.S. equity ETFs has evolved over time. The total AUM in U.S. equity ETFs has been growing rapidly, and since 2008 ETF trading has comprised approximately one-third of all U.S. equity trading activity. The vast majority of ETF trading activity occurs in the secondary market and does not lead to AP arbitrage (Fulkerson, Jordan, and Travis, 2022). Figure 1(b) shows how ETF primary flows – the daily flows in and out of ETFs via creation and redemption of ETF shares – have grown over time. Compared with Figure 1(a), between 3% and 8% of quarterly ETF trading volume results in creation or redemption activity, while the rest is secondary market trading and does not result in inflows or outflows.

Insert Figure 1 About Here

III. ETF mechanics and trading activity

This section describes exchange-traded funds (ETFs) and the features of ETFs that motivate our empirical approach. Subsection A describes the creation and redemption mechanism that ensures ETF shares track their target index closely. Subsection B describes daily net ETF fund flows ("primary flow") and ETF index arbitrage which is how ETFs interface with underlying asset markets.

A. ETF basket choice

An exchange-traded fund (ETF) is an investment intermediary that tracks a basket of underlying securities. ETF shares are listed on an exchange and trade throughout the day. ETF shares track the underlying basket because of the arbitrage activity of authorized participants (APs), which are large market making firms. APs have access to the creation and redemption mechanism which allows them to exchange ETF shares for the basket of underlying securities with the ETF provider at the end of each trading day. If the ETF's shares trade sufficiently above the basket's net asset value, an AP sells ETF shares and buys the underlying basket of securities and vice versa. APs thus provide a liquid two-sided market for the ETF's shares, which facilitates secondary trading activity (Lettau and Madhavan, 2018; Evans et al., 2019).

The creation and redemption baskets are set daily by the ETF provider. The ETF provider has complete freedom to define the baskets, and even "heartbeat" transactions for a single index asset are common.[4] In setting the baskets, the basic tradeoff that the ETF provider faces is to minimize expected transaction costs, which reduces the bid-ask spread that APs are willing to offer and promotes trading, and to simultaneously minimize expected tracking error against the target index.

[4] A heartbeat trade occurs when an ETF provider exchanges ETF shares for a large quantity of a single asset in a one-off transaction. These are commonly used to reflect an index reconstitution or rebalancing (Moussawi, Shen, and Velthuis, 2020).

In the Internet Appendix we present a model of the ETF provider's problem and optimal basket weight. Underweighting a stock relative to the index reduces the expected transaction costs of trading it, at the cost of increasing tracking error. By contrast, overweighting a stock relative to the index increases both expected transaction costs and tracking error. Thus, the ETF provider strategically underweights index assets that are relatively illiquid.

1. Stock liquidity and ETF basket inclusion

To gauge the extent of sampling, we first examine the basket weights of the ETFs in our sample relative to their benchmark index weights.

Table 2 lists the ETFs for which we obtained matched historical index weight data from the index provider. For Russell and CRSP indexes we obtained monthly historical index weights; for the S&P indexes we obtained yearly data. The list contains popular funds from the three largest fund providers, namely BlackRock, State Street, and Vanguard. The sample covers four of the five largest ETFs as of 2019 and covers 37% of total AUM and 44% of total trading volume for U.S. equity ETFs over the sample period.

We first give an example with the creation-redemption basket of VTI, the Vanguard Total U.S. Stock Market fund, relative to its benchmark index, the CRSP value-weighted U.S. stock market index. On January 2, 2015, the first date in our sample, the CRSP index consisted of 3,695 stocks, while VTI's basket consisted of 1,497 stocks. However, it is not clear from this example what fraction of ETFs actually use sampling and to what extent they use it in practice. Even if some ETFs tracking broad indexes use sampling quite heavily, the practice of sampling could be much more limited in practice, especially for large-cap ETFs that do not track such broad indexes.

To systematically examine how ETF baskets differ from their benchmark indexes, we construct a fund-stock-month panel. This panel consists of all stocks in each ETF's benchmark index at the beginning of a each month, for all ETFs in the sample for which we have historical basket and index data (Table 2). We use the stock i's basket weight relative to its index weight for ETF j in each month m to compute its Relative Weight:[5]

$$\text{RelWeight}_{i,j,m} = \frac{\text{BasketWeight}_{i,j,m}}{\text{IndexWeight}_{i,j,m}}$$

Table 2 displays the average relative weight for each ETF in the sample. 36 of the 56 ETFs in our sample have an average relative weight below 95% and 35 out of 56 have an average relative weight below 90%. Thus, for various thresholds, the majority of ETFs in our sample plausibly follow a sampling approach. We provide additional evidence supporting the prevalence of ETF sampling in the Internet Appendix.

Next, we sort all sample stocks into terciles each month based on their liquidity as measured by their prior month average effective spread. Figure 2 shows the average basket weight compared to the index weight in relative terms, for stocks in each tercile, across all ETFs in the sample. Consistent with our model's prediction, the relative weight of an index stock in an ETF's basket is much lower, on average, for less liquid stocks.

Since the most illiquid stocks are most underweighted, a further prediction of our model is that the dispersion in relative weights within a given ETF should reflect the dispersion in liquidity across stocks in the index. In unreported tests, available on request, we check this prediction. We

[5]When we calculate each stock's relative weight in each basket, we drop observations where the index weight is less than 1 basis point (0.0001), so that our results are not driven by observations with very low index weights. The results are similar if we winsorize the relative weights or if we omit this filter entirely.

find that ETFs with more dispersion of liquidity across the stocks within their index have more dispersion in their relative weights as well, and this relationship holds across ETF categories as well as individual funds.

Next we examine formal regression estimates which address a variety of potential confounders. The effect of a stock's lagged liquidity on its $\text{RelWeight}_{i,j,m}$ can be interpreted as the effect on that stock's basket weight holding its index weight fixed. We also control for stocks' index weight in a flexible fashion, to absorb any (potentially nonlinear) effects that index weight might have on ETFs' basket weighting decisions. To do so, we estimate the following equation:

$$\text{RelWeight}_{i,j,m} = \text{Liquidity}_{i,m-1} + \text{Index Weight Control}_{i,j,m} + \gamma_{j,m} + \epsilon_{i,j,m}$$

where $\text{Liquidity}_{i,m-1}$ is the liquidity of stock i in prior month measured by effective spread or bid-ask spread, standardized to a standard deviation of 1 for ease of interpretation. $\text{Index Weight Control}_{i,j,m}$ denotes whether the model contains the index weight (Linear), weight and squared weight (Square), or the weight, squared weight, and cubed weight (Cubic) of stock i in the index in month m. Since the model's prediction is based on a stock's relative liquidity, we include fund-year-month fixed effects $\gamma_{j,m}$, which means that the variation is across stocks within each ETF's index at each point in time.

Table 3 panel A reports the results and shows that, controlling for its index weight and including fund-by-year-month fixed effects, a stock's liquidity strongly negatively predicts its ETF basket weight. The effect is economically large and highly statistically significant. For example, in Column 3 the coefficient of the stock's relative weight on its prior month effective spread is -1.90. Put differently, a stock in the ETF's benchmark index with a one-half standard deviation lower liquidity has approximately a 95% higher likelihood relative to the average of being omitted from the ETF basket.[6] Panel B reports the results within the subset of funds for which we do not need to impute the benchmark index. This sample contains more large-cap funds, so that the magnitude of the relation between liquidity and basket weight is smaller, but still highly statistically significant and robust.

The Internet Appendix presents results when we compute an alternative measure of sampling behavior that accounts for the relative significance of each stock within the index. This approach reduces the disproportionate influence of stocks with minimal index weights. We find again that sampling behavior is widespread across our sample of ETFs.

Overall, the ETF basket weight data are consistent with the prediction that ETFs strategically underweight or omit illiquid index assets in their creation-redemption baskets. These results bolster our second prediction that the effects of ETF index arbitrage on underlying asset markets should be concentrated in liquid underlying assets, and should be weaker or absent in illiquid underlying assets.

B. ETF primary flow and index arbitrage

1. Primary flow

The net creation and redemption of ETF shares at the end of the day, through which assets and investor funds flow in or out of the ETF, is referred to as "primary flow." Primary flow is how

[6]In untabulated estimates we find that the time series variation in the relative weights is an order of magnitude smaller. Thus, the variation in relative weights is mostly in the cross-section of each ETF basket.

ETFs interface with underlying asset markets. The effects of ETF holdings on stocks and firms have been extensively studied (Ben-David et al., 2018; Glosten et al., 2021; Sağlam et al., 2019). Yet, an ETF with large and stable holdings may have a large or small (or zero) primary flow on any given day. Total trading volume in an ETF can also be misleading, because most of the daily trading volume in ETFs reflects bilateral trades (Lettau and Madhavan, 2018) that do not impact the underlying markets. By contrast, ETF primary flow directly reflects APs making use of the creation/redemption mechanism and exchanging ETF shares for index assets (Brown, Davies, and Ringgenberg, 2021).

We calculate each ETF j's primary flow on day t as the absolute change in shares outstanding on day t compared to day $t-1$ times the closing price on day $t-1$:[7]

$$\text{Primary Flow}_{j,t} = (\text{Shares Outstanding}_{j,t} - \text{Shares Outstanding}_{j,t-1}) \times \text{Price}_{j,t} \ .$$

[7]We exclude observations when the ETF provider performed a heartbeat trade, a one-off exchange of ETF shares for one or a few assets (Moussawi et al., 2020). We filter out heartbeat trades because they do not originate from APs' market making activities in the ETF, and they introduce considerable noise in the daily primary flow data. We define a heartbeat trade as a paired inflow and outflow that satisfy: (i) both are greater than 1% of the fund's AUM, (ii) the inflow is the largest primary flow during the surrounding two months, and (iii) the outflow occurs less than five trading days following the inflow.

2. Index Arbitrage

We calculate the index arbitrage activity on the stock-by-day level by projecting the ETF's primary flow onto the weight matrix of the daily ETF basket. Specifically:

$$|\text{APArb}_{i,t}| = \frac{\sum_j |\text{PrimaryFlow}_{j,t}| \times w_{i,j,t}}{\text{MarketCap}_{i,m-1}}$$

where $w_{i,j,t}$ is ETF j's basket weight in stock i on day t scaled by the market cap of the stock as of the previous month. The absolute value is taken on ETFs' primary flow before the projection to reflect that index arbitrage occurs at an ETF-by-ETF level. For example, if a stock is a constituent of two ETFs with 1% basket weight and the two ETFs have a \$100 and -\$100 primary flow respectively, the stock is subject to \$2 index arbitrage activity. The assumption underlying the construction of $|\text{APArb}_{i,t}|$ is that primary flows are netted within each ETF-day, but that primary flows in opposite directions for different ETFs or on different days do not net with one another. In reality, netting of primary flows could occur within APs (Pan and Zeng, 2023) and/or across multiple trading days (Evans et al., 2019).

3. AP arbitrage affects intraday trading activity

To examine the effects of ETF index arbitrage on underlying markets, we first compare trading days with low AP arbitrage activity to days with high AP arbitrage activity. The prediction is that days with high AP arbitrage activity should see an increase in trading activity in underlying assets, but (i) only in the most liquid assets and (ii) toward the end of the trading day, when APs close out their residual positions.

Figure 3 plots the difference in intraday share turnover for U.S. equities between days on which they experience high AP arbitrage activity versus days on which they experience low AP arbitrage activity. We see that when a stock is subject to higher AP arbitrage activity, trading activity surges in the last half hour of the trading session, consistent with the idea that AP arbitrage activity is concentrated late in the day near the market close.

Insert Figure 3 About Here

The intraday patterns in trading volume are suggestive of how APs perform ETF index arbitrage. Specifically, APs internalize their trading in ETF shares during the day. At the end of the trading day, APs adjust their net position using the creation/redemption mechanism. Since this involves delivering or receiving a wide basket of assets, ETF index arbitrage is thus reflected in the underlying markets.

This delay and pooling of trading activity in the underlying assets suggests that ETFs have effects on asset liquidity and market quality. This finding aligns with that of Box et al. (2021) that price transmission between the ETF and the underlying constituents is achieved via quote adjustments rather than arbitrage trades throughout the trading day. During days when AP arbitrage is high, the delay of activity in the underlying market until the end of the trading day is likely to consume liquidity and worsen information efficiency for the underlying assets (Kyle, 1985; Glosten and Milgrom, 1985). On the other hand, it is not clear how large these effects should be, since AP arbitrage activity still accounts for a small fraction (less than 1%) of total trading volume for the stocks in our sample.

Next, we examine the effects of ETF index arbitrage on underlying asset markets. In order to identify causal effects, we isolate exogenous variation in AP arbitrage activity with a novel "return-chasing" instrumental variable.

IV. Research Design

There are many market forces that affect both ETF index arbitrage and underlying asset markets, which makes causal inference challenging. We develop a novel instrumental variable to overcome the identification problem. Specifically, we instrument each ETF's signed daily primary flow with its lagged (i.e. stale) prior-month return. Dannhauser and Pontiff (2021), among others, show that uninformed investors "return-chase", and this pattern is especially strong in ETFs relative to open-ended mutual funds. As a result, investor dollars flow into ETFs with high lagged returns and out of ETFs with low lagged returns, even though the lagged returns do not predict future returns (Broman, 2022; Box et al., 2021).

First, we estimate the predictive regression:

$$\text{PrimaryFlow}_{j,t} = \sum_{k=1}^{4} \beta_k \text{Return}_{j,t-k} + \beta \text{Return}_{j,t-26\to t-5} + \gamma_j + \kappa_t + \epsilon_{j,t} \tag{1}$$

where j denotes ETF and t denotes trading day. $\text{Return}_{j,t-k}$ is the lagged individual daily return of ETF i for the last four trading days and $\text{Return}_{j,t-26\to t-5}$ is the cumulative return of ETF j over the last month from $t-5$ to $t-26$ (22 trading days). γ_j is ETF fixed effects and κ_t is day fixed effects.

We use unadjusted fund returns instead of risk-adjusted returns because the literature suggests that retail investors respond to unadjusted returns (Odean, 1998; Ben-David and Hirshleifer, 2012). The key independent variable is the fund's return from trading day $t-26$ to trading day $t-5$, i.e. the return over the prior month, lagged by five trading days. We lag the return by five trading days (at least one calendar week) for three reasons. First, the lag avoids the possibility of reverse causation from ETF flows (for example, due to investors rebalancing) to ETF returns. Second, the lag rules out omitted variables such as information arrival, which could drive both investor demand for a particular ETF and that ETF's returns. In the modern equity market where high-frequency traders operate on a time horizon of microseconds, it is safe to assume that information that was known five trading days ago is stale and already incorporated into prices. Third, the multi-day lag allows for discretion in the funds' operational strategy and reporting.[8]

Insert Table 4 About Here

Table 4 Column 1 shows that a higher lagged prior-month return predicts a higher ETF primary flow. To verify that the relationship is driven by uninformed investors return-chasing, first, we examine the daily net flow of trades by retail investors in each ETF using the methodology in Boehmer, Jones, Zhang, and Zhang (2021). Column 2 shows that indeed, the net flow of trades by retail investors is strongly positively predicted by the lagged prior-month return. Second, we examine the daily holdings of users of Robinhood, a smartphone trading app that is popular among retail investors. Table 4 Column 3 shows that an ETF's lagged prior-month return even more strongly predicts its popularity with Robinhood users. In sum, the results are consistent with our proposed mechanism of return-chasing by uninformed retail investors.[9]

[8]Evans et al. (2019) document that APs can create "operational shorts" by creating new ETF shares and waiting several business days to deliver the basket of underlying assets; and Staer (2017) documents that some ETFs use T+1 accounting while others use T+0 accounting to report their shares outstanding. We thank Markus Broman for pointing this out.

[9]Most evidence suggests Robinhood users are typical retail investors whose trades are unlikely to contain market-relevant information, even though in some cases their trading might be profitable for other reasons. Welch (2022) documents that during the Covid period, Robinhood traders overall timed the market and generated alpha.

We use the predicted values from the regression (1) as an exogenous shifter of ETF index arbitrage. Specifically, we multiply the coefficient on $\text{Return}_{j,t-26\rightarrow t-5}$ from Table 4 Column 1 by each ETF's lagged prior-month return to produce a daily predicted primary flow for each ETF on each day. We take the absolute value, because a large positive or negative primary flow implies more index arbitrage activity. The first stage F-statistic is 13.5 (p=0.0004); thus, the condition of instrument relevance is plausibly satisfied (Angrist and Kolesár, 2023).

$$|\widehat{\text{PrimaryFlow}}_{j,t}| = |\beta_{FirstStage} \times \text{Return}_{j,t-26\rightarrow t-5}|$$

$$|\widehat{\text{APArb}}_{i,t}| = \frac{\sum_j |\widehat{\text{PrimaryFlow}}_{j,t}| \times w_{i,j,m(t)-1}}{\text{MarketCap}_{i,m(t)-1}}$$

The main variable of interest $|\widehat{\text{PrimaryFlow}}_{i,t}|$ is plausibly unrelated to information on day t and therefore plausibly satisfies the exclusion restriction, for two reasons. First, $|\widehat{\text{PrimaryFlow}}_{j,t}|$ only reflects fund flows that are predicted based on each ETF's prior-month return lagged by five trading days. Figure 4 plots the values of the summation of all ETFs' actual (realized) and predicted primary flow on daily level. We see that the predicted absolute primary flow accounts for only a small fraction of the overall variation in realized primary flow. In fact, the two time

This does not necessarily contradict our interpretation that Robinhood trades do not contain information. The apparent market timing in Welch (2022) comes from the lack of selling by Robinhood traders in the market crash during the initial Covid crisis. This one observation could be an artifact of slow moving retail investor behavior (Barber, Huang, Odean, and Schwarz, 2022) with subsequent price impact (Frazzini, Israel, and Moskowitz, 2018; Gabaix and Koijen, 2021). Indeed, Welch (2022) states that his findings may represent a one-time phenomenon with limited external validity.

series are uncorrelated with one another ($\rho = 0.02$), suggesting that any information in actual ETF primary flows is not reflected in our instrumented ETF primary flows.

Next, we multiply the predicted ETF-level primary flows for ETF j by its basket weights as of the end of the prior month $w_{i,j,m(t)-1}$ to obtain the predicted level of AP arbitrage activity in stock i on day t, scaled by the stock's market cap as of the end of the prior month. This stock-by-day variable $|\widehat{\text{APArb}}_{i,t}|$ resembles "flow-induced trading" in the mutual fund literature, which isolates trading activity by mutual funds due to fund flows (Coval and Stafford, 2007; Lou, 2012). Similarly, we isolate trading activity due to the (predicted) primary flows of ETFs.

In sum, our independent variable of interest $|\widehat{\text{APArb}}_{i,t}|$ is driven by predictable variation in ETF fund flows based on stale returns. This activity is passed through to the underlying stocks depending on their prior-month basket weights. The basket weights are stable over time; the cross-sectional variation within each ETF is two orders of magnitude more than the variation in the weights over time. As a result, instrumented AP arbitrage activity is unlikely to be correlated with any stock-specific or weight-specific information on day t.[10]

All these features suggest that our research design plausibly satisfies the exclusion restriction. On the other hand, the fact that, at some point in the future, ETF premiums and returns reverse is true of most financial data and trading strategies given the arrival of new information. Though we attempt to rule out alternative explanations, information or return predictability could explain part of our results. We provide additional evidence and discussion supporting the validity of the instrument in the Internet Appendix.

Insert Figure 4 About Here

[10]We conduct additional robustness checks in Section VI.

V. The effects of AP arbitrage on underlying assets

In this section, we examine the effects of ETF index arbitrage on underlying asset markets. We first describe the estimation strategy and investigate the effects on asset market liquidity and quality. We then examine effects on asset returns. Lastly, we contrast the instrumental variables estimates with the corresponding ordinary least squares (OLS) estimates, which shows the magnitude of the endogenous market forces that jointly affect ETF flows and asset market activity.

A. Effects on asset market quality

In this section, we investigate the heterogeneous effects of ETF index arbitrage on the quality of underlying asset markets. The key prediction is that the effects will be strongest in liquid stocks, and weaker or even opposite in illiquid stocks.

For each month from January 2015 through December 2019, we sort stocks into terciles based on their average effective spread over the previous month. The first tercile contains the most liquid stocks and the third tercile contains the most illiquid stocks *ex ante*.

We estimate the following equation:

$$\begin{aligned} Y_{i,t} = &\sum_{q=1}^{3} \beta^q \times |\widehat{\text{APArb}}_{i,t}| \times \text{Liquid}^q_{i,t} \\ &+ \text{Liquid}^q_{i,t} + \boldsymbol{X}_{i,m(t)-1} + \boldsymbol{Y}_{i,t-1} + \gamma_i + \kappa_t + \epsilon_{i,t} \end{aligned} \tag{2}$$

$Liquid^q_{i,t}$ is a set of three dummy variables that equal 1 if stock i was in liquidity tercile q in the previous month. $\boldsymbol{X}_{i,m-1}$ includes lagged stock-level controls – turnover, market capitalization, close price, and all market quality measures (effective spread, pricing error, volatility, and index correlation) as of the previous month. We control for lagged market quality measures to account for potential autocorrelation. Stock fixed effects, γ_i, sweep out any time-invariant stock-specific factors. The year-month fixed effects κ_t sweep out any time trends in the sample.

The key independent variable, $|\widehat{\text{APArb}}_{i,t}|$, is the instrumented magnitude of ETF index arbitrage in stock i on day t (Table 2). We standardize i.e. scale this variable to have a standard deviation of one over the entire sample. Thus, the coefficients β^1, β^2, β^3 capture the effect of a one standard deviation increase in AP arbitrage activity on the outcome variable of interest, for stocks in the high-, medium- and low-liquidity terciles respectively.

The dependent variables $Y_{i,t}$ include effective spread, pricing error, intraday volatility, and intraday correlation of stock i on day t. The effective spread is the signed distance between trade price and the midpoint of NBBO, scaled by trade price and averaged across the entire trading day weighted by the dollar value of each trade. The pricing error is calculated according to Hasbrouck (1993). The intraday volatility is the standard deviation of 1-minute returns over the entire trading day. The intraday correlation is the correlation between the 1-minute returns of the stock and the 1-minute returns of the SPDR S&P 500 ETF over the entire trading day. [11]

We begin by examining stock liquidity. On one hand, trading due to AP arbitrage is plausibly uncorrelated with information about any particular stock, and uninformed trades are liquidity-improving (Kyle, 1985; Glosten and Milgrom, 1985). Thus, trading due to AP arbitrage could be liquidity-improving. On the other hand, our data show that ETF-driven trades appear

[11] We also calculate these outcome variables using only data from the last 30-minutes of the trading day because the primary market activity due to AP arbitrage is concentrated near market close. We report these results in the Internet Appendix Section IA2 and they yield qualitatively the same results.

mostly at the end of the trading day, pooling with other rebalancing activity and possibly crossing bid-ask spreads to get executed.[12] Thus, trading driven by ETF index arbitrage could be liquidity-consuming.

Table 5 Column 1 shows the differential effect of AP arbitrage activity on underlying stocks' liquidity. We see that more AP arbitrage activity leads to wider effective spreads – reduced liquidity – in the most liquid tercile of stocks. The effect becomes weaker in the second tercile, in which stocks are mostly included in the creation/redemption basket of the ETF but are underweighted. The effect flips sign for the least liquid tercile – that is, more AP arbitrage activity lead to slightly *better* liquidity in illiquid assets – perhaps because the AP arbitrage directs liquidity-consuming trading activity out of those assets. A Wald test for the high versus low liquidity terciles yields an F-statistic of 95.9, which rejects the null that the effects are the same with $p < 0.01$. These results suggest that APs' index arb activity consumes liquidity in the underlying stocks, consistent with the results of Dannhauser (2017) and Eaton, Green, Roseman, and Wu (2021), and this effect is concentrated in *ex ante* liquid stocks.

Insert Table 5 About Here

Table 5 Columns 2 to 4 show the differential effect of AP arbitrage activity on other aspects of underlying market quality. In Column 2 the outcome variable is the short-run price inefficiency (volatility of pricing error) following Hasbrouck (1993). We find that AP arbitrage activity leads to worse price efficiency, but again, only for stocks in the most liquid tercile. A one standard deviation increase leads to a 1.9% increase in the short-run pricing error, comparable to the estimates in Israeli et al. (2017). However, the effect is weaker in the second tercile and is

[12] The Internet Appendix Section IA3 presents IV estimates that show the same patterns as in Figure 3.

insignificant in the least liquid tercile. A Wald test rejects with $p < 0.01$ the null that the treatment effects are the same for the high-liquidity and low-liquidity tercile.

In Table 5 Column 3 the outcome variable is the intraday return volatility of the stock. Consistent with Ben-David et al. (2018), who find a positive effect of ETF ownership on monthly return volatility, we find that more AP arbitrage activity is accompanied by higher return volatility in the two most liquid terciles. A one standard deviation increase leads to a 2.5% increase in liquid stocks' volatility. The volatility-inducing effect is present in all three terciles but is monotonically smaller when liquidity is lower.

In Table 5 Column 4 the outcome variable is the correlation between the intraday returns of the individual stock and the SPDR S&P 500 ETF (SPY). Consistent with Greenwood (2007) and Da and Shive (2018), who find a positive effect of ETF ownership on index correlation, we find that AP arbitrage activity leads to a higher intraday correlation with the market – but only in the two most liquid terciles of stocks. In the least liquid tercile of stocks, the effect is economically and mostly statistically insignificant.

Our simple model and mechanism predict that excluded stocks should see zero treatment effects. Instead, for the least liquid stocks we see treatment effects that are near-zero and mostly statistically insignificant. This may be because our model does not capture many other aspects of market behavior such as substitutability between stocks, other holdings of the ETF sponsor (Pan and Zeng, 2023), and effects on noise trader behavior (Foucault, Sraer, and Thesmar, 2011).

To sum up, we find that ETF index arbitrage has negative effects on liquidity and price efficiency and positive effects on volatility and return correlation. The unconditional effects we find are close to the magnitudes documented in the existing literature. For example, Ben-David et al. (2018) find a 1 standard deviation increase in ETF ownership is associated with a 0.16

standard deviation higher daily stock volatility; we find that a 1 standard deviation increase in ETF index arbitrage is followed by a 0.75 / 9.42 = 0.08 standard deviation change. Our unconditional estimates for price efficiency and comovement are also similar to those of Glosten et al. (2021) and Greenwood (2007). Since our research design and sample are different from these prior studies, our findings represent out-of-sample confirmation of their results.

More importantly, we find that these effects are all present for the tercile of liquid stocks only and are weak, absent, or even opposite in illiquid stocks. This finding is consistent with our main prediction. The mechanism driving this prediction is that ETFs on average systematically tilt their baskets in the same direction, namely toward liquid assets and away from illiquid assets. Put differently, the treatment effects of ETFs on asset markets that prior research has documented are confined to a minority of liquid underlying assets.

This heterogeneity also means that unconditional estimates of average treatment effects, which are implicitly equal-weighted across all sample stocks, will tend to understate the effects of ETFs on asset markets. The bottom row of Table 5 reports the average treatment effects (i.e. without splitting stocks into liquidity terciles). We see that all of the estimates are significantly smaller than the treatment effect within the most liquid tercile; the understatement varies from -35% for volatility to -71% for pricing error. Thus, the fact that the treatment effects vary systematically across assets based on those assets' liquidity is of first-order importance for understanding the effects that ETFs have on asset markets.

B. Price impact and information content

Since AP arbitrage have impacts on trading activity and market quality, a natural question is whether they impact asset returns. First, if the trading activity of APs has significant price impact, we would expect to see an effect on closing prices and returns on day t. Second, if our instrumented AP arbitrage activity does contain information about future fund flows or asset values, violating the exclusion restriction, then we would expect to see an effect on returns on days $t+1$, $t+2$, and so on. Thus, looking for effects of our instrument on future stock returns represents a test of the exclusion restriction.

Table 6 presents estimates with daily stock returns as the outcome variable. In column 3, we see that AP arbitrage activity in stock i has no significant relationship with its contemporaneous return. The lack of statistical and economic significance suggests that contemporaneous AP arbitrage activities have limited price impact.

AP arbitrage activity also has an economically insignificant relation with returns on future trading days, and with no differential effect between liquid and illiquid terciles. A Wald test fails to reject equality with an F-stat of less than 2. The lack of predictive power for future returns leads us to conclude that our instrument is not confounded by market-relevant information. Instead, it reflects flow-driven, non-discretionary trading activity driven by the pre-announced ETF basket weights. We conduct additional robustness checks in Section VI which also cut against the possibility that instrumented AP arbitrage activity could potentially contain market-relevant information.

We observe that instrumented AP arbitrage activity is slightly positively associated with the stock's return over the two prior trading days leading up to day t. The magnitude of the differential relations with prior-day returns is quite small – 3 to 4 basis points per standard deviation of AP arbitrage activity. This observation is consistent with APs potentially taking time to smooth out their execution over multiple days (Evans et al., 2019), and motivates our lagging the stale monthly return instrument by several trading days.

In sum, the relation of instrumented ETF index arbitrage with returns on day t suggests no significant price impact, and its relation with returns on days $t+1$ and $t+2$ suggests no meaningful information content.[13] The latter findings are consistent with the findings of other studies of ETF market impact which use different research designs (Dannhauser and Pontiff, 2021; Box et al., 2021).

C. Comparing IV and OLS estimates

In this section, we compare our instrumental variables IV with the corresponding ordinary least squares (OLS) regressions. The OLS estimates use the realized variation in AP arbitrage activity (Figure 4). In contrast to our instrumented ETF flows, which use only stale lagged ETF returns, realized ETF flows are largely driven by endogenous market forces – for example, information arrival and investor rebalancing, which affect both market quality and fund flows simultaneously. By contrasting the OLS estimates with the IV estimates, we can observe the magnitude of these endogenous market forces. In other words, this comparison gives a sense of how confounded the relationship between ETF index arbitrage and asset market quality is, and how necessary the IV strategy is for accurate inference.

[13] We provide additional robustness checks in Internet Appendix Section IA4 by running the same tests for creation and redemption flows separately. The effects are economically insignificant in both directions, consistent with the results in Table 6.

The Internet Appendix Section IA5 presents the OLS regression results. The specification is identical to the IV estimates except that the predicted value of $|\widehat{\text{APArb}}_{i,t}|$ is replaced with the realized value of $|\text{APArb}_{i,t}|$. In each case, the OLS estimates have the same sign but larger magnitudes than our IV estimates. Thus, the OLS regressions, even with high dimensional fixed effects, would lead researchers to overstate the impact of ETFs on asset markets because a significant part of their association is driven by omitted market factors.

More importantly, due to endogeneity, the OLS estimates (i) overstate the effects of ETF index arbitrage on assets, and (ii) understate the differential treatment effects caused by ETF basket choice.

VI. Alternative Explanations

One concern with our findings is that market dynamics are changing over time and could drive both AP arbitrage activity and also stock turnover and liquidity. First, market structure evolves over time and time-varying factors such as high-frequency trading activity and market fragmentation are known to impact market quality (Weller, 2018; Haslag and Ringgenberg, 2022). Second, the arrival of market-relevant information could drive increased AP arbitrage activity and at the same time cause market makers and participants to alter their trading activity in individual stocks. This section examines these three alternative explanations for our findings – high-frequency trading (HFT), market fragmentation, and market-moving news.

First, it could be that the increase in HFT activity over our sample period differentially affected stocks' liquidity. This is because high-frequency traders potentially improve stock liquidity but endogenously choose to trade stocks that are inherently liquid Brogaard, Hendershott, and Riordan (2014). Greater HFT activity or competition also plausibly affects how APs carry out ETF index arbitrage. To capture HFT activity in each stock over time, we construct the trade-to-order ratio using the SEC MIDAS data following the same procedure in Weller (2018). For each stock, we sum the trade volume and the order volume over the month and calculate the ratio of the two:

$$TOR_{i,m} = \frac{\sum_{t\in m} \text{Trading Volume}_{i,t}}{\sum_{t\in m} \text{Order Volume}_{i,t}}$$

where i denotes stock and m denotes month.[14] This yields a measure, TOR, that is highly associated with HFT activity as demonstrated by Weller (2018).[15]

Insert Table 7 About Here

We then rerun our main estimates on measures of market quality, adding the contemporaneous stock-by-month values of TOR. We interact TOR with the liquidity terciles so that we capture and control for differential effects of HFT activity. Table 7 shows the results. HFT activity, as measured by TOR, has strong differential effects on market quality. For example, higher HFT activity is associated with significantly improved liquidity for all stocks, with the

[14]We calculate this measure on monthly level instead of daily level for the main specification. This is because we are interested in controlling the structural change in HFT throughout the sample period rather than the day-to-day changes in HFT intensity. We provide a robustness check using daily-level HFT measure as control in the Internet Appendix Section IA6. The results are similar.

[15]We also use other measures proposed by Weller (2018), including odd-lot ratio (OLR), cancel-to-trade (CTR), and average trade size (ATS), as proxies for HFT activity and obtain similar results. We report these results in the Internet Appendix Section IA7.

effect being strongest in the most liquid tercile of stocks (Table 7 Column 1). Columns 2 to 4 show HFT's differential effects on the other measures of market quality as well.

Despite the differential effects of HFT trading on market quality, the differential effects of AP arbitrage survive almost unchanged. Compared to Table 5 the coefficients and significance levels for the most and least liquid terciles and the difference between them remain similar in all cases. Thus, variation in high-frequency trading activity during our sample does not explain the effects that we find.

Second, it could be that the increase in market fragmentation over our sample period affected stocks' liquidity and also ETF index arbitrage activity. To examine this possibility, we construct a stock-by-month Herfindahl-Hirschman Index (HHI) of trading volumes across market venues:

$$HHI_{i,m} = \sum_{j,t \in m} \left(\frac{\text{Trading Volume}_{i,j,t}}{\sum_j \text{Trading Volume}_{i,j,t}} \right)^2$$

where i denotes stock, j denotes market venue, and m denotes month.[16] This measure captures the degree of realized market fragmentation, for each stock in each month individually.

Insert Table 8 About Here

We then rerun our main estimates on measures of market quality, adding the contemporaneous stock-by-month controls for market fragmentation (HHI). Table 8 shows the

[16]We calculate this measure on monthly level instead of daily level for the main specification. This is because we are interested in controlling the structural change in market fragmentation throughout the sample period rather than the day-to-day changes in HFT intensity. We provide a robustness check using daily-level fragmentation in the Internet Appendix Section IA6. The conclusions are the same.

results. Market fragmentation has strong differential effects on market quality. For example, greater market fragmentation is associated with significantly worse liquidity for the most liquid tercile of stocks, but with significantly improved liquidity for the least liquid tercile (Table 8 Column 1). Columns 2 to 4 show strong differential effects on the other measures of market quality as well.

However, despite the differential effects of market fragmentation on market quality, the differential effects of AP arbitrage survive almost unchanged. Compared to Table 5 the coefficients and significance levels for the most and least liquid terciles and the difference between them remain similar in all cases. Thus, variation in market fragmentation does not explain the effects that we find.

Third, it could be that when there is market-moving news (or the risk of market-moving news arriving), market makers and participants change their trading behavior differentially in liquid versus illiquid assets, at the same time affecting ETF index arbitrage activity. To examine this possibility we add the magnitude (absolute value) of the CRSP value-weighted U.S. market index on each day as an additional explanatory variable, interacted with each stock's liquidity tercile.

Insert Table 9 About Here

Table 9 shows the results. The magnitude of the market return has its own differential effects on market quality. However, the differential effects of AP arbitrage again survive, and are again similar compared to the estimates in Table 5. Compared to Table 5 the coefficients and significance levels for the most and least liquid terciles and the difference between them remain similar in most cases, albeit the statistical significance of the differential effect is weaker for correlation.

In untabulated tests we restrict our sample to days on which the CRSP value-weighted U.S. market index return was within the range [-1%, 1%]. This robustness check drops days on which the market moved significantly, likely when some significant event or news occurred. Restricted to “quiet market days”, our estimates are again similar to those in Table 5. This is an additional robustness check against the possibility that our instrument might carry market-relevant information that it could bias our inference. Taken together, we conclude that market-moving news does not explain the differential effects that we find.

Thus, for the three major market forces that we examine – high-frequency trading, market fragmentation, and market-moving news – we conclude that our estimates are not driven by these alternative explanations.

VII. Conclusion

The objective of an exchange-traded fund (ETF), and passive index funds more generally, is to accurately track a target index at low cost. In this paper, we examine the direct consequences of ETFs’ index replication strategy.

We model funds as facing a fundamental tradeoff between tracking error and transaction costs. The model predicts that ETF providers are better off underweighting or omitting assets that are illiquid and expensive to trade. As a result, the effects of ETF index arbitrage on underlying assets are heterogeneous and determined by the asset’s relative liquidity within the ETF’s target index. This is an important point for research on passive investing and the rise of ETFs, because the tilt away from illiquid assets is systematic across funds.

We examine ETF basket choice empirically using a sample of U.S. equity ETFs that spans different fund providers and styles. We find that in large-cap, small-cap, and total-market funds, ETFs systematically tilt their baskets away from illiquid stocks.

Using a novel instrument for ETF primary flow – the stale lagged prior-month return of the fund – we document that funds' sampling of their underlying index causes ETF index arbitrage to have differential impact on underlying asset markets. Specifically, on days with more instrumented AP arbitrage activity, in the tercile of stocks with the highest *ex ante* liquidity, liquidity and price efficiency fall, and volatility and correlation with the market rise. Conversely, in the tercile of stocks with the lowest *ex ante* liquidity, these effects are absent or opposite in sign. These results are robust to a variety of specifications and are not explained by other market factors.

Our results also show that unconditional estimates of the treatment effects, which implicitly average across all sample assets, understate the effects of AP arbitrage on markets. In our sample, the unconditional average effects understate the effects on liquid assets by up to 58% and overstate the effects on illiquid assets.

These facts have implications for both the theoretical and empirical literature on passive investing and the rise of ETFs. First, many theoretical models of passive investing assume that the fund replicates its benchmark *pro rata*. By contrast, we show that ETFs replicate their index strategically and their portfolio is tilted relative to their target index in a common direction across ETFs. Models of the impact of ETFs can be made richer and more realistic by incorporating this feature. Second, a growing empirical literature investigates the growth of passive investing and the effects on asset markets. We show that ignoring this aspect of ETF trading activity mistakenly classifies liquid and illiquid assets as equally “treated”, and results in biased estimates of the effects of ETF index arbitrage on asset markets.

While we find that ETF index arbitrage increases volatility and reduces liquidity and market quality in liquid index assets, this does not imply that ETFs are a net negative socially. To evaluate the effect on investor welfare, outcomes must be compared to a counterfactual in which these funds are not available, and any costs they impose must be weighed against the benefits they provide, which include letting investors trade and invest in diversified investment products with unprecedented efficiency.

References

Agarwal, V.; P. Hanouna; R. Moussawi; and C. W. Stahel. “Do ETFs increase the commonality in liquidity of underlying stocks?” *Working Paper*.

Angrist, J., and M. Kolesár. “One instrument to rule them all: The bias and coverage of just-id IV.” *Journal of Econometrics*.

Asparouhova, E.; H. Bessembinder; and I. Kalcheva. “Noisy prices and inference regarding returns.” *Journal of Finance*, 68 (2013), 665–714.

Barber, B. M.; X. Huang; T. Odean; and C. Schwarz. “Attention-Induced Trading and Returns: Evidence from Robinhood Users.” *Journal of Finance*, 77 (2022), 3141–3190.

Basak, S., and A. Pavlova. “Asset Prices and Institutional Investors.” *American Economic Review*, 103 (2013), 1728–1758.

Basak, S.; A. Pavlova; and A. Shapiro. “Optimal asset allocation and risk shifting in money management.” *Review of Financial Studies*, 20 (2007), 1583–1621.

Ben-David, I.; F. Franzoni; and R. Moussawi. “Do ETFs increase volatility?” *Journal of Finance*, 73 (2018), 2471–2535.

Ben-David, I., and D. Hirshleifer. “Are investors really reluctant to realize their losses? Trading responses to past returns and the disposition effect.” *Review of Financial Studies*, 25 (2012), 2485–2532.

Bhattacharya, A., and M. O’Hara. “Can ETFs Increase Market Fragility? Effect of Information Linkages in ETF Markets.” *Working Paper*.

Bhattacharya, A., and M. O'Hara. "ETFs and Market Runs." *Working Paper*.

Boehmer, E.; C. M. Jones; X. Zhang; and X. Zhang. "Tracking retail investor activity." *Journal of Finance*, 76 (2021), 2249–2305.

Box, T.; R. Davis; R. Evans; and A. Lynch. "Intraday arbitrage between ETFs and their underlying portfolios." *Journal of Financial Economics*, 141 (2021), 1078–1095. ISSN 0304-405X.

Brogaard, J.; T. Hendershott; and R. Riordan. "High-frequency trading and price discovery." *Review of Financial Studies*, 27 (2014), 2267–2306.

Broman, M. S. "Naïve Style-level Feedback Trading in Passive Funds." *Journal of Financial and Quantitative Analysis*, 57 (2022), 1083–1114.

Brown, D. C.; S. W. Davies; and M. C. Ringgenberg. "ETF arbitrage, non-fundamental demand, and return predictability." *Review of Finance*, 25 (2021), 937–972.

Carpenter, J. N. "Does option compensation increase managerial risk appetite?" *Journal of Finance*, 55 (2000), 2311–2331.

Cespa, G., and T. Foucault. "Illiquidity contagion and liquidity crashes." *Review of Financial Studies*, 27 (2014), 1615–1660.

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

Cremers, K. M., and A. Petajisto. "How Active Is Your Fund Manager? A New Measure That Predicts Performance." *Review of Financial Studies*, 22 (2009), 3329–3365.

Da, Z., and S. Shive. “Exchange traded funds and asset return correlations.” *European Financial Management*, 24 (2018), 136–168.

Dannhauser, C. D. “The impact of innovation: Evidence from corporate bond exchange-traded funds (ETFs).” *Journal of Financial Economics*, 125 (2017), 537–560.

Dannhauser, C. D., and J. Pontiff. “FLOW.” *Working Paper*.

Easley, D.; D. Michayluk; M. O’Hara; and T. J. Putniņš. “The active world of passive investing.” *Review of Finance*, 25 (2021), 1433–1471.

Eaton, G. W.; T. C. Green; B. Roseman; and Y. Wu. “Zero-Commission Individual Investors, High Frequency Traders, and Stock Market Quality.” *Working Paper*.

Evans, R. B.; R. Moussawi; M. S. Pagano; and J. Sedunov. “ETF Short Interest and Failures-to-Deliver: Naked Short-Selling or Operational Shorting?” *Working Paper*.

Foucault, T.; D. Sraer; and D. J. Thesmar. “Individual investors and volatility.” *The Journal of Finance*, 66 (2011), 1369–1406.

Frazzini, A.; R. Israel; and T. J. Moskowitz. “Trading costs.” *Working Paper*.

Fulkerson, J. A.; S. D. Jordan; and D. H. Travis. “ETF Arbitrage and Daily Cash Flow.” *Journal of Investing*.

Gabaix, X., and R. S. Koijen. “In search of the origins of financial fluctuations: The inelastic markets hypothesis.” National Bureau of Economic Research (2021).

Glosten, L., and P. R. Milgrom. “Bid, ask and transaction prices in a specialist market with heterogeneously informed traders.” *Journal of financial economics*, 14 (1985), 71–100.

Glosten, L.; S. Nallareddy; and Y. Zou. "ETF trading and informational efficiency of underlying securities." *Management Science*, 67 (2021), 22–47.

Greenwood, R. "Excess comovement of stock returns: Evidence from cross-sectional variation in Nikkei 225 weights." *Review of Financial Studies*, 21 (2007), 1153–1186.

Hasbrouck, J. "Assessing the quality of a security market: A new approach to transaction-cost measurement." *Review of Financial Studies*, 6 (1993), 191–212.

Haslag, P. H., and M. Ringgenberg. "The Demise of the NYSE and NASDAQ: Market Quality in the Age of Market Fragmentation." *Journal of Financial and Quantitative Analysis*, Forthcoming.

Holden, C. W., and S. Jacobsen. "Liquidity measurement problems in fast, competitive markets: Expensive and cheap solutions." *Journal of Finance*, 69 (2014), 1747–1785.

Hortaçsu, A., and C. Syverson. "Product differentiation, search costs, and competition in the mutual fund industry: A case study of S&P 500 index funds." *Quarterly Journal of Economics*, 119 (2004), 403–456.

Israeli, D.; C. M. Lee; and S. A. Sridharan. "Is there a dark side to exchange traded funds? An information perspective." *Review of Accounting Studies*, 22 (2017), 1048–1083.

Koont, N.; Y. Ma; L. Pastor; and Y. Zeng. "Steering a Ship in Illiquid Waters: Active Management of Passive Funds." *Working Paper*.

Kyle, A. S. "Continuous auctions and insider trading." *Econometrica*, 53 (1985), 1315–1335.

Lee, C. M., and M. J. Ready. “Inferring trade direction from intraday data.” *Journal of Finance*, 46 (1991), 733–746.

Lettau, M., and A. Madhavan. “Exchange-traded funds 101 for economists.” *Journal of Economic Perspectives*, 32 (2018), 135–54.

Li, S.; X. Wang; and M. Ye. “Who provides liquidity, and when?” *Journal of Financial Economics*, 141 (2021), 968–980.

Lou, D. “A flow-based explanation for return predictability.” *Review of Financial Studies*, 25 (2012), 3457–3489.

Malamud, S. “A dynamic equilibrium model of ETFs.” *Working Paper*.

Marta, T. “Fixed Income ETFs, Bond Liquidity, and Stressed Markets.” *Working Paper*.

Moussawi, R.; K. Shen; and R. Velthuis. “ETF Heartbeat Trades, Tax Efficiencies, and Clienteles: The Role of Taxes in the Flow Migration from Active Mutual Funds to ETFs.” *Working Paper*.

Odean, T. “Are investors reluctant to realize their losses?” *Journal of Finance*, 53 (1998), 1775–1798.

Pan, K., and Y. Zeng. “ETF arbitrage under liquidity mismatch.” *Working Paper*.

Sağlam, M.; T. Tuzun; and R. Wermers. “Do ETFs Increase Liquidity?” *Working Paper*.

Shim, J. J., and K. Todorov. “ETFs, Illiquid Assets, and Fire Sales.” *Working Paper*.

Staer, A. “Fund flows and underlying returns: The case of ETFs.” *International Journal of Business*, 22 (2017), 275–304.

Statista. “Total net assets under management (AUM) of Exchange Traded Funds (ETFs) in the United States.”

Welch, I. “The wisdom of the Robinhood crowd.” *Journal of Finance*, 77 (2022), 1489–1527.

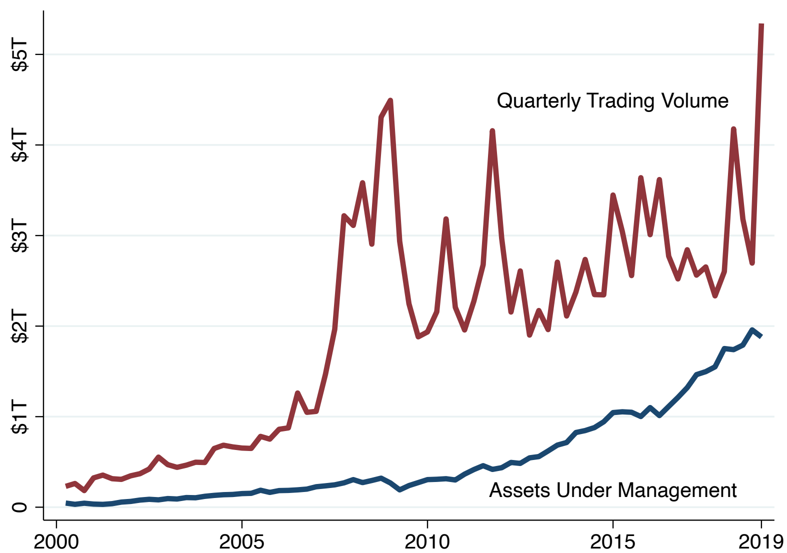
Weller, B. M. “Does algorithmic trading reduce information acquisition?” *Review of Financial Studies*, 31 (2018), 2184–2226.

FIGURE 1

ETF Trading Volume and Primary Flows

Figure (a) plots quarterly total trading volume and total assets under management across U.S. equity ETFs. Figure (b) plots quarterly primary flows (total dollar creation and redemption activity for all funds that experienced a net creation or a net redemption, respectively) across U.S. equity ETFs.

(a) **ETF trading volume and AUM**



(b) **ETF primary flows**

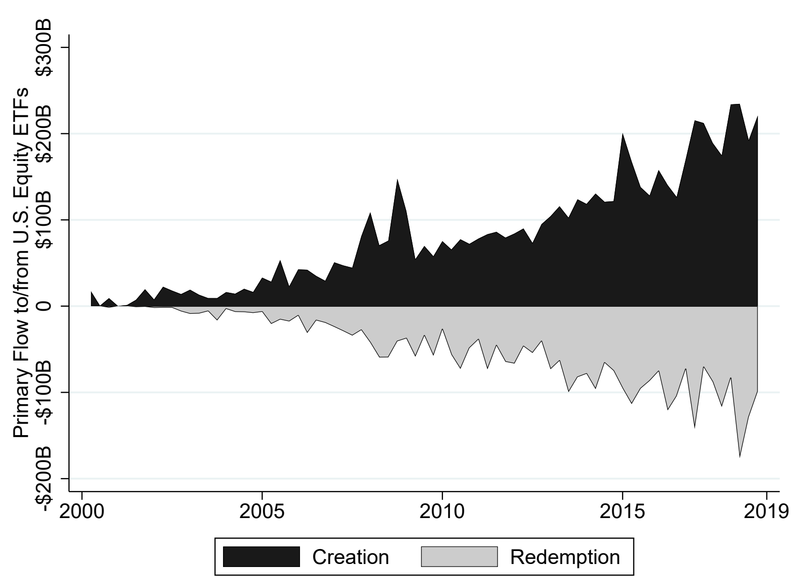


FIGURE 2

Stock Liquidity and ETF Basket Underweighting

The figure displays the relative weight difference (basket weight divided by index weight) in ETF creation/redemption baskets across all the ETFs in our sample from 2015-2019. Sample stocks are sorted into terciles by their liquidity, measured as the prior month's average effective spread.

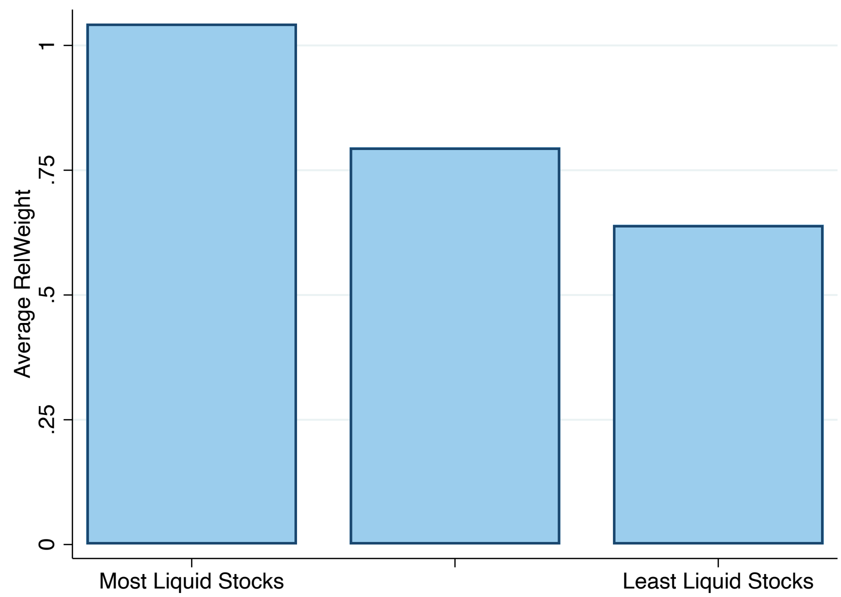


FIGURE 3

ETF Primary Flow and Asset Turnover during the Trading Day

The figure plots the difference in intraday share turnover for U.S. equities between days on which they experience high AP arbitrage activities and days on which they experience low AP arbitrage activities. Within each stock, we sort the sample days using the predicted AP arbitrage activities for the stock and calculate the average share turnovers during all intraday intervals for top decile days and those for bottom decile days. We then average them across all sample stocks for all intraday intervals. The sample consists of U.S. common stocks from 2015-2019.

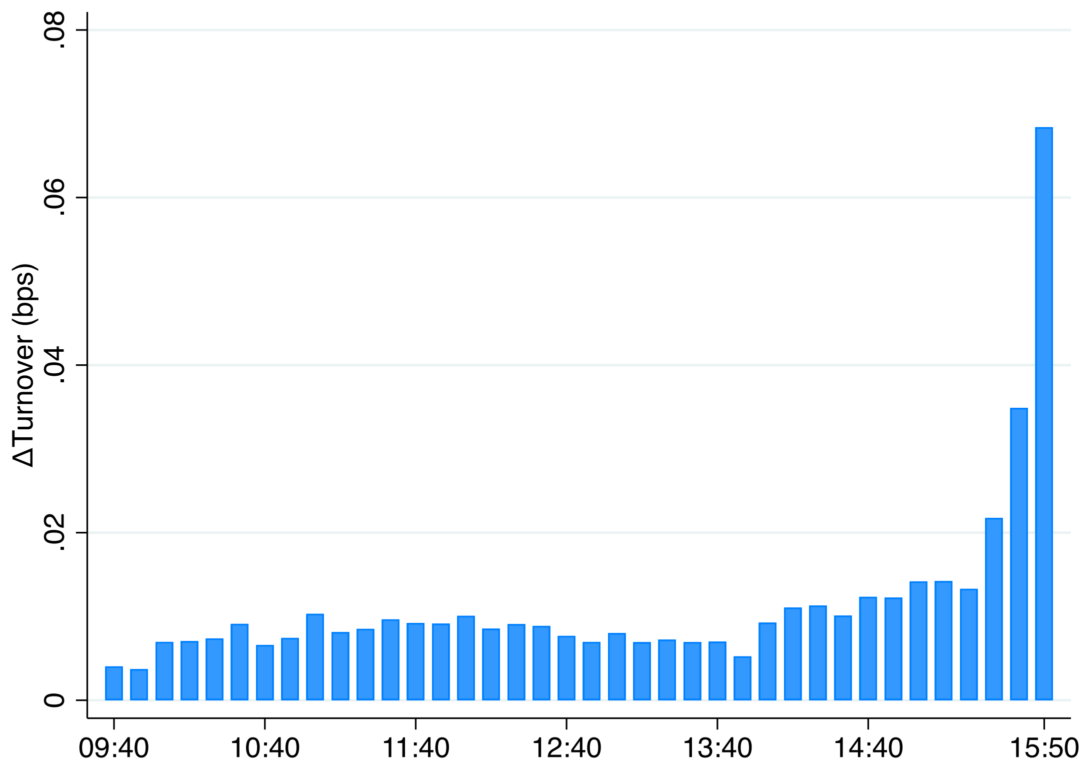


FIGURE 4

Realized Versus Instrumented ETF Primary Flows

The figure compares realized ETF primary flows daily during 2015-2019, to ETF primary flows predicted using each individual ETF's lagged return from t-26 to t-5. The sample consists of U.S. equity ETFs.

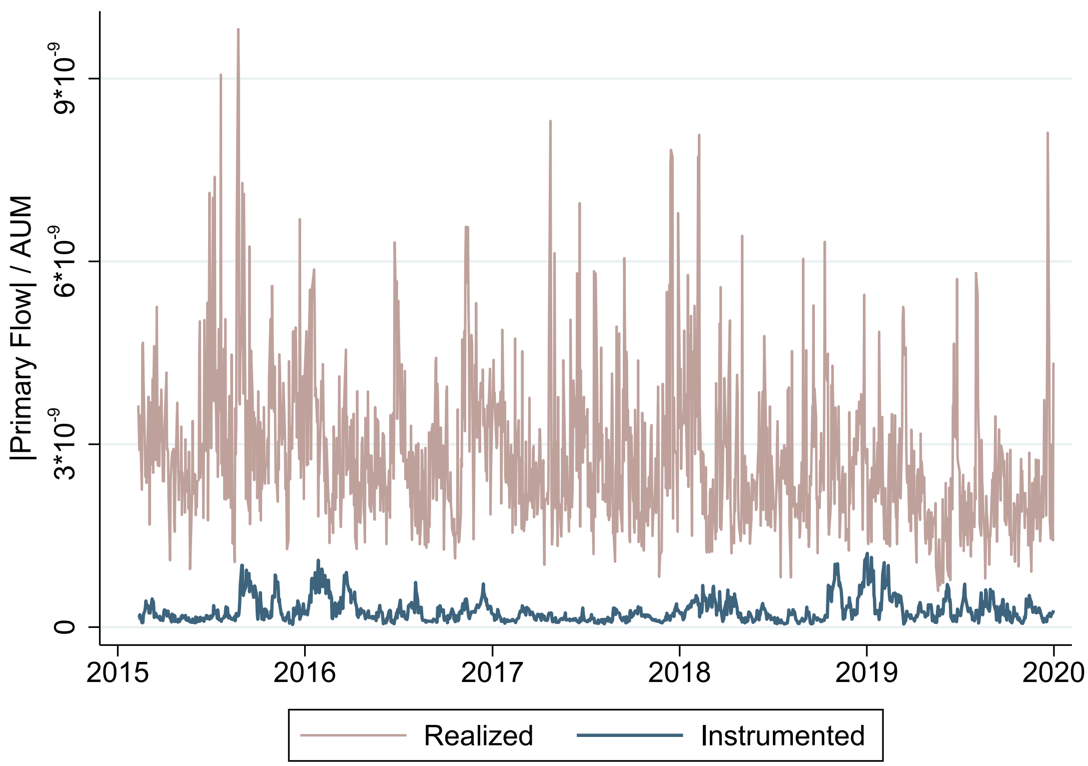


TABLE 1

Summary Statistics

Panel A presents summary statistics of the sample stocks, which consists of all U.S. common stocks with a lagged market capitalization of at least $300 million from January 2015 through December 2019. The sample contains 3,404 unique stocks. |Arbitrage| is the authorized participants' arbitrage activity on stock level, calculated as the absolute value of primary flows of all sample ETFs projected on the fund-by-stock basket weight matrix. Trade turnover is measured as the daily trading volume in shares divided by shares outstanding expressed in basis points. Effective spread is value-weighted and expressed as a fraction of stock price in basis points. Pricing error is estimated using the method in Hasbrouck (1993). Intraday volatility and correlation with SPY are measured at 1-minute frequency. Panel B displays summary statistics of ETFs in the sample, which consists of U.S. equity ETFs in the CRSP mutual fund database with at least $100 million of assets under management from January 2015 through September 2019. The sample contains 102 unique ETFs. Assets under management is the price times total shares outstanding expressed in millions. Expense ratio is the amount shareholders pay for the ETF's operation expense divided by total investment. Turnover ratio is the minimum of aggregated sales or aggregated purchases divided by the average 12-month total net assets. Tracking error is the standard deviation of the daily difference between the ETF return and the index return. Detailed variable definitions can be found in Appendix A.

Panel A: Sample Stocks

	Mean	StDev	P10	Median	P90	N
|APArb| ($ Millions)						
Most liquid tercile	3.84	10.94	0.14	1.11	8.42	887,398
	0.63	1.45	0.04	0.32	1.35	887,423
Least liquid tercile	0.21	0.67	0.00	0.09	0.48	887,436
|APArb| to Market Capitalization (bps)						
Most liquid tercile	0.99	0.55	0.26	0.97	1.70	887,398
	0.99	0.59	0.22	0.95	1.79	887,423
Least liquid tercile	0.75	0.53	0.06	0.65	1.48	887,436
|APArb| to Trading Volume (%)						
Most liquid tercile	1.28	0.88	0.19	1.19	2.46	887,398
	1.27	0.91	0.15	1.13	2.51	887,423
Least liquid tercile	1.14	0.81	0.05	1.02	2.21	887,436
Market Capitalization ($ Millions)	10,866	38,741	462	2,046	21,009	2,686,887
Trade Turnover (bps)	0.10	0.16	0.02	0.07	0.19	2,686,887
Effective Spread (bps)	17	116	3	10	34	2,686,887
log(Pricing Error)	-7.26	1.63	-8.80	-7.54	-5.81	2,686,887
Intraday Volatility (bps)	13.90	9.42	5.67	11.59	24.41	2,686,887
Intraday Correlation with SPY	0.19	0.17	0.00	0.16	0.42	2,686,887

Panel B: Sample Exchange Traded Funds

	Mean	StDev	P10	Median	P90	N
Assets under Management ($ Millions)	14,533	27,544	2,080	6,587	30,322	123,252
Expense Ratio (bps)	19	14	5	15	39	123,252
Portfolio Turnover Ratio	0.19	0.18	0.04	0.14	0.43	123,252
Tracking Error (bps)	7.8	5.3	3.2	6.1	13.0	88,447

TABLE 2

ETF Sample, Benchmarks, and Statistics

The table lists ETFs in the sample and their benchmark used in the analysis in Table 3. We use the benchmark index disclosed in the ETF's prospectus for the indices whose weights are obtainable from index providers such as S&P, Russell, and CRSP. For ETFs whose benchmark index weights are not obtainable from the index providers (indicated by the * next to the benchmark index), we impute the benchmark following Cremers and Petajisto (2009). Relative Weight is calculated as a stock's basket weight divided by its index weight, averaged across all basket assets. The average ETF daily volume is reported in millions of dollars. |APArb| to Volume is the absolute value of daily primary flows on the ETF level divided by the ETF's daily trading volumes.

Ticker	ETF Name	Benchmark Index	Relative Weight	Volume ($M)	|APArb| to Volume
DGRO	iShares Core Div Growth	Russell 1000 Value*	0.688	22	0.429
DVY	iShares Select Div	Russell 1000 Value*	0.687	69	0.157
FNDA	Schwab U.S. S.C.	Russell 2000	0.468	8	0.493
FNDX	Schwab U.S. L.C.	Russell 1000	0.850	11	0.498
FVD	Value Line Div	Russell 1000 Value*	0.769	18	0.531
GSLC	Goldman Sachs U.S. L.C.	Russell 1000*	0.691	14	0.589
HDV	iShares Core High Div	Russell 1000 Value*	0.317	37	0.183
IJH	iShares S&P M.C. 400	S&P 400	0.968	209	0.217
IJR	iShares S&P S.C. 600	S&P 600	0.951	204	0.345
IJS	iShares S&P S.C. 600 Value	Russell 2000 Value*	0.799	25	0.453
IJT	iShares S&P S.C. 600 Growth	Russell 2000 Growth*	0.686	22	0.551
ITOT	iShares Core S&P Total Market	S&P Total Market	0.955	61	0.421
IUSG	iShares Core S&P U.S. Growth	Russell 1000 Growth*	0.927	24	0.411
IUSV	iShares Core S&P U.S. Value	Russell 1000 Value*	0.894	19	0.544
IVE	iShares S&P 500 Value	Russell 1000 Value*	0.645	97	0.148
IVV	iShares S&P 500	S&P 500	0.993	1,002	0.199
IVW	iShares S&P 500 Growth	Russell 1000 Growth*	0.684	109	0.582
IWB	iShares Russell 1000	Russell 1000	0.990	155	0.175
IWD	iShares Russell 1000 Value	Russell 1000 Value	0.989	248	0.145
IWF	iShares Russell 1000 Growth	Russell 1000 Growth	0.995	222	0.132
IWM	iShares Russell 2000	Russell 2000	0.975	3,524	0.087
IWN	iShares Russell 2000 Value	Russell 2000 Value	0.972	129	0.116
IWO	iShares Russell 2000 Growth	Russell 2000 Growth	0.989	126	0.175
IWP	iShares Russell M.C. Growth	Russell Mid Cap Growth	0.993	39	0.336

Continued on next page

TABLE 2 – continued from previous page

Ticker	ETF Name	Benchmark Index	Relative Weight	Volume ($M)	|APArb| to Volume
IWR	iShares Russell M.C.	Russell Mid Cap	0.989	54	0.607
IWS	iShares Russell M.C. Value	Russell Mid Cap Value	0.986	46	0.166
IWV	iShares Russell 3000	Russell 3000	1.005	43	0.119
MDY	SPDR S&P M.C. 400	S&P 400	0.998	412	0.211
MGK	Vanguard Mega Cap Growth	Russell 1000 Growth*	0.439	14	0.259
MTUM	iShares MSCI USA Momentum	Russell 1000*	0.577	62	0.326
NOBL	ProShares S&P 500 Aristocrats	Russell 1000 Value*	0.330	21	0.331
PRF	Invesco FTSE RAFI US 1000	Russell 1000*	1.018	16	0.174
QUAL	iShares MSCI USA Quality	Russell 1000*	0.569	40	0.527
SCHA	Schwab U.S. S.C.	Russell 2000*	0.526	26	0.226
SCHB	Schwab U.S. Broad Market	Russell 3000*	0.732	47	0.184
SCHG	Schwab U.S. L.C. Growth	Russell 1000 Growth*	1.048	24	0.225
SCHV	Schwab U.S. L.C. Value	Russell 1000 Value*	0.681	18	0.275
SCHX	Schwab U.S. L.C.	Russell 1000*	0.734	52	0.245
SDY	SPDR S&P Div	Russell 1000 Value*	0.640	59	0.159
SPHD	Invesco S&P High Div Low Vol	Russell 1000 Value*	0.425	19	0.242
SPLV	Invesco S&P Low Vol	Russell 1000 Value*	0.599	105	0.198
SPY	SPDR S&P 500	S&P 500	0.999	21,972	0.064
USMV	iShares MSCI USA Min Vol	Russell 1000*	0.707	139	0.252
VB	Vanguard S.C.	CRSP Small Cap	0.718	83	0.465
VBK	Vanguard S.C. Growth	Russell 2000 Growth*	0.380	25	0.625
VBR	Vanguard S.C. Value	Russell 2000 Value*	0.341	39	0.567
VO	Vanguard M.C.	CRSP Mid Cap	0.689	64	0.591
VOE	Vanguard M.C. Value	Russell Mid Cap Value*	0.764	29	0.476
VOO	Vanguard S&P 500	S&P 500	0.996	560	0.232
VOT	Vanguard M.C. Growth	Russell Mid Cap Growth*	0.730	17	0.373
VTI	Vanguard Total Market	CRSP Total Market	0.381	348	0.296
VTV	Vanguard Value	Russell 1000 Value*	0.612	138	0.330
VUG	Vanguard Growth	Russell 1000 Growth*	0.864	100	0.405
VV	Vanguard L.C.	CRSP Large Cap	0.839	33	0.325
VYM	Vanguard High Div Yield	Russell 1000 Value*	0.587	75	0.311
XLK	SPDR Technology Sector	S&P Tech Sector	0.999	671	0.119

TABLE 3

Stock Liquidity and Inclusion in ETF Baskets

The table presents regressions of stocks' weight in the ETF creation/redemption basket on lagged stock liquidity. The sample consists of all stocks in each fund's benchmark index, monthly for Russell and CRSP indexes and yearly for S&P indexes, from 2015 to 2019. The dependent variable is $\text{RelWeight}_{i,j,t}$, calculated as the ratio of the basket weight to the index weight for stock i in fund j in month t. $\log(\text{Effective Spread}_{i,t-1})$ is the log average effective spread of the stock, and $\log(\text{Bid-Ask Spread}_{i,t-1})$ is the log average percent bid-ask spread of the stock, over the prior month. Variable definitions are in Appendix A. We standardize the independent variables to have a mean of 0 and a standard deviation of 1. Standard errors are double clustered by year-month and ETF. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A: All funds

	(1)	(2)	(3)	(4)	(5)	(6)
$\log(\text{Effective Spread}_{i,t-1})$	-0.13***	-0.13***	-0.13***			
	(0.009)	(0.009)	(0.009)			
$\log(\text{Bid-Ask Spread}_{i,t-1})$				-0.12***	-0.12***	-0.12***
				(0.009)	(0.009)	(0.009)
Index Weight Control	Linear	Quadratic	Cubic	Linear	Quadratic	Cubic
Fund x Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,813,218	1,813,218	1,813,218	1,813,218	1,813,218	1,813,218
Adjusted R-squared	0.016	0.016	0.016	0.016	0.016	0.016

Panel B: Funds with non-imputed benchmark

	(1)	(2)	(3)	(4)	(5)	(6)
$\log(\text{Effective Spread}_{i,t-1})$	-0.04***	-0.04***	-0.04***			
	(0.003)	(0.003)	(0.003)			
$\log(\text{Bid-Ask Spread}_{i,t-1})$				-0.04***	-0.04***	-0.04***
				(0.003)	(0.003)	(0.003)
Index Weight Control	Linear	Quadratic	Cubic	Linear	Quadratic	Cubic
Fund x Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	676,320	676,320	676,320	676,320	676,320	676,320
Adjusted R-squared	0.103	0.103	0.103	0.102	0.102	0.102

TABLE 4

ETF Flows and Lagged ETF Returns

The table presents regressions of daily ETF flows on the funds' lagged returns:

$$y_{j,t} = \boldsymbol{\beta} \sum_{k=0}^{4} \text{Return}_{j,t-k} + \beta_{lm} \text{Return}_{j,t-26 \to t-5} + \gamma_j + \kappa_m + \epsilon_{j,t}$$

where j denotes ETF and t denotes day. Primary $\text{Flow}_{j,t}$ is the daily creation/redemption of ETF j on day t. Retail $\text{Flow}_{j,t}$ is daily retail order flow following Boehmer et al. (2021). Flows are scaled by the ETF's lagged assets under management. $\text{Robinhood}_{j,t}$ is the daily number of Robinhood users who held stock j. The independent variables are the daily returns for days t to $t-4$ and the prior month's lagged return from $t-5$ to $t-26$ (22 trading days). All variables are standardized to a mean of zero and standard deviation of one. Definitions can be found in Appendix A. γ_j is an ETF fixed effect and κ_m is a year-month fixed effect. The sample consists of U.S. equity ETFs in the CRSP Mutual Fund database with at least \$100 million AUM, daily from 2015 to 2019. Standard errors in parentheses are robust and clustered by fund and day. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1) Primary $\text{Flow}_{j,t}$	(2) Retail $\text{Flow}_{j,t}$	(3) $\text{Robinhood}_{j,t}$
$\text{Return}_{j,t}$	0.076***	0.051***	0.009
	(0.017)	(0.010)	(0.014)
$\text{Return}_{j,t-1}$	0.059***	0.032***	0.006
	(0.008)	(0.007)	(0.028)
$\text{Return}_{j,t-2}$	0.046***	0.026***	0.060***
	(0.007)	(0.006)	(0.018)
$\text{Return}_{j,t-3}$	0.029***	0.020***	0.023
	(0.007)	(0.006)	(0.016)
$\text{Return}_{j,t-4}$	0.023***	0.011**	0.023*
	(0.008)	(0.005)	(0.013)
$\text{Return}_{j,t-26 \to t-5}$	0.059***	0.050***	0.092***
	(0.016)	(0.011)	(0.027)
F-stat ($\text{Return}_{j,t-26 \to t-5}$)	13.5***		
p-value	0.0004		
Fund FE	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes
Observations	97,354	77,739	25,652
Adj. R-squared	0.035	0.034	0.248

TABLE 5

ETF Index Arbitrage and Asset Market Quality: IV Estimates

The table shows instrumental variables estimates of the effects of ETF index arbitrage on individual stocks' market quality. Effective Spread$_{i,t}$ is the share-weighted and measured as percentage of the price, Pricing Error$_{i,t}$ is based on Hasbrouck (1993), Volatility$_{i,t}$ is the intraday volatility based on 1-minute returns of stock i, and Correlation$_{i,t}$ is the intraday correlation between the 1-minute returns of stock i and the SPDR S&P 500 ETF. The main independent variable is the instrumented magnitude of ETF index arbitrage, standardized to a standard deviation of 1. Coefficients and standard errors are scaled by 100 to display percent changes. Variable definitions can be found in Appendix A. The sample is daily from 2015 to 2019. Standard errors in parentheses are robust and clustered by stock and day. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A: Sort by Past Liquidity

	(1) Effective Spread$_{i,t}$	(2) Pricing Error$_{i,t}$	(3) Volatility$_{i,t}$	(4) Correlation$_{i,t}$
$\lvert\widehat{\text{APArb}}_{i,t}\rvert \times \text{Liquid}^1_{i,t}$ (Most Liquid)	2.31***	1.91***	2.48***	0.56**
	(0.36)	(0.46)	(0.38)	(0.26)
$\lvert\widehat{\text{APArb}}_{i,t}\rvert \times \text{Liquid}^2_{i,t}$	1.04***	0.60**	1.51***	0.22
	(0.26)	(0.30)	(0.24)	(0.13)
$\lvert\widehat{\text{APArb}}_{i,t}\rvert \times \text{Liquid}^3_{i,t}$ (Least Liquid)	-0.96***	-0.57	1.04***	0.14
	(0.30)	(0.44)	(0.28)	(0.13)
Stock-level Controls	Yes	Yes	Yes	Yes
Tercile Main Effects	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Observations	2,521,966	2,522,319	2,522,316	2,522,316
Adj. R-squared	0.828	0.806	0.744	0.627
$\beta^1 - \beta^3$	3.27***	2.48***	1.44***	0.42**
F-stat	95.92	17.84	22.99	3.79
Unconditional Treatment Effect	0.75***	0.56*	1.60***	0.27*
	(0.26)	(0.30)	(0.26)	(0.15)

Panel B: Sort by Past Relative Weight (RW)

	(1) Effective Spread$_{i,t}$	(2) Pricing Error$_{i,t}$	(3) Volatility$_{i,t}$	(4) Correlation$_{i,t}$
$\lvert\widehat{\text{APArb}}_t\rvert \times \text{RW}^1_{i,t}$ (Highest RW)	1.14***	1.46***	1.93***	0.20
	(0.43)	(0.46)	(0.43)	(0.26)
$\lvert\widehat{\text{APArb}}_t\rvert \times \text{RW}^2_{i,t}$	0.81*	0.85*	1.66***	0.23
	(0.42)	(0.44)	(0.42)	(0.24)
$\lvert\widehat{\text{APArb}}_t\rvert \times \text{RW}^3_{i,t}$ (Lowest RW)	0.02	0.72	1.41***	0.20
	(0.42)	(0.44)	(0.41)	(0.22)
Stock-level Controls	Yes	Yes	Yes	Yes
Tercile Main Effects	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Observations	2,460,989	2,461,338	2,461,338	2,461,338
Adj. R-squared	0.825	0.804	0.743	0.624
$\beta^1 - \beta^3$	1.13***	0.74***	0.53***	-0.01
F-stat	(26.78)	(5.61)	(12.90)	(0.00)
Unconditional Treatment Effect	0.65	1.00**	1.66***	0.21
	(0.41)	(0.42)	(0.41)	(0.24)

TABLE 6

ETF Index Arbitrage and Stock Returns: IV Estimates

The table shows instrumental variables estimates of the effects of ETF index arbitrage on individual stocks' daily returns. $\text{Return}_{i,t}$ is the return to stock i on day t. The main independent variable is the instrumented magnitude of ETF index arbitrage, standardized to a standard deviation of 1. Variable definitions can be found in Appendix A. The sample is daily from 2015 to 2019. Standard errors in parentheses are robust and clustered by stock and day. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1) $\text{Return}_{i,t-2}$	(2) $\text{Return}_{i,t-1}$	(3) $\text{Return}_{i,t}$	(4) $\text{Return}_{i,t+1}$	(5) $\text{Return}_{i,t+2}$
$\lvert\widehat{\text{APArb}}_{i,t}\rvert \times \text{Liquid}^1_{i,t}$ (Most Liquid)	0.0002	-0.0001	-0.0004	0.0004	0.0004
	(0.0004)	(0.0004)	(0.0003)	(0.0004)	(0.0004)
$\lvert\widehat{\text{APArb}}_{i,t}\rvert \times \text{Liquid}^2_{i,t}$	-0.0000	-0.0002	-0.0004	0.0005	0.0005
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
$\lvert\widehat{\text{APArb}}_{i,t}\rvert \times \text{Liquid}^3_{i,t}$ (Least Liquid)	-0.0001	-0.0005	-0.0005*	0.0005	0.0006*
	(0.0004)	(0.0003)	(0.0003)	(0.0004)	(0.0003)
Stock-level Controls	Yes	Yes	Yes	Yes	Yes
Tercile Main Effects	Yes	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes
Observations	2,522,290	2,522,294	2,522,319	2,521,762	2,521,208
Adj. R-squared	0.016	0.015	0.013	0.015	0.013
$\beta^1 - \beta^3$	0.0003*	0.0004**	0.0002	-0.0001	-0.0002
F-stat	3.04	4.61	1.19	0.52	1.68
Unconditional Treatment Effect	0.0000	-0.0003	-0.0004	0.0005	0.0005
	(0.0004)	(0.0003)	(0.0003)	(0.0003)	(0.0003)

TABLE 7

ETF Index Arbitrage and Asset Market Quality: Controlling for HFT

The table repeats the analyses in Table 5 with contemporaneous controls for the intensity of high-frequency trading, $\text{TOR}_{i,m}$, measured as the trade-to-order ratio of stock i in the same month following Weller (2018), standardized to a standard deviation of 1. The main independent variable is the instrumented magnitude of ETF index arbitrage, standardized to a standard deviation of 1. Coefficients and standard errors are scaled by 100 to display percent changes. Standard errors in parentheses are robust and clustered by stock and day. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1) Effective Spread$_{i,t}$	(2) Pricing Error$_{i,t}$	(3) Volatility$_{i,t}$	(4) Correlation$_{i,t}$
$\lvert\widehat{\text{APArb}}_{i,t}\rvert \times \text{Liquid}^1_{i,t}$ (Most Liquid)	2.34***	1.89***	2.49***	0.60**
	(0.36)	(0.46)	(0.38)	(0.25)
$\lvert\widehat{\text{APArb}}_{i,t}\rvert \times \text{Liquid}^2_{i,t}$	1.05***	0.61**	1.51***	0.23*
	(0.26)	(0.30)	(0.24)	(0.13)
$\lvert\widehat{\text{APArb}}_{i,t}\rvert \times \text{Liquid}^3_{i,t}$ (Least Liquid)	-0.97***	-0.58	1.04***	0.13
	(0.31)	(0.44)	(0.28)	(0.13)
$\text{TOR}_{i,m} \times \text{Liquid}^1_{i,t}$ (Most Liquid)	-2.66***	-1.22**	-0.20	-2.35***
	(0.37)	(0.48)	(0.41)	(0.11)
$\text{TOR}_{i,m} \times \text{Liquid}^2_{i,t}$	-1.96***	-1.65***	0.20	-0.85***
	(0.37)	(0.37)	(0.34)	(0.05)
$\text{TOR}_{i,m} \times \text{Liquid}^3_{i,t}$ (Least Liquid)	-0.85**	-2.28***	0.00	-0.23***
	(0.41)	(0.48)	(0.28)	(0.05)
Stock-level Controls	Yes	Yes	Yes	Yes
Tercile Main Effects	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Observations	2,521,687	2,522,040	2,522,037	2,522,037
Adjusted R-squared	0.829	0.806	0.744	0.630
$\beta^1 - \beta^3$	3.31***	2.48***	1.45***	0.48***
F-stat	99.86	17.57	23.29	5.31
Unconditional Treatment Effect	0.75***	0.56*	1.60***	0.27*
	(0.26)	(0.30)	(0.26)	(0.15)

TABLE 8

ETF Index Arbitrage and Asset Market Quality: Controlling for Market Structure

The table repeats the analyses in Table 5 with contemporaneous controls for market structure, $\text{HHI}_{i,m}$, measured as the Herfindahl–Hirschman Index on trading volume across venues for stock i in the same month, standardized to a standard deviation of 1. The main independent variable is the instrumented magnitude of ETF index arbitrage, standardized to a standard deviation of 1. Coefficients and standard errors are scaled by 100 to display percent changes. Standard errors in parentheses are robust and clustered by stock and day. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1) Effective Spread$_{i,t}$	(2) Pricing Error$_{i,t}$	(3) Volatility$_{i,t}$	(4) Correlation$_{i,t}$
$\lvert\widehat{\text{APArb}}_{i,t}\rvert \times \text{Liquid}^1_{i,t}$ (Most Liquid)	2.50***	1.95***	2.51***	0.58**
	(0.36)	(0.46)	(0.38)	(0.26)
$\lvert\widehat{\text{APArb}}_{i,t}\rvert \times \text{Liquid}^2_{i,t}$	1.07***	0.62**	1.51***	0.20
	(0.26)	(0.30)	(0.24)	(0.13)
$\lvert\widehat{\text{APArb}}_{i,t}\rvert \times \text{Liquid}^3_{i,t}$ (Least Liquid)	-1.05***	-0.59	1.02***	0.13
	(0.30)	(0.44)	(0.28)	(0.13)
$\text{HHI}_{i,m} \times \text{Liquid}^1_{i,t}$ (Most Liquid)	-0.25	-0.10	1.76***	1.23***
	(0.43)	(0.53)	(0.27)	(0.13)
$\text{HHI}_{i,m} \times \text{Liquid}^2_{i,t}$	-2.92***	0.04	1.55***	0.58***
	(0.35)	(0.35)	(0.21)	(0.05)
$\text{HHI}_{i,m} \times \text{Liquid}^3_{i,t}$ (Least Liquid)	-3.75***	-1.10***	-0.02	0.25***
	(0.36)	(0.30)	(0.17)	(0.05)
Stock-level Controls	Yes	Yes	Yes	Yes
Tercile Main Effects	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Observations	2,521,687	2,522,040	2,522,037	2,522,037
Adjusted R-squared	0.829	0.806	0.744	0.627
$\beta^1 - \beta^3$	3.55***	2.54***	1.49***	0.45**
F-stat	117.37	18.68	24.41	4.40
Unconditional Treatment Effect	0.78***	0.57*	1.59***	0.27*
	(0.26)	(0.30)	(0.26)	(0.15)

TABLE 9

ETF Index Arbitrage and Asset Market Quality: Controlling for Market News

The table repeats the analyses in Table 5 with contemporaneous controls for market-wide news arrival, $|\text{Market Return}_t|$, measured as the absolute value of with-distribution return of CRSP Value-Weighted Index at day t, standardized to a standard deviation of 1. The main independent variable is the instrumented magnitude of ETF index arbitrage, standardized to a standard deviation of 1. Coefficients and standard errors are scaled by 100 to display percent changes. Standard errors in parentheses are robust and clustered by stock and day. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1) Effective Spread$_{i,t}$	(2) Pricing Error$_{i,t}$	(3) Volatility$_{i,t}$	(4) Correlation$_{i,t}$
$\lvert\widehat{\text{APArb}}_{i,t}\rvert \times \text{Liquid}^1_{i,t}$ (Most Liquid)	2.19***	1.58***	2.35***	0.42*
	(0.36)	(0.43)	(0.36)	(0.25)
$\lvert\widehat{\text{APArb}}_{i,t}\rvert \times \text{Liquid}^2_{i,t}$	1.05***	0.68**	1.55***	0.23*
	(0.26)	(0.30)	(0.23)	(0.13)
$\lvert\widehat{\text{APArb}}_{i,t}\rvert \times \text{Liquid}^3_{i,t}$ (Least Liquid)	-0.85***	-0.30	1.14***	0.26**
	(0.30)	(0.43)	(0.27)	(0.13)
$\lvert\text{Market Return}_t\rvert \times \text{Liquid}^1_{i,t}$ (Most Liquid)	2.63***	6.49***	4.64***	1.90***
	(0.42)	(0.61)	(0.50)	(0.29)
$\lvert\text{Market Return}_t\rvert \times \text{Liquid}^2_{i,t}$	1.55***	2.89***	3.11***	0.61***
	(0.36)	(0.50)	(0.41)	(0.20)
$\lvert\text{Market Return}_t\rvert \times \text{Liquid}^3_{i,t}$ (Least Liquid)	0.71*	1.26***	2.61***	-0.30**
	(0.39)	(0.46)	(0.37)	(0.14)
Stock-level Controls	Yes	Yes	Yes	Yes
Tercile Main Effects	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Observations	2,521,966	2,522,319	2,522,316	2,522,316
Adjusted R-squared	0.829	0.807	0.747	0.631
$\beta^1 - \beta^3$	3.05***	1.88**	1.21***	0.16
F-stat	84.56	11.74	17.96	0.57
Unconditional Treatment Effect	0.75***	0.56*	1.60***	0.27*
	(0.26)	(0.30)	(0.25)	(0.15)

Appendix

A. Variable Definitions

Variable Names	Description
$1_{included,it}$	A dummy variable that equals 1 if stock i is included in the creation/redemption basket of Vanguard Total Market Index ETF (VTI) on day t and 0 otherwise.
Asset under Management$_{i,t}$ (ETF)	Price of ETF i times shares outstanding of ETF i on day t.
Bid-Ask Spread$_{i,t}$	Let i denote stock, t denote day, and s denote intraday time. The dollar-weighted percentage bid-ask spread of stock i on day t is calculated as following: $$\sum_s \frac{(\text{ask}_{its} - \text{bid}_{its})}{\text{price}_{its}} \cdot \frac{\text{price}_{its} \cdot \text{size}_{its}}{\sum_s \text{price}_{its} \cdot \text{size}_{its}}$$
Correlation$_{i,t}$	The correlation between the minute-to-minute returns of stock i and the minute-to-minute returns of S&P 500 Index on day t.
Correlation$_{i,t}$ (monthly)	The correlation between daily returns of stock i and the daily returns of the index that includes stock i in month t.
Effective Spread$_{i,t}$	Let i denote stock, t denote day, and s denote intraday time. The dollar-weighted percentage effective spread of stock i on day t is calculated as following: $$\sum_s \frac{2D_{its} \cdot (\text{price}_{its} - \text{midpoint}_{its})}{\text{midpoint}_{its}} \cdot \frac{\text{price}_{its} \cdot \text{size}_{its}}{\sum_s \text{price}_{its} \cdot \text{size}_{its}}$$ where D_{its} is a buy-sell indicator that equals 1 if the transaction is a buy and -1 if the transaction is a sell. Transactions are signed using the method in Lee and Ready (1991).
HHI$_{i,t}$	Let i denote stock, j denote exchange, and t denote month. Herfindahl-Hirschman Index is calculated monthly as following: $$\text{HHI}_{i,t} = \sum_j \left(\frac{\text{Trading Volume}_{ijt}}{\sum_j \text{Trading Volume}_{ijt}} \right)^2$$
Liquid$^q_{i,t}$	A dummy variable that equals 1 if stock i is included in the qth tercile as of previous month on day t and 0 otherwise.

Continued on next page

TABLE A1 – continued from previous page

Variable Definitions	Description
Market Capitalization$_{i,t}$	Closing price of stock *i* times shares outstanding of stock *i* on day *t*.
Pricing Error$_{i,t}$	The volatility of pricing error estimated using method in Hasbrouck (1993).
$\lvert$Primary Flow$_{j,t}\rvert$	The absolute value of the change in the shares outstanding of ETF *j* from day *t-1* to day *t* times closing price of ETF *j* on day *t*., divided by the lagged assets under management.
Return$_{j,t-k}$	The return of ETF *j* on day t-k.
Return$_{j,t-26\rightarrow t-5}$	The cumulative return of ETF *j* from day $t-26$ to $t-5$.
Retail Flow$_{j,t}$	Let *i* denote stock, *t* denote day, and *s* denote intraday time. The retail flow of ETF *j* on day *t* is calculated as following: $$\text{Retail Flow}_{j,t} = \frac{D_{jts} \cdot \text{size}_{jts}}{\text{Shares Outstanding}_{j,t}}$$ where D_{its} is a retail buy-sell indicator that equals 1 if the transaction is a buy and -1 if the transaction is a sell. Retail transactions are identified and signed using method in Boehmer et al. (2021).
Robinhood$_{j,t}$	The number of Robinhood users that held ETF *j* on day t, scaled by Shares Outstanding$_{j,t}$.
APArb$_{i,t}$	The amount of AP arbitrage activity in stock i on day t, calculated as the primary flow on day t for each ETF that posted stock i in its basket, multiplied by the basket weight as of the end of the previous month, scaled by the stock's market cap as of the end of the previous month.
TOR$_{i,t}$	Trade-to-order ratio, calculated as the trading volume of stock *i* on day *t* divided by the order volume of stock *i* on day *t*.
Turnover$_{i,t}$	Trading volume of stock *i* divided by shares outstanding of stock *i* on day *t*.
Volatility$_{i,t}$	The standard deviation of the minute-to-minute returns of stock i on day t.
Volatility$_{i,t}$ (monthly)	Standard deviation of daily return of stock *i* in month *t*.
Volume$_{i,t}$	Trading volume of stock i on day t.

B. A model of optimal ETF basket weights

In this subsection we present a model of the ETF provider's problem. The main prediction of the model is that the ETF provider strategically underweights index assets that are relatively illiquid. By contrast, the model predicts that index assets are almost never strategically overweighted. Intuitively, underweighting an asset increases tracking error but reduces transaction costs, while overweighting an asset increases both tracking error and transaction costs.

We consider an economy in which there are N assets and cash (indexed by zero) in the market with a vector of prices p (with $p_0 = 1$), and one-period excess returns $\tilde{r}$ (with $r_0 = 0$) which are normally distributed with expectation 0 and covariance matrix Σ. There are three types of agents: ETF providers, authorized participants, and investors. We consider the market for ETFs that track a specified index such as the S&P 500, the Russell 2000, or the CRSP value-weighted U.S. market index. The index is a vector of weights v (with cash weight $v_0 = 0$) that add to 1 and are exogenous and fixed.

An ETF provider enters the market by publishing a basket, which is a vector of weights w. She agrees to create or redeem shares of the ETF in exchange for that basket of individual assets. The cash weight w_0 does not need to be zero to reflect the fact that ETF providers allow authorized participants to create or redeem ETF shares in exchange for a basket of index assets plus some cash. The basket asset value (BAV) corresponding to one ETF share is BAV $= w'p$. The provider incurs administrative costs and collects a management fee. There is free entry, so in equilibrium ETF providers' fees equal their costs.

The ETF provider nominates one or more authorized participants (APs) who have access to the creation and redemption mechanism. The AP is risk neutral and makes the market for ETF shares by posting a bid price and an ask price around the BAV. The quotes are executed against by order flows from index investors. At the end of the period the AP nets the buy and sell orders that arrived within the period. She then closes out her netted position by trading in the individual index assets in the basket and making basket-ETF exchange with the provider.

The AP's expected profit from posting the offer BAV $+ b$ and being lifted is:

$$E[(\text{BAV} + b) - (w'p + C(w))] = b - E[C(w)],$$

where b is the spread and $\mathrm{E}[C(w)]$ is the expected transaction cost incurred by the netted order flow. As long as the provider nominates at least two authorized participants, if one posts an offer that is above the zero-profit bound, the other will undercut them. That is, competition between authorized participants is Bertrand. Thus, the bid and offer prices that investors face to trade the ETF shares are BAV $\pm E[C(w)]$.

In the competition among ETFs, all investors prefer an ETF that has a lower bid-ask spread and a lower tracking error. Thus, in equilibrium the ETF that captures the market is the one that minimizes

$$U = \mathrm{E}[C(w)] + \lambda(w - v)'\Sigma(w - v), \tag{3}$$

where λ is the shadow price that investors attach to a higher tracking error (term 2) relative to a higher bid-ask spread (term 1).[17]

The first order condition for the optimal weight of the fund in asset i is:

$$\frac{Cov(\tilde{r}_i, \tilde{r}_{ETF} - \tilde{r}_{Index})}{\partial \mathrm{E}[C(w)]/\partial w_i} = -1/2\lambda.$$

That is, the optimality condition for each asset is that the marginal increase in trading costs equals the marginal decrease in the expected tracking error, which is a product of the covariance of $\tilde{r}_i$ with the other index assets.

[17] We assume that all investors have the same preference λ. Relaxing this assumption would result in a frontier of ETFs that express different tradeoffs between tracking error and transaction costs and cater to investors with different preferences. In practice, the multiplicity of funds tracking the same index seems to be driven by investor search costs and not by different investor preferences (Hortaçsu and Syverson, 2004).

To solve for w explicitly, we specify the trading cost as linear and additively separable[18] and assume that index investors have independently normally distributed exogenous flow (or, effectively in a one-period model, the end-of-day inventory imbalance of the authorized participants) $\tilde{f} \sim N(0, \sigma_f^2)$ and that authorized participants liquidate their entire inventory every day:

$$\mathrm{E}[C(w)] = 2\mathrm{E}\left[\left|\tilde{f}\right|\right] \sum_i c_i w_i = 2\sqrt{\frac{2}{\pi}} \sum_i c_i w_i,$$

where c_i measures the trading cost of stock i.[19] The first order condition of asset i can be re-written as:

$$\frac{c_i}{2\lambda'} = \sum_j (v_j - w_j)\sigma_{ij}, \tag{4}$$

where $\lambda' = \lambda\sqrt{\pi/2}$ and σ_{ij} is the covariance of the returns between asset i and j, or the element on ith row and jth column in covariance matrix Σ. We immediately see that the cash weight is unconstrained, i.e., w_0 can take any positive value, since $c_0 = 0$ and $\sigma_{0j} = 0$. For index assets 1 to

[18] Almgren, Thum, Hauptmann, and Li (2005) propose that trading cost is exponential in order size for each asset, and using a large set of execution data they estimate the exponent to be 1.375. The model's predictions remain the same irrespective of the choice of the exponent.

[19] The inventory imbalance is linearly separable from the expected transaction costs and therefore can have a more generalized distribution, which leads to qualitatively the same functional form of the expected transaction costs up to a scalar. In addition, the authorized participants can choose to carry over their inventory across days, deviating from our stylized assumption of total liquidation of inventory. The cross-period optimization adds an additional tradeoff based on APs' expectation of the time series correlation of the day-to-day aggregate flow from the investors. This tradeoff is orthogonal to the cost-error tradeoff that we focus on in this paper. In sum, we assume normality without loss of generality and assume total liquidation for tractability.

N, it is easy to solve a symmetric system of N equations with N unknowns using Cramer's rule:

$$w_i = v_i - \frac{D_i}{D}$$

where D is the determinant of the covariance matrix Σ and D_i is the determinant of matrix Σ_i, which is the covariance matrix Σ with the ith column replaced by the vector $c_i/(2\lambda')$.

To show the intuition, consider the case with $N = 2$. We have:

$$v_2 - w_2 = \frac{c_2\sigma_{11} - c_1\sigma_{12}}{2\lambda'(\sigma_{11}\sigma_{22} - \sigma_{12}^2)}$$

in which the denominator is positive by the Cauchy-Schwarz inequality. When $c_2 \gg c_1$, i.e., asset 2 is *relatively* illiquid, $v_2 - w_2 > 0$, i.e., asset 2 is underweighted in the basket. In other words, the model predicts that the optimal basket weight w_i^* decreases as the index asset becomes more illiquid relative to other index assets.

Online Appendix to
ETF Sampling and Index Arbitrage

Jonathan Brogaard
David Eccles School of Business
University of Utah

Davidson Heath
David Eccles School of Business
University of Utah

Da Huang
D'Amore-McKim School of Business
Northeastern University

May 10, 2024

.

This online appendix presents results and details not in the main paper, "ETF Sampling and Index Arbitrage".

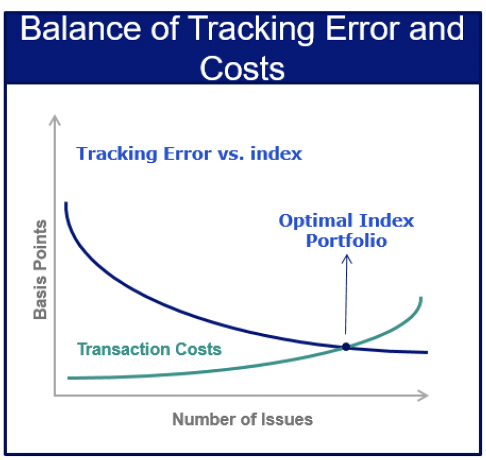
IA1. Broad Evidence of Equity ETF Sampling

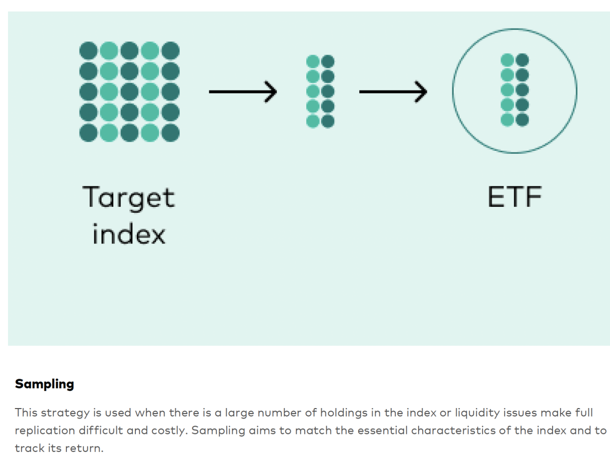
This section provides additional support for sampling behavior being widespread in equity ETFs. We show that both industry practitioners and regulators state that ETF sampling is an important feature of ETF operations.

1.1. Industry practitioners

In our conversations with several large ETF providers, they expressed support for our model and agreed that sampling is a common way to reduce transaction costs and improve liquidity for the ETFs they sponsor.

Below we reproduce two figures from industry white papers about ETF construction. The first is from Invesco, and illustrates the tradeoff between tracking error and transaction costs given the number of index constituents included in the ETF. The second is from Vanguard, and graphically shows that ETFs with many index assets or illiquid index assets sample their underlying index.





1.2. Regulators

The U.S. Securities and Exchange Commission (SEC) has referenced in their regulatory documents the longstanding practice of sampling in the ETF industry dating back to 2001. In a concept release titled "Actively Managed Exchange-Traded Funds"[1], the SEC gave a short introduction of how passive ETFs work:

However, because ETFs organized as open-end funds employ investment advisers, some of these ETFs instead may use "sampling strategies" to track the performance of an index. Using a sampling strategy, an investment adviser can construct a portfolio that is a subset of the component securities in the corresponding index, rather than a replication of the index.

Indeed, in the famous ETF rule (commonly known as rule 6c-11) passed by the SEC in 2019[2], the SEC also found it necessary to specify policies on ETF sampling:

Rule 6c-11 therefore will require an ETF to adopt and implement written policies and procedures that govern the construction of baskets and the process that will be used for the acceptance of baskets as proposed. These policies and procedures must cover the methodology that the ETF will use to construct baskets. For example, the policies and procedures should detail the circumstances under which the basket may omit positions that are not operationally feasible to transfer in kind. The policies and procedures also should detail when the ETF would use representative sampling of its portfolio to create its basket, and how the ETF would sample in those circumstances.

[1] https://www.sec.gov/rules-regulations/2001/11/actively-managed-exchange-traded-funds
[2] ff

IA2. An Alternate Measure of Sampling

The main paper presents results that are suggestive of widespread sampling behavior by many (though not all) of our sample ETFs, based on the $RelativeWeight$ measure of the ETF's basket weight divided by the index weight, for all index assets. However, $RelativeWeight$ could potentially present an incomplete picture of sampling because it is more driven by the index assets with the smallest index weight (and therefore a smaller denominator in the measure).

To address this concern, we compute an alternative measure of sampling behavior which we term $WeightedDiff$. This alternative measure is important because it accounts for the relative significance of each stock within the index. Stocks with larger index weights will have a greater impact on the measure, aligning the evaluation more closely with the actual market influence of each stock. This approach reduces the disproportionate influence of stocks with minimal index weights and provides a more accurate reflection of how the ETF's composition deviates from its benchmark index.

We first compute the deviation between the basket weight of stock i in ETF j in month m from its weight in the index tracked by ETF j in month m. Call this $WeightDiff_{i,j,m}$. We then weight this difference by the weight of the stock in the index to obtain

$$WeightedDiff_{i,j,m} = WeightDiff_{i,j,m} * w_{i,j,m} \tag{IA1}$$

and sum this weighted difference across all stocks in ETF j to obtain a measure of the deviation of the weights in ETF j. We then average across all monthly observations t for that ETF. If the ETF basket overall tends to underweight index assets, the average WeightedDiff should be negative.

Table IA1 presents the results. We see that WeightedDiff is negative on average for the vast majority of sample ETFs (52 out of 56), suggesting that sampling and in particular underweighting of index assets is widespread.

How should we assess the magnitude of this index weighted difference between the basket and the index? Consider an index with 90 large assets that each have 1% index weight and 100 small assets that each have a 0.1% index weight. If the ETF posts a 1% basket weight for each large asset, and completely omits all small assets, this basket design would result in an average WeightedDiff of (90*(0.01-0.01)*0.01 + 100*(0-0.001)*0.001) = -0.0001.

36 out of 56 ETFs have an average WeightedDiff below this -0.0001 threshold. Put differently, 36 out of 56 ETFs are sampling as aggressively, or more, than a fund that omits all index assets with weight below 0.1%.

We emphasize that the fictitious index we refer to is realistic – it has the bottom half of index assets making up 10% of the total index value. For comparison, the Russell 3000 has its bottom 2500 stocks making about 15% of the total index value.

We emphasize that WeightedDiff should not be interpreted as the deviation (or 1 minus fidelity) from full replication. To estimate the fidelity of index replication, one should refer to the RelativeWeight measure that we report in the main paper, Table 2. The WeightedDiff measure is most of all a robustness check on RelativeWeight, to mitigate the possibility of index assets with trivial weights driving the results.

Table IA1
WeightedDiff Measure of Sampling

The table lists ETFs in the sample and their average WeightedDiff as defined above. WeightedDiff is calculated as a stock's basket weight minus its index weight, multiplied by its index weight, summed across all index assets.

Ticker	Avg WeightedDiff
VO	-0.001412
NOBL	-0.001354
SDY	-0.001235
HDV	-0.001139
FVD	-0.001138
DVY	-0.001124
SPHD	-0.001106
VBK	-0.001082
VBR	-0.001001
MTUM	-0.000950
SPLV	-0.000892
QUAL	-0.000846
IJT	-0.000746
IJS	-0.000728
DGRO	-0.000726
GSLC	-0.000718
IUSG	-0.000595
VOE	-0.000528
FNDA	-0.000469
VOT	-0.000464
IUSV	-0.000444
VYM	-0.000411
USMV	-0.000407
VB	-0.000401
SCHG	-0.000367
IVW	-0.000357
VV	-0.000354
SCHA	-0.000314
SCHV	-0.000290

Continued on next page

Table IA1 – continued from previous page

Ticker	Avg WeightedDiff
PRF	-0.000286
VTI	-0.000240
FNDX	-0.000233
IVE	-0.000230
VTV	-0.000185
IJR	-0.000129
IJH	-0.000106
IWD	-0.000080
VUG	-0.000070
IWN	-0.000068
IWF	-0.000057
IWV	-0.000048
IWS	-0.000046
IWB	-0.000045
ITOT	-0.000043
SCHB	-0.000041
IVV	-0.000038
IWO	-0.000034
IWM	-0.000029
IWP	-0.000024
IWR	-0.000021
XLK	-0.000019
MDY	-0.000008
SPY	0.000005
VOO	0.000005
MGK	0.000063
SCHX	0.000080

IA3. Effects on ETF Markets

In this section we examine the relation between instrumented ETF arbitrage activity and market quality for the ETF itself. One important question for our instrumental variables design is whether larger ETF primary flows have effects on underlying asset markets because of the effects of ETF arbitrage (as our evidence suggests), or whether larger predicted ETF primary flows contain or correlate with information that is relevant to market quality.

If *predicted* ETF flows impact market quality or returns on the ETF itself, then the information could be not that stale but contain market-relevant information. As one way to test for the existence of the alternative channels, we investigate if the instrument predicts ETF mispricing, market volatility, or fund returns. In Table IA2 we rerun our baseline estimates of the effects of instrumented ETF primary flow on measures of 1) market quality and 2) returns, for the ETFs themselves.

Table IA2 Column 1 shows that on days where the (predicted) ETF primary flow is larger, there is a statistically significant increase in that ETF's daily price volatility. This is consistent with some level of price impact. However, while this effect is statistically significant, it is economically tiny: A one standard deviation higher predicted ETF primary flow results in ETF volatility that is 0.2 basis points higher, less than 1/20 of one sample standard deviation. Reflecting the relation's economic insignificance, the incremental R-squared that the IV contributes is also tiny (0.002 relative to the overall model R^2 of 0.155).

Table IA2 Columns 2 and 3 investigate other measures of ETF market quality, in particular, ETF liquidity and Hasbrouck pricing error on the day. We see that the relationship is economically and statistically zero, that is, there is no effect on the market for the ETF itself. Columns 4-6 investigate how our instrumental variable relates to ETF returns. We see that our IV has zero significant relationship with lagged, contemporaneous and future ETF returns.

In sum, higher predicted ETF primary flows are statistically associated with higher price volatility for the fund on that day, but the economic significance is tiny. For other measures of ETF market quality and also ETF returns, the relationship is statistically and economically zero. Thus, we conclude that it is unlikely that lagged ETF returns – and thus predicted ETF primary flows – pick up omitted variables related to market conditions or market-relevant information.

However, it is still a potential concern if changes in ETF market quality explain some part of our results, in which case the exclusion restriction would be violated. To examine this possibility we repeat our main estimates controlling for contemporaneous variation in ETF market quality. That is, we take the measures of volatility and effective spread (outcomes in columns 1 and 3 in Table IA2) for every ETF-day in the sample. We multiply the ETF-day measures with the ETF basket weights and sum them for each stock on that day, in parallel with the calculation of our predicted primary flow measure $\widehat{\text{APArb}}_{i,t}$. We then add the weighted stock-by-day measures of ETF market quality as control variables to our main estimates to see if they explain any part of our results.

Table IA3 shows the results. From the coefficients on the control variables, we see that variation in ETF market quality is strongly related to market quality in the underlying stocks. Nevertheless our main estimates are effectively unchanged. While the exclusion restriction is in principle untestable, this result provides some reassurance that variation in ETF market quality is not driving a significant part of our findings.

Table IA2
ETF arbitrage and ETF market quality

The table presents regressions of ETF market quality on the magnitude of daily predicted arbitrage activity in that ETF. $|\widehat{\text{Primary Flow}}_{j,t}|$ is the predicted magnitude of each ETF j's primary flow on day t, standardized to a standard deviation of 1. The sample is fund-by-day from January 2015 to December 2019. Standard errors in parentheses are robust and clustered by fund and day. Incremental adjusted R^2 is the adjusted R-squared that is contributed by the independent variable $|\widehat{\text{Primary Flow}}_{j,t}|$. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
	$\text{Volatility}_{j,t}$	$\text{Pricing Error}_{j,t}$	$\text{Effective Spread}_{j,t}$	$\text{Return}_{j,t-1}$	$\text{Return}_{j,t}$	$\text{Return}_{j,t+1}$
$\lvert\widehat{\text{Primary Flow}}_{j,t}\rvert$	0.00002***	0.00000	0.00001	0.0004	0.0002	0.0002
	(0.00001)	(0.00000)	(0.00000)	(0.0002)	(0.0002)	(0.0002)
Incremental Adj. R^2	0.002	0.000	0.000	0.001	0.000	0.000
Fund FE	Yes	Yes	Yes	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	99,547	99,547	99,547	98,349	99,547	98,042
Adjusted R-squared	0.155	0.041	0.136	0.032	0.026	0.030

Table IA3
Main estimates controlling for contemporaneous ETF market quality

The table shows instrumental variables estimates of the effects of ETF index arbitrage on individual stocks' market quality, controlling for contemporaneous variation in ETF market quality (intraday volatility and liquidity), multiplied by the basket weights per ETF and summed for each stock-day. Variable definitions can be found in Appendix **??**. The sample is daily from 2015 to 2019. Standard errors in parentheses are robust and clustered by stock and day. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1) Effective Spread$_{i,t}$	(2) Pricing Error$_{i,t}$	(3) Volatility$_{i,t}$	(4) Correlation$_{i,t}$
$\lvert\widehat{\text{APArb}}_{i,t}\rvert \times \text{Liquid}^1_{i,t}$ (Most Liquid)	2.11***	1.71***	2.20***	0.48*
	(0.34)	(0.44)	(0.34)	(0.25)
$\lvert\widehat{\text{APArb}}_{i,t}\rvert \times \text{Liquid}^2_{i,t}$	0.85***	0.39	1.18***	0.10
	(0.25)	(0.29)	(0.23)	(0.13)
$\lvert\widehat{\text{APArb}}_{i,t}\rvert \times \text{Liquid}^3_{i,t}$ (Least Liquid)	-1.08***	-0.73*	0.77***	0.01
	(0.30)	(0.43)	(0.26)	(0.13)
$\sum Basket_{ijt} \times ETFVolatility_{i,t}$	7,258.96***	10,624.64***	16,755.21***	6,729.91***
	(2,051.34)	(1,346.69)	(3,629.02)	(1,883.57)
$\sum Basket_{ijt} \times ETFSpread_{i,t}$	8,003.57***	5,634.43***	4,762.15	-1,084.34
	(2,317.81)	(1,977.60)	(2,955.99)	(1,436.07)
Stock-level Controls	Yes	Yes	Yes	Yes
Tercile Main Effects	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Observations	2,496,569	2,496,920	2,496,920	2,496,920
Adjusted R-squared	0.829	0.805	0.750	0.634
$\beta^1 - \beta^3$	3.20***	2.44***	1.43***	0.47**
F-stat	94.74	17.52	24.52	5.07

IA4. End of Day Estimates

This section repeats the analyses in table **??** using only the last 30 minutes to calculate the market quality as the outcome variable. We skip the pricing error because the fewer observations make the calculation infeasible. We see that the results are qualitatively the same compared to table **??**. The point estimates are generally larger and the differential effect is present, albeit statistically weaker in the case of intraday correlation.

Table IA4
ETF primary flows and asset market quality at the end of day: IV estimates

The table repeats the analyses of columns (1), (3), and (4) in table **??** in the main paper, with the dependent variables calculated using only the last 30 minutes of the trading day. Coefficients and standard errors are scaled by 100 to display percent changes. Variable definitions can be found in Appendix **??**. The sample is daily from 2015 to 2019. Standard errors in parentheses are robust and clustered by stock and day. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1) Effective Spread$_{i,t}$	(2) Volatility$_{i,t}$	(3) Correlation$_{i,t}$
$\lvert\widehat{\text{Primary Flow}}_{i,t}\rvert \times \text{Liquid}^1_{i,t}$ (Most Liquid)	3.32***	4.51***	0.78**
	(0.42)	(0.62)	(0.34)
$\lvert\widehat{\text{Primary Flow}}_{i,t}\rvert \times \text{Liquid}^2_{i,t}$	0.77***	1.16***	0.67***
	(0.29)	(0.35)	(0.23)
$\lvert\widehat{\text{Primary Flow}}_{i,t}\rvert \times \text{Liquid}^3_{i,t}$ (Least Liquid)	-2.28***	-1.27***	0.47**
	(0.39)	(0.38)	(0.22)
Stock-level Controls	Yes	Yes	Yes
Tercile Main Effects	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes
Month FE	Yes	Yes	Yes
Observations	2,521,092	2,521,058	2,521,058
Adj. R-squared	0.853	0.672	0.413
$\beta^1 - \beta^3$	5.60***	5.78***	0.31
F-stat	109.78	129.78	1.42
Unconditional Treatment Effect	0.47*	1.22***	0.63**
	(0.27)	(0.38)	(0.24)

IA5. Effects on Trading Activity

This section examines the relationship between ETF primary flows and trading activity in individual stocks. Figure **??** in the main paper suggests that APs absorb intraday flows and trade in the underlying assets near the end of the trading day. As a result, on days with more ETF index arbitrage, end-of-day trading activity in liquid stocks increases relative to their mid-day trading activity.

To formally test whether ETF primary flows cause a change in trading activity during the trading day, we examine the effects on intraday share turnover. We compute the share turnover for each stock for each day divided into 10 minute time-of-day blocks. $\text{TurnoverClose}_{it}$ is share turnover in the last ten minutes (15:50 to 16:00), while TurnoverMid_{it} is the average turnover per ten minutes from 9:40 to 15:50. $\text{Turnover(Close-Mid)}_{i,t}$ is the difference between the two on for stock i on day t.

Table IA5 displays estimates of the causal effects of ETF index arbitrage on trading activity. The results show clear differential effects of predicted primary flow on stock-level share turnover. Column 1 shows that a one standard deviation larger ETF primary flow is associated with a 2.5 standard deviation higher turnover for stocks in the most liquid tercile and a 2 standard deviation higher turnover for stocks in the least liquid tercile. The results are intuitive – when the stock is subject to higher ETF arbitrage activities, it sees a higher share turnover. However, we see statistically insignificant difference between the effects in liquid and illiquid terciles.

The picture is clearer when we examine changes in intraday trading activity. In Table IA5 Column 2 the dependent variable is the amount of trading activity in each stock in the last 10 minutes of trading on day t, relative to the average trading activity per 10 minutes over the rest of the day. Thus, a positive coefficient reflects a relative delay and pooling of orders intraday, away from the midday and toward the close. The differential effect between the most liquid and least liquid stocks is very clear – ETF arbitrage activity has a much stronger effect in liquid tercile to shift trading activities from mid-day to the last 10 minutes of the trading day. The null hypothesis that the effect is equal between the liquid and illiquid terciles is rejected with a Wald test (a difference of 1.2 standard deviation change with an F-stat=4.2 and a p-value < 0.05). The differential effect is consistent with the fact that ETF basket overweight liquid stocks and underweight illiquid stocks.

Table IA5
ETF primary flows and asset trading activity: IV estimates

The table presents regressions of trading activity in individual stocks on the magnitude of ETF primary flow for stock i. $\text{Liquid}^q_{i,t}$ equals one if stock i is in liquidity tercile q the prior month and zero otherwise. Tercile 1 contains the most liquid stocks and tercile 3 contains the least liquid stocks. $\text{Turnover}_{i,t}$ is the trade volume of stock i on day t divided by its shares outstanding. $\text{Turnover(Close-Mid)}_{i,t}$ is the difference between the trade volume in the last 10 minutes of the day and the average 10-minute trade volume from 9:40 to 15:50, divided by shares outstanding and standardized to a standard deviation of 1. $|\widehat{\text{Primary Flow}}_{i,t}|$ is the instrumented ETF primary flow distributed through the ETF basket weights, standardized to a standard deviation of 1. Variable definitions can be found in Appendix **??**. The sample is stock-by-day from January 2015 to September 2019. Standard errors in parentheses are robust and clustered by stock and day. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1) $\text{Turnover}_{i,t}$	(2) $\text{Turnover(Close-Mid)}_{i,t}$
$\lvert\widehat{\text{Primary Flow}}_{i,t}\rvert \times \text{Liquid}^1_{i,t}$ (Most Liquid)	2.49***	3.40***
	(0.51)	(0.63)
$\lvert\widehat{\text{Primary Flow}}_{i,t}\rvert \times \text{Liquid}^2_{i,t}$	1.99***	2.05***
	(0.47)	(0.55)
$\lvert\widehat{\text{Primary Flow}}_{i,t}\rvert \times \text{Liquid}^3_{i,t}$ (Least Liquid)	2.04***	2.18***
	(0.52)	(0.57)
Stock-level Controls	Yes	Yes
Tercile Controls	Yes	Yes
Stock FE	Yes	Yes
Month FE	Yes	Yes
Observations	2,522,319	2,522,317
Adj. R-squared	0.455	0.387
$\beta^1 - \beta^3$	0.44	1.22**
F-stat	0.69	4.20
Unconditional Treatment Effect	2.12***	2.39***
	(0.42)	(0.51)

IA6. Robustness of Fixed Effects

This section repeats the analyses in table **??** using daily fixed effects instead of the monthly fixed effects. We choose monthly fixed effects in the main table to avoid the "over-differencing" problem from using daily fixed effects. Daily fixed effects sweep out the market-wide day-to-day variation in liquidity, pricing error, volatility, and correlation with the index, which "re-centers" the estimates every day and makes the tercile-specific point estimates hard to interpret. Therefore, with daily fixed effects, the focus should be the differential effects between the most and least terciles. We see that the differential effects are economically and statistically significant and remain largely unchanged compared to the results in the main paper.

Table IA6
ETF index arbitrage and asset market quality: IV estimates

The table repeats the analyses in table **??** with daily fixed effects. Coefficients and standard errors are scaled by 100 to display percent changes. Variable definitions can be found in Appendix **??**. The sample is daily from 2015 to 2019. Standard errors in parentheses are robust and clustered by stock and day. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1) Effective Spread$_{i,t}$	(2) Pricing Error$_{i,t}$	(3) Volatility$_{i,t}$	(4) Correlation$_{i,t}$
$\lvert\widehat{\text{APArb}}_{i,t}\rvert \times \text{Liquid}^1_{i,t}$ (Most Liquid)	1.88***	1.05***	1.33***	0.12
	(0.25)	(0.40)	(0.21)	(0.15)
$\lvert\widehat{\text{APArb}}_{i,t}\rvert \times \text{Liquid}^2_{i,t}$	0.71***	-0.08	0.66***	-0.16***
	(0.18)	(0.27)	(0.11)	(0.06)
$\lvert\widehat{\text{APArb}}_{i,t}\rvert \times \text{Liquid}^3_{i,t}$ (Least Liquid)	-1.35***	-1.29***	0.09	-0.28***
	(0.26)	(0.45)	(0.18)	(0.10)
Stock-level Controls	Yes	Yes	Yes	Yes
Tercile Main Effects	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes
Daily FE	Yes	Yes	Yes	Yes
Observations	2,521,966	2,522,319	2,522,316	2,522,316
Adj. R-squared	0.839	0.810	0.770	0.714
$\beta^1 - \beta^3$	3.23***	2.33***	1.24***	0.40***
F-stat	102.94	17.01	21.87	4.41
Unconditional Treatment Effect	0.31*	-0.24	0.61***	-0.15**
	(0.17)	(0.28)	(0.12)	(0.06)

IA7. Separate Creation and Redemption Flows

This section repeats the analyses on stock returns but separates the creation flow and the redemption flow. In table **??** of the main paper, we find statistically and economically insignificant effect of ETF arbitrage activities on stock returns. But this could be due to treating creation and redemption flows in the same way when taking absolute value to calculate the arbitrage activity. We separate the two in the following way.

We run the same first stage specification of ETF primary flow on past returns and predict the ETF-level primary flow. We then keep only the positive (negative) predicted primary flows in the sample, distribute them through the basket weight matrix, and calculate an ETF arbitrage activity measure on stock level that comes from only creation (redemption) activity on ETF level. We then repeat the same specification in table **??** using the creation-specific and redemption-specific ETF arbitrage activity on stock level. Table IA7 reports the creation-specific results and table IA8 reports the redemption-specific results on stock returns.

We see that the effect of ETF arbitrage activities on stock returns is economically small even though statistically significant. The economic magnitude of the effect on liquid stocks is consistently smaller than 3 basis points, suggesting that there is little to no price impact or information content in the predicted ETF arbitrage activities. This evidence is especially useful since liquid stocks are where we find the main effect of interests concentrated. In addition, the unconditional effect is also economically insignificant and only barely statistically significant on a 5% level. The lack of economic significance suggests that the instrumented primary flows are stale and uninformed.

In sum, we emphasize that there is little to no effect of ETF arbitrage activities on returns of the liquid stocks, in which the effect of ETF arbitrage activities on market quality is concentrated, even after splitting the creation and redemption primary flows. The results further suggest that our instrument is unlikely to contain stock-specific fundamental information and the exclusion restriction is likely satisfied.

Table IA7
ETF creation primary flows and stock returns: IV estimates

The table repeats the analyses in table **??** using creation-specific ETF primary flows to calculate the ETF arbitrage activities on stock level. Variable definitions can be found in Appendix **??**. The sample is daily from 2015 to 2019. Standard errors in parentheses are robust and clustered by stock and day. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1) Return$_{i,t-2}$	(2) Return$_{i,t-1}$	(3) Return$_{i,t}$	(4) Return$_{i,t+1}$	(5) Return$_{i,t+2}$
$\lvert\widehat{\text{Primary Flow}}_{i,t}\rvert \times \text{Liquid}^1_{i,t}$ (Most Liquid)	-0.0002**	-0.0003***	-0.0002***	-0.0002**	-0.0001
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
$\lvert\widehat{\text{Primary Flow}}_{i,t}\rvert \times \text{Liquid}^2_{i,t}$	-0.0009**	-0.0013***	-0.0011**	-0.0009**	-0.0009**
	(0.0004)	(0.0005)	(0.0005)	(0.0004)	(0.0004)
$\lvert\widehat{\text{Primary Flow}}_{i,t}\rvert \times \text{Liquid}^3_{i,t}$ (Least Liquid)	-0.0018***	-0.0023***	-0.0019***	-0.0015**	-0.0014**
	(0.0006)	(0.0007)	(0.0007)	(0.0006)	(0.0006)
Stock-level Controls	Yes	Yes	Yes	Yes	Yes
Tercile Main Effects	Yes	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes
Observations	2,522,237	2,522,241	2,522,266	2,521,709	2,521,155
Adj. R-squared	0.016	0.015	0.013	0.015	0.013
$\beta^1 - \beta^3$	0.0016***	0.0020***	0.0016**	0.0014**	0.0013**
F-stat	7.14	8.09	6.52	5.09	5.05
Unconditional Treatment Effect	-0.0002*	-0.0003**	-0.0003**	-0.0002*	-0.0001
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)

Table IA8
ETF redemption primary flows and stock returns: IV estimates

The table repeats the analyses in table **??** using redemption-specific ETF primary flows to calculate the ETF arbitrage activities on stock level. Variable definitions can be found in Appendix **??**. The sample is daily from 2015 to 2019. Standard errors in parentheses are robust and clustered by stock and day. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1)	(2)	(3)	(4)	(5)
	$\text{Return}_{i,t-2}$	$\text{Return}_{i,t-1}$	$\text{Return}_{i,t}$	$\text{Return}_{i,t+1}$	$\text{Return}_{i,t+2}$
$\lvert\widehat{\text{Primary Flow}}_{i,t}\rvert \times \text{Liquid}^1_{i,t}$ (Most Liquid)	0.0003***	0.0003***	0.0002**	0.0002*	0.0001
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
$\lvert\widehat{\text{Primary Flow}}_{i,t}\rvert \times \text{Liquid}^2_{i,t}$	0.0025***	0.0026***	0.0019***	0.0016**	0.0016**
	(0.0009)	(0.0008)	(0.0007)	(0.0008)	(0.0008)
$\lvert\widehat{\text{Primary Flow}}_{i,t}\rvert \times \text{Liquid}^3_{i,t}$ (Least Liquid)	0.0038*	0.0040**	0.0034**	0.0031**	0.0032**
	(0.0020)	(0.0017)	(0.0015)	(0.0015)	(0.0016)
Stock-level Controls	Yes	Yes	Yes	Yes	Yes
Tercile Main Effects	Yes	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes
Observations	2,522,237	2,522,241	2,522,266	2,521,709	2,521,155
Adj. R-squared	0.017	0.015	0.013	0.015	0.013
$\beta^1 - \beta^3$	-0.0035*	-0.0036**	-0.0032**	-0.0029*	-0.0031*
F-stat	3.2112	4.5484	4.7973	3.8360	3.9815
Unconditional Treatment Effect	0.0003**	0.0003**	0.0002*	0.0002*	0.0002*
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)

IA8. OLS Estimates

This section presents OLS versions of the IV estimates presented in the main body of the paper. These estimates do not rely on our instrumental variables strategy of predicting ETF primary flow using the lagged prior-month return on the ETF. However, they are potentially confounded by omitted variables that drive both ETF primary flows and asset market quality. In each case, the results are consistent with the IV estimates presented in the main paper. This observation suggests that our conclusions in the paper are robust to the identification strategy.

Table IA9
ETF primary flows and asset market quality: OLS Estimates

The table presents the OLS version of the IV estimates in Table ?? in the main paper. Standard errors in parentheses are robust and clustered by stock and day. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1) Effective Spread$_{i,t}$	(2) Pricing Error$_{i,t}$	(3) Volatility$_{i,t}$	(4) Correlation$_{i,t}$
$\|\text{Primary Flow}_{i,t}\| \times \text{Liquid}^1_{i,t}$ (Most Liquid)	2.89***	3.55***	3.30***	0.98***
	(0.44)	(0.65)	(0.47)	(0.23)
$\|\text{Primary Flow}_{i,t}\| \times \text{Liquid}^2_{i,t}$	1.99***	0.97***	1.78***	0.32***
	(0.27)	(0.31)	(0.25)	(0.11)
$\|\text{Primary Flow}_{i,t}\| \times \text{Liquid}^3_{i,t}$ (Least Liquid)	1.05***	0.34	1.43***	0.29***
	(0.30)	(0.28)	(0.24)	(0.09)
Stock-level Controls	Yes	Yes	Yes	Yes
Tercile Controls	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Observations	2,521,966	2,522,319	2,522,316	2,522,316
Adj. R-squared	0.829	0.806	0.745	0.627
$\beta^1 - \beta^3$	1.85***	3.21***	1.88***	0.69***
F-stat	17.98	27.66	20.91	11.10
Unconditional Treatment Effect	1.74***	1.13***	1.89***	0.42***
	(0.27)	(0.30)	(0.25)	(0.11)

IA9. Using Daily-Level Controls

This section reports a robustness check of Tables **??** and **??** with control variables calculated on daily frequency. The results are qualitatively the same.

Table IA10
ETF primary flows and asset market quality: Controlling for high-frequency trading on daily frequency

The table repeats the analyses in Table **??** with trade-to-order ratio calculated on daily level. Standard errors in parentheses are robust and clustered by stock and day. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1) Effective Spread$_{i,t}$	(2) Pricing Error$_{i,t}$	(3) Volatility$_{i,t}$	(4) Correlation$_{i,t}$
$\lvert\widehat{\text{Primary Flow}}_{i,t}\rvert \times \text{Liquid}^1_{i,t}$ (Most Liquid)	2.32***	1.82***	2.38***	0.63***
	(0.36)	(0.47)	(0.39)	(0.23)
$\lvert\widehat{\text{Primary Flow}}_{i,t}\rvert \times \text{Liquid}^2_{i,t}$	1.05***	0.59**	1.48***	0.23*
	(0.25)	(0.30)	(0.25)	(0.12)
$\lvert\widehat{\text{Primary Flow}}_{i,t}\rvert \times \text{Liquid}^3_{i,t}$ (Least Liquid)	-0.97***	-0.55	1.06***	0.11
	(0.30)	(0.44)	(0.28)	(0.12)
$\text{TOR}_{i,d} \times \text{Liquid}^1_{i,t}$ (Most Liquid)	-2.03***	4.09***	7.01***	-6.57***
	(0.44)	(0.65)	(0.59)	(0.19)
$\text{TOR}_{i,d} \times \text{Liquid}^2_{i,t}$	-3.07***	-0.89*	5.70***	-3.17***
	(0.39)	(0.48)	(0.45)	(0.10)
$\text{TOR}_{i,d} \times \text{Liquid}^3_{i,t}$ (Least Liquid)	-1.42*	-2.23**	0.58*	-0.36*
	(0.78)	(1.13)	(0.31)	(0.19)
Stock-level Controls	Yes	Yes	Yes	Yes
Tercile Main Effects	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Observations	2,521,494	2,521,847	2,521,844	2,521,844
Adjusted R-squared	0.829	0.806	0.747	0.653
$\beta^1 - \beta^3$	3.28***	2.37***	1.31***	0.52***
F-stat	97.04	15.94	17.58	7.80
Unconditional Treatment Effect	0.75***	0.56*	1.60***	0.26*
	(0.26)	(0.30)	(0.26)	(0.14)

Table IA11
ETF primary flows and asset market quality: Controlling for market structure on daily frequency

The table repeats the analyses in Table **??** with the Herfindahl-Hirschman Index calculated on daily level. Standard errors in parentheses are robust and clustered by stock and day. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1) Effective Spread$_{i,t}$	(2) Pricing Error$_{i,t}$	(3) Volatility$_{i,t}$	(4) Correlation$_{i,t}$
$\lvert\widehat{\text{Primary Flow}}_{i,t}\rvert \times \text{Liquid}^1_{i,t}$ (Most Liquid)	2.38***	1.93***	2.53***	0.56**
	(0.36)	(0.46)	(0.38)	(0.25)
$\lvert\widehat{\text{Primary Flow}}_{i,t}\rvert \times \text{Liquid}^2_{i,t}$	1.04***	0.62**	1.54***	0.20
	(0.26)	(0.30)	(0.24)	(0.13)
$\lvert\widehat{\text{Primary Flow}}_{i,t}\rvert \times \text{Liquid}^3_{i,t}$ (Least Liquid)	-1.01***	-0.59	1.00***	0.13
	(0.30)	(0.44)	(0.28)	(0.13)
$\text{HHI}_{i,d} \times \text{Liquid}^1_{i,t}$ (Most Liquid)	1.30***	-0.37	0.30	1.61***
	(0.31)	(0.39)	(0.21)	(0.10)
$\text{HHI}_{i,d} \times \text{Liquid}^2_{i,t}$	-0.32	0.23	0.45***	0.64***
	(0.25)	(0.24)	(0.15)	(0.04)
$\text{HHI}_{i,d} \times \text{Liquid}^3_{i,t}$ (Least Liquid)	-0.95***	-0.85***	-1.24***	0.15***
	(0.19)	(0.17)	(0.10)	(0.03)
Stock-level Controls	Yes	Yes	Yes	Yes
Tercile Main Effects	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Observations	2,521,494	2,521,847	2,521,844	2,521,844
Adjusted R-squared	0.829	0.806	0.745	0.637
$\beta^1 - \beta^3$	3.38***	2.52***	1.53***	0.43**
F-stat	104.30	18.32	25.68	4.38
Unconditional Treatment Effect	0.75***	0.56*	1.60***	0.26*
	(0.26)	(0.30)	(0.26)	(0.14)

IA10. Alternative Measures of High-Frequency Trading

This section reports a robustness check of Table **??** using alternative measures of high-frequency trading (HFT) in Weller (2018). Specifically, we conduct the analyses using odd-lot ratio (OLR), cancel-to-trade ratio (CTR), and average trade size (ATS) in addition to trade-to-order ratio (TOR) used in the main paper. Tables IA12, IA13, and IA14 report the results. The effects of ETF arbitrage activities are qualitatively the same using different HFT measures.

Table IA12
ETF primary flows and asset market quality: Controlling for high-frequency trading using odd-lot ratio

The table repeats the analyses in Table ?? using odd-lot ratio to measure high-frequency trading intensity. Standard errors in parentheses are robust and clustered by stock and day. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1) Effective Spread$_{i,t}$	(2) Pricing Error$_{i,t}$	(3) Volatility$_{i,t}$	(4) Correlation$_{i,t}$
$\lvert \widehat{\text{Primary Flow}}_{i,t} \rvert \times \text{Liquid}^1_{i,t}$ (Most Liquid)	2.04***	1.82***	2.39***	0.57**
	(0.36)	(0.46)	(0.38)	(0.25)
$\lvert \widehat{\text{Primary Flow}}_{i,t} \rvert \times \text{Liquid}^2_{i,t}$	1.11***	0.71**	1.56***	0.21
	(0.26)	(0.30)	(0.24)	(0.13)
$\lvert \widehat{\text{Primary Flow}}_{i,t} \rvert \times \text{Liquid}^3_{i,t}$ (Least Liquid)	-0.71**	-0.55	1.10***	0.12
	(0.30)	(0.44)	(0.28)	(0.13)
$\text{OLR}_{i,m} \times \text{Liquid}^1_{i,t}$ (Most Liquid)	4.87***	8.78***	3.38***	0.56***
	(0.61)	(0.60)	(0.37)	(0.16)
$\text{OLR}_{i,m} \times \text{Liquid}^2_{i,t}$	1.92***	6.80***	2.12***	0.90***
	(0.52)	(0.52)	(0.33)	(0.07)
$\text{OLR}_{i,m} \times \text{Liquid}^3_{i,t}$ (Least Liquid)	0.36	6.47***	1.74***	0.92***
	(0.55)	(0.50)	(0.31)	(0.10)
Stock-level Controls	Yes	Yes	Yes	Yes
Tercile Main Effects	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Observations	2,521,634	2,521,987	2,521,984	2,521,984
Adjusted R-squared	0.829	0.806	0.745	0.636
$\beta^1 - \beta^3$	2.75***	2.38***	1.30***	0.45***
F-stat Fstat	67.34	15.93	17.62	4.75
Unconditional Treatment Effect	0.75***	0.57*	1.60***	0.27*
	(0.26)	(0.30)	(0.26)	(0.15)

Table IA13
ETF primary flows and asset market quality: Controlling for high-frequency trading using cancel-to-trade ratio

The table repeats the analyses in Table **??** using cancel-to-trade ratio to measure high-frequency trading intensity. Standard errors in parentheses are robust and clustered by stock and day. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1) Effective Spread$_{i,t}$	(2) Pricing Error$_{i,t}$	(3) Volatility$_{i,t}$	(4) Correlation$_{i,t}$
$\lvert\widehat{\text{Primary Flow}}_{i,t}\rvert \times \text{Liquid}^1_{i,t}$ (Most Liquid)	2.31***	1.83***	2.47***	0.60**
	(0.36)	(0.46)	(0.38)	(0.24)
$\lvert\widehat{\text{Primary Flow}}_{i,t}\rvert \times \text{Liquid}^2_{i,t}$	1.03***	0.58*	1.50***	0.22*
	(0.26)	(0.30)	(0.24)	(0.13)
$\lvert\widehat{\text{Primary Flow}}_{i,t}\rvert \times \text{Liquid}^3_{i,t}$ (Least Liquid)	-0.96***	-0.52	1.06***	0.10
	(0.30)	(0.44)	(0.28)	(0.13)
$\text{CTR}_{i,m} \times \text{Liquid}^1_{i,t}$ (Most Liquid)	1.21**	-0.52	-2.06***	2.70***
	(0.48)	(0.57)	(0.30)	(0.16)
$\text{CTR}_{i,m} \times \text{Liquid}^2_{i,t}$	0.51	0.82**	-1.16***	0.28***
	(0.44)	(0.38)	(0.24)	(0.08)
$\text{CTR}_{i,m} \times \text{Liquid}^3_{i,t}$ (Least Liquid)	0.57**	1.52***	-0.60***	-0.14**
	(0.29)	(0.44)	(0.14)	(0.06)
Stock-level Controls	Yes	Yes	Yes	Yes
Tercile Main Effects	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Observations	2,521,634	2,521,987	2,521,984	2,521,984
Adjusted R-squared	0.829	0.806	0.744	0.638
$\beta^1 - \beta^3$	3.27***	2.34***	1.41***	0.51***
F-stat	96.66	15.56	21.20	6.56
Unconditional Treatment Effect	0.75***	0.56*	1.60***	0.26*
	(0.26)	(0.30)	(0.26)	(0.14)

Table IA14
ETF primary flows and asset market quality: Controlling for high-frequency trading using average trade size

The table repeats the analyses in Table **??** using average trade size to measure high-frequency trading intensity. Standard errors in parentheses are robust and clustered by stock and day. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1) Effective Spread$_{i,t}$	(2) Pricing Error$_{i,t}$	(3) Volatility$_{i,t}$	(4) Correlation$_{i,t}$
$\lvert\widehat{\text{Primary Flow}}_{i,t}\rvert \times \text{Liquid}^1_{i,t}$ (Most Liquid)	2.27***	1.77***	2.32***	0.52**
	(0.36)	(0.46)	(0.38)	(0.25)
$\lvert\widehat{\text{Primary Flow}}_{i,t}\rvert \times \text{Liquid}^2_{i,t}$	1.09***	0.66**	1.56***	0.21
	(0.26)	(0.30)	(0.24)	(0.13)
$\lvert\widehat{\text{Primary Flow}}_{i,t}\rvert \times \text{Liquid}^3_{i,t}$ (Least Liquid)	-1.03***	-0.49	1.19***	0.17
	(0.31)	(0.44)	(0.28)	(0.13)
$\text{ATS}_{i,m} \times \text{Liquid}^1_{i,t}$ (Most Liquid)	2.41***	-1.66***	-4.67***	-0.93***
	(0.62)	(0.61)	(0.57)	(0.16)
$\text{ATS}_{i,m} \times \text{Liquid}^2_{i,t}$	3.65***	0.47	-2.76***	-0.99***
	(0.52)	(0.50)	(0.37)	(0.09)
$\text{ATS}_{i,m} \times \text{Liquid}^3_{i,t}$ (Least Liquid)	1.89***	0.73	-1.27***	-0.49***
	(0.50)	(0.55)	(0.31)	(0.10)
Stock-level Controls	Yes	Yes	Yes	Yes
Tercile Main Effects	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Observations	2,521,634	2,521,987	2,521,984	2,521,984
Adjusted R-squared	0.829	0.806	0.745	0.636
$\beta^1 - \beta^3$	3.31***	2.27***	1.13***	0.35**
F-stat	99.44	14.93	13.59	2.82
Unconditional Treatment Effect	0.74***	0.56*	1.60***	0.27*
	(0.26)	(0.30)	(0.26)	(0.14)