Signaling of ETF Flow: A Decomposition of Fundamental and Non-Fundamental Demand Shocks

Jinhwan Kim[a], Hoon Cho[a], Sangik Seok[b,*]

[a]College of Business, Korea Advanced Institute of Science and Technology, 85, Hoegi-ro, Dongdaemun-gu, Seoul 02455, Republic of Korea

[b]College of Business Administration, University of Ulsan, 93, Daehak-ro, Nam-gu, Ulsan 44610, Republic of Korea

*Corresponding Author. E-mail: siseok@ulsan.ac.kr

Highlights

- ETF flows reflect both fundamental and non-fundamental demand shocks
- The fundamental component of ETF flows shows short-term or insignificant return predictability
- The non-fundamental component of ETF flows exhibits long-term price reversals
- Fund flows in ETFs that track the same index exhibit low commonality on a daily basis

Abstract

This study introduces a novel approach to using exchange-traded funds (ETF) flows as a tool for identifying both fundamental and non-fundamental demand shocks in underlying asset markets. We propose a theoretical framework that decomposes ETF flows and emphasize the importance of this decomposition. In the empirical analysis, we investigate the return predictability of the decomposed flows to understand the term structure of demand shocks. Our findings reveal that while daily fundamental flows have either insignificant or negative return predictability over short-term, non-fundamental flows exhibit significantly negative predictability across various time horizons. Additionally, we observe greater commonality in weekly and monthly ETF flows compared to daily flows, suggesting that lower frequency measurements are more effective in capturing demand shocks. Overall, we demonstrate that ETF flows offer a valuable means of detecting demand shocks that are otherwise challenging to observe.

Keywords: Exchange-traded funds; Demand shocks; Fundamental mispricing; Flow decomposition; Flow commonality
JEL classification: G11, G12, G23

1. Introduction

Exchange-traded funds (ETFs) are investment vehicles listed on exchanges that are designed to track the performance of specific indices. Since the launch of the SPDR S&P 500 ETF Trust (commonly referred to as 'SPY' or 'spider'), which tracks the S&P 500 index, the ETF market has experienced continuous growth in both the number of funds and the total assets under management (AUM). This growth accelerated significantly after the COVID-19 pandemic, driven in part by a surge in retail investor activity. ETFs have become highly attractive to both retail and institutional investors due to their relatively low costs, high liquidity, and access to a wide range of asset classes. The unique structural characteristics of ETFs create a close relationship between ETFs and their underlying asset markets, with information flow between the two even observable at the intraday level (Bhattacharya and O'Hara, 2018; Box et al., 2021). Particularly, ETF fund flows capture important information about demand shocks in the underlying markets, making ETF flows a valuable source of information for financial institutions and portfolio managers seeking to understand market dynamics. While mutual fund flows, often considered an alternative to ETF flows, also reflect information about demand shocks (Kamstra et al., 2017), they are influenced by the skill of the fund managers (Chen, Jegadeesh, and Wermers, 2000; Wermers, 2000; Berk and Green, 2004). Additionally, mutual fund flows and ETF flows differ in their ability to predict future asset returns (Ben-Rephael, Kandel, and Wohl, 2012; Lou, 2012; Brown, Davies, and Ringgenberg, 2021). Due to the more complex nature of mutual fund flows, they are less suitable for analyzing the term structure of demand shocks. Despite the growing importance of ETF flows in understanding the dynamics of underlying asset markets, research on the informational content embedded in ETF flows remains limited. In response to this gap, this study aims to explore how ETF flows reflect information about demand shocks in underlying asset markets.

Among the two types of demand shocks that occur in underlying asset markets—fundamental and non-fundamental demand shocks—non-fundamental demand shocks play a crucial role. These shocks induce temporary mispricing relative to the fundamental value, and over the long term, they exhibit predictability as prices revert to their fundamental values. Despite their significance, non-fundamental demand shocks are difficult to observe directly, making their estimation a key, yet unresolved, challenge. Brown, Davies, and Ringgenberg (2021) are the first to suggest that ETF flows could serve as an observational tool for detecting non-fundamental demand shocks in underlying asset markets, providing a breakthrough in this area. They argue that ETFs and their underlying assets respond differently to non-fundamental demand shocks, which results in mispricing between the ETF's net asset value (NAV) and its market price. This mispricing triggers ETF flows, which signal the presence of non-fundamental demand shocks. In this paper, we extend and refine the theoretical framework introduced by Brown, Davies, and Ringgenberg (2021). We not only investigate how ETF flows capture information about non-fundamental demand shocks but also explore how they incorporate information about fundamental demand shocks. Additionally, we theoretically and empirically examine the term structure of demand shocks by analyzing return predictability linked to ETF flows.

We pose several research questions regarding the informational content of ETF flows. Brown, Davies, and Ringgenberg (2021) assume that fundamental demand shocks affect both ETFs and their underlying asset markets equally, implying that they do not cause an imbalance between NAV and ETF prices. Therefore, these shocks are not reflected in ETF flows and they do not consider fundamental demand shocks in their analysis. However, when fundamental demand shocks occur, both ETFs and underlying asset markets may overreact or underreact (Shiller, 2003). If ETFs and their underlying assets differ in their sensitivity to fundamental demand shocks, ETF flows may indirectly capture information about fundamental demand shocks. Ben-David, Franzoni, and Moussawi's (2018) price discovery hypothesis supports this idea, suggesting that ETFs and NAVs respond differently to fundamental shocks. This hypothesis reinforces the potential for ETF flows to reflect fundamental shock indirectly. This leads to our first research question: Can the influence of fundamental-driven flows truly be disregarded? If not, how can we disentangle fundamental-driven flows from aggregate ETF flows? By providing both theoretical and empirical evidence in response to this question, we aim to enhance our understanding of the term structure of demand shocks in underlying asset markets and refine the interpretation of the information embedded in ETF flows.

Our second research question arises from the contradictory findings in the prior literature regarding the return predictability of ETF flows. Brown, Davies, and Ringgenberg (2021) demonstrate a negative correlation between ETF flows and future asset returns at the monthly level. In contrast, Xu, Yin, and Zhao (2022) present empirical evidence showing a positive correlation between ETF flows and future returns over a short-term, one- to four-day horizon at the daily level. Xu, Yin, and Zhao (2022) argue that this positive return predictability is driven by authorized participants (APs) incorporating private information into their trading decisions, which is reflected in daily ETF flows. Holding large ETF units created by APs without immediate consumption can be costly, suggesting that liquidity and other factors play a crucial role in such information-based trading through ETFs. Given these dynamics, the noise introduced by factors such as APs' private information, liquidity constraints, and various costs, can distort high-frequency ETF flow signals. Consequently, the empirical findings may vary depending on the frequency at which ETF flows are measured and whether periods of market instability, such as financial crises, are included in the analysis.[1] These conflicting findings in previous research likely stem from the unique arbitrage mechanism within ETFs.

In addition, we believe that to properly assess whether ETF flows can serve as reliable indicators of demand shocks, it is essential to examine flow dynamics across multiple ETFs that track the same or similar indices. If ETF flows effectively capture information about demand shocks, then ETFs tracking the same or similar indices should exhibit common fluctuations. This analysis is particularly important in today's ETF market, where the number of ETFs tracking similar indices has grown exponentially. Therefore, our second research question is: Whether there is a common movement among fund flows of ETFs that track the same index? Whether this commonality varies depending on the frequency at which flow is measured? By addressing this question, we aim to further clarify the value of ETF flows as an observational tool for demand shocks.

The key findings and contributions of this paper are as follows. First, this is the first study to separate information related to fundamental demand shocks from aggregate ETF flows. Through a theoretical model, we demonstrate that ETF flows consist of two orthogonal components: one driven by non-fundamental demand shocks and the other by fundamental demand shocks. These two types of flows exhibit distinct return predictability patterns due to the different underlying information they reflect, emphasizing the importance of decomposing ETF flows. Our empirical analysis supports the model's predictions. Using recent ETF data,

[1] Ben-David et al. (2018) discuss the variation in arbitrage activity due to various market frictions, such as the bid-ask spread and stock lending fees.

which largely does not overlap with previous studies, we find that ETF flows driven by non-fundamental demand shocks exhibit significant return predictability over long horizons – at daily, weekly, and monthly intervals. In contrast, flows driven by fundamental demand shocks show return predictability only over short horizons at the daily level. Furthermore, we demonstrate that the decomposed flows associated with non-fundamental demand shocks exhibit more pronounced negative return predictability over long horizons than aggregate ETF flows. Our findings provide empirical evidence that the speed of price reversals caused by non-fundamental demand shocks slows over time, with reversals occurring over periods of one year or longer, particularly at the monthly level. Both our theoretical and empirical findings highlight that fundamental demand shocks are indeed reflected in ETF flows and that their influence cannot be ignored, underscoring the critical importance of flow decomposition in ETF flow analysis. These findings complement prior literature, which has largely overlooked fundamental demand shocks and suggest that both fundamental and non-fundamental demand shocks can be observed through ETF flows.

Second, we explore the value of ETF flows as indicators of demand shocks in underlying asset markets. For ETFs to serve as effective observational tools for demand shocks, there should be common movements across fund flows in ETFs tracking the same index. While previous studies are limited by the lack of multiple ETFs tracking the same indices, the current ETF markets provide a more suitable environment for such analysis. We analyze five major market indices replicated by numerous ETFs and find that ETF flows driven by non-fundamental demand shocks show low commonality across ETFs tracking the same index at the daily level but higher commonality at the monthly level. In contrast, ETF flows driven by fundamental demand shocks exhibit high flow commonality even at the daily level, consistent with the predictions of our theoretical model. Although the number of significantly correlated pairs decreases at the monthly level, the commonality of monthly ETF flows driven by fundamental demand shocks remains higher than at the daily level. These findings suggest that daily ETF flows contain not only information about demand shocks in underlying asset markets but also significant noise from other factors, such as individual ETFs' liquidity conditions and the arbitrage decisions of APs. As ETF flows are aggregated over longer periods, the common movements across flows become clearer, indicating that ETF flows should be analyzed at monthly or longer intervals to effectively capture information about demand shocks in underlying asset markets. We emphasize the distinct differences in flow commonality between fundamental and non-fundamental flows, highlighting ETF flows as signals of demand shocks. To our knowledge, no prior research has explored the commonality of ETF flow, making our findings a novel contribution to the literature on ETF flows and their signaling properties. We believe our results will not only deepen the understanding of ETF flows but also provide practical insights for policymakers, financial institutions, and asset managers by offering guidance on how to interpret ETF flows in underlying asset markets.[2]

This paper is organized as follows: Section 2 presents the theoretical framework and develops the hypotheses related to ETF flows. Section 3 describes the data and variables used in the study. Section 4 discusses the empirical results, and Section 5 concludes.

2. Theoretical background

In this section, we modify the ETF trading model of Brown, Davies, and Ringgenberg (2021) (hereafter BDR model) to explain how ETF flows reflect information about both

[2] The only study related to fund flow commonality is the working paper by Nguyen and Rakowski (2023), which examines commonality in mutual fund flows.

fundamental and non-fundamental demand shocks. The primary distinction between our model and the BDR model is that we incorporate information from fundamental demand shocks, in addition to non-fundamental shocks. The BDR model argues that fundamental demand shocks do not create mispricing because both ETF prices and NAVs share the same fundamental value, and thus no arbitrage opportunities arise from these shocks. Based on this reasoning, the BDR model asserts that ETF flows are driven solely by non-fundamental demand shocks and analyzes return predictability using aggregate flows without decomposing them. However, differences in liquidity can lead to varying responses from ETFs and their underlying asset markets to fundamental demand shocks. Under this scenario, arbitrage opportunities arising from the fundamental demand shocks may emerge and eventually be reflected in ETF flows. Then, using aggregate flows to test return predictability is inappropriate for analyzing the term structure of non-fundamental demand shocks.

How could information about fundamental demand shocks be reflected in ETF flows? Shiller (2003) explains that investors may overreact or underreact to fundamental changes due to behavioral biases. Similarly, Brown, Davies, and Ringgenberg (2021) suggest that fundamental demand shocks can lead to over- or under-reactions in both ETFs and NAVs. These arguments imply that part of the price pressure observed in ETFs and NAVs is driven by fundamental demand shocks. Furthermore, Israeli, Lee, and Sridharan (2017) demonstrate that ETFs, being more liquid than their underlying securities, are more attractive to uninformed traders, making them potentially more susceptible to demand-driven shocks than the underlying assets. Ben-David, Franzoni, and Moussawi's (2018) price discovery hypothesis further suggests that ETFs and NAVs respond differently to fundamental shocks. Previous studies indicate that ETFs tend to react more quickly to fundamental demand shocks, while the underlying assets may not fully reflect these shocks immediately due to lower trading volumes and liquidity constraints. As a result, over- or under-reactions to fundamental demand shocks manifest differently in ETFs and NAVs. These discrepancies, in turn, lead to arbitrage activity by APs, which reflects information about fundamental demand shocks in ETF flows.

Before detailing our model, we briefly outline the sequence of events in the ETF trade model. We construct a four-period model where $T = 0$, $T = 1$, $T = 2$, and $T = Long\ Term$ represent different time points. At $T = 0$, both the ETF price and NAV are aligned with the initial fundamental value. At $T = 1$, both a non-fundamental demand shock and a fundamental demand shock are realized. At $T = 2$, arbitrage activity by APs occurs due to the mispricing between the ETF price and NAV that arose from the demand shocks at $T = 1$. Finally, at $T = Long\ Term$, the NAV converges to a new fundamental value, which incorporates both the initial fundamental value from $T = 0$ and the fundamental shock realized at $T = 1$. Figure 1 visualizes four scenarios based on the ETF's sensitivity to the demand shocks. High sensitivity indicates that the ETF exhibits greater volatility compared to the NAV, while low sensitivity means the ETF shows less volatility compared to the NAV. The top half of the figure represents the case of a non-fundamental demand shock occurs, while the bottom half represents the case of a fundamental demand shock.

[Figure 1 here]

We now provide a detailed explanation of each point in time. First, at $T = 0$, the ETF price, NAV, and fundamental value satisfy the following equation.

$$p_0 = \pi_0 = \Omega_0, \tag{1}$$

where p_t, π_t, and Ω_t are indicate the ETF price, NAV, and fundamental value at time t, respectively. At $T = 0$, the market is in a stable state before any demand shocks are realized. During this period, the ETF price, NAV, and fundamental value are equal, indicating that the market is in equilibrium. At $T = 1$, both a non-fundamental demand shock and a fundamental demand shock occur, affecting the ETF price and NAV. The ETF price and NAV at time $T = 1$ are governed by the following equations.

$$
\begin{aligned}
\Omega_1 &= \Omega_0 + w^f, \quad w^f \sim N(0, \sigma_f^2), \\
p_1 &= \Omega_1 + \varepsilon^{etf} + \gamma^{etf} w^f, \quad \varepsilon^{etf} \sim N(0, \sigma_e^2), \\
\pi_1 &= \Omega_1 + \varepsilon^{nav} + \gamma^{nav} w^f, \quad \varepsilon^{nav} \sim N(0, \sigma_n^2),
\end{aligned} \tag{2}
$$

where w^f is the fundamental demand shock, while ε^{etf} and ε^{nav} are the non-fundamental demand shocks affecting the ETF and NAV, respectively. The parameters γ^{etf} and γ^{nav} capture the over- or under-reaction of investors to the fundamental demand shocks. Our model is a generalized ETF trading model, accounting for the possibility that aggregate ETF flows may include portions driven by fundamental demand shocks. In this setup, we separate the variation caused by fundamental demand shocks from the overall demand shocks, ensuring that w^f is independent of both ε^{nav} and ε^{etf} ($Cov(w^f, \varepsilon^{nav}) = Cov(w^f, \varepsilon^{nav}) = 0$). We omit time subscripts for simplicity because demand shocks only occur at $T = 1$.

According to equation (2), the mispricing between the ETF price and NAV at $T = 1$, denoted as μ_1, and the fundamental mispricing between the NAV and the fundamental value, denoted as φ_1, are defined as follows.

$$\mu_1 = p_1 - \pi_1 = \varepsilon^{etf} - \varepsilon^{nav} + (\gamma^{etf} - \gamma^{nav}) w^f, \tag{3}$$

$$\varphi_1 = \pi_1 - \Omega_1 = \varepsilon^{nav} + \gamma^{nav} w^f. \tag{4}$$

Due to the arbitrage activity of APs at $T = 2$, ETF flows are generated.[3] From

[3] Events occurring at $T = 1$ and $T = 2$ happen rapidly and repeatedly. To help clarify, our model aims to distinguish between the stage of demand shocks realization and arbitrage activities that occur at the intraday level (Bassiouny and Tooma, 2021; Box et al., 2021).

equation (3), we see that the aggregate ETF flow is divided into two orthogonal components: $\varepsilon^{etf} - \varepsilon^{nav}$ and $(\gamma^{etf} - \gamma^{nav})w^{f}$. We define the ETF flow driven by $\varepsilon^{etf} - \varepsilon^{nav}$ as the non-fundamental flow and the ETF flow driven by $(\gamma^{etf} - \gamma^{nav})w^{f}$ as the fundamental flow. We can estimate the relationship between non-fundamental flow, fundamental flow, and the fundamental mispricing that occurs at $T = 1$ as shown in equation (4). This estimated relationship provides insight into the return predictability of both fundamental and non-fundamental flows as the NAV converges to the latent fundamental value at $T = Long\ Term$. Let ρ represent the correlation coefficient between ε^{nav} and ε^{etf}. Then, equations (5) and (6) illustrate the relationships between non-fundamental flow and fundamental mispricing, and between fundamental flow and fundamental mispricing, respectively.

$$\begin{aligned} Cov(-\varphi_1, \varepsilon^{etf} - \varepsilon^{nav}) &= Cov(-(\varepsilon^{nav} + \gamma^{nav} w^{f}), \varepsilon^{etf} - \varepsilon^{nav}) \\ &= -Cov(\varepsilon^{nav}, \varepsilon^{etf}) + \sigma_n^2 = \sigma_n(\sigma_n - \rho\sigma_e), \end{aligned} \tag{5}$$

$$\begin{aligned} Cov(-\varphi_1, (\gamma^{etf} - \gamma^{nav})w^{f}) &= Cov(-(\varepsilon^{nav} + \gamma^{nav} w^{f}), (\gamma^{etf} - \gamma^{nav})w^{f}) \\ &= -\gamma^{nav}(\gamma^{etf} - \gamma^{nav})\sigma_f^2. \end{aligned} \tag{6}$$

Based on equations (5) and (6), we emphasize that there is a distinct difference in the return predictability of fundamental and non-fundamental flows. This distinction strongly supports the need to separate aggregate flows to analyze the term structures of demand shocks more accurately. From equation (5), we find that when the condition $\sigma_n < \rho\sigma_e$ holds, the non-fundamental flow exhibits negative return predictability. This condition is consistent with the one established by the BDR model, where $\rho = 1$ and $\sigma_n < \sigma_e$, and can be considered a more general case. As mentioned earlier, prior studies suggest that ETFs' liquidity advantage enables them to respond more quickly and significantly to non-fundamental demand shocks. Therefore, it is expected that non-fundamental flows exhibit negative return predictability. In contrast, from equation (6), we see that the direction of return predictability for fundamental flows is determined by the coefficients γ^{nav} and γ^{etf}, which remain largely obscured. This formula presents one of the primary objectives of this study: exploring the return predictability of fundamental flows.

Our ETF trade model highlights that if the return predictability of non-fundamental flows differs significantly from that of fundamental flows, then analyzing the aggregate flows may yield results that differ substantially from those obtained using decomposed flows. In situations where most variables are unobservable, how can we empirically isolate fundamental flow from aggregate flow using observable factors? We propose a simple method to decompose aggregate ETF flows, drawing from our theoretical framework. The basic idea is to utilize the common term w^f, which represents fundamental demand shocks, in conjunction with the observable and relatively stable value – NAV return. Equation (7) expresses the observable value of NAV returns, denoted as $\Delta\pi$.[4]

$$\Delta\pi = \pi_1 - \pi_0 = \Omega_1 + \varepsilon^{nav} + \gamma^{nav} w^f - \Omega_0 = \varepsilon^{nav} + (1 + \gamma^{nav}) w^f. \quad (7)$$

First, we estimate the portion of the aggregate flow that can be explained by NAV returns.

$$\frac{Cov\left(\varepsilon^{etf} - \varepsilon^{nav} + \left(\gamma^{etf} - \gamma^{nav}\right) w^f, \varepsilon^{nav} + (1 + \gamma^{nav}) w^f\right)}{Var\left(\varepsilon^{nav} + (1 + \gamma^{nav}) w^f\right)} = \frac{(1 + \gamma^{nav})\left(\gamma^{etf} - \gamma^{nav}\right)\sigma_f^2 + \sigma_n(\rho\sigma_e - \sigma_n)}{(1 + \gamma^{nav})^2\sigma_f^2 + \sigma_n^2} = \gamma. \quad (8)$$

Based on the conventional understanding that the magnitude of non-fundamental demand shocks affecting NAV is relatively small, we can deduce that the estimated value, γ, approximates $(\gamma^{etf} - \gamma^{nav})/(1 + \gamma^{nav})$. Second, we multiply the estimated value, γ, by the

[4] Since the size of underlying asset markets is significantly larger than that of the ETFs, the price pressure resulting from arbitrage activity is substantially greater for ETFs. Therefore, we assume for the sake of computational convenience that the ETF price converges to the NAV, given that the impact of arbitrage activity on NAV is relatively small. For more detailed information related to this, please refer to https://www.ishares.com/us/insights/global-etf-facts-q1-2024.

NAV return. The resulting value is as follows.

$$\gamma\Delta\pi \approx \frac{(\gamma^{etf}-\gamma^{nav})}{(1+\gamma^{nav})}\{\varepsilon^{nav} + (1+\gamma^{nav})w^{f}\} = (\gamma^{etf}-\gamma^{nav})w^{f} + \frac{(\gamma^{etf}-\gamma^{nav})}{(1+\gamma^{nav})}\varepsilon^{nav}. \qquad (9)$$

Equation (9) shows that multiplying γ by the NAV return yields a form of the fundamental flow with some added noise. Since ε^{nav} is relatively small, the resulting value from this calculation closely approximates the fundamental flow. While this method introduces some noise into the exact measurement of the fundamental flow, it provides an efficient approximation using observable factors in an ETF trade model where most variables are unobservable. In the remainder of this paper, we empirically explore the decomposition of aggregate flows into fundamental and non-fundamental components, based on the theoretical framework outlined above, and analyze their characteristics and term structures.

3. Data and flow decomposition

3.1. U.S. ETF sample

We evaluate the return predictability of fundamental flows and non-fundamental flows using data from the U.S. market, which has the largest and most diverse range of ETFs globally. Our empirical analysis focused on non-leveraged passive equity ETFs listed on the U.S. market. To gather the necessary information, we merge data from multiple sources. Daily ETF prices, shares outstanding, and NAVs are collected from Bloomberg, which is widely recognized for providing the most accurate ETF data (Ben-David, Franzoni, and Moussawi, 2018; Xu, Yin, and Zhao, 2022). We also obtain ETF identification details, such as inception dates and CUSIP numbers. The daily AUM for each ETF is calculated as the product of daily NAV and daily shares outstanding. We compute the daily aggregate ETF flow as described in equation (10).

$$AF_{i,t} = \frac{Shares\ Outstanding_{i,t} - Shares\ Outstanding_{i,t-1}}{Shares\ Outstanding_{i,t-1}}, \tag{10}$$

where $AF_{i,t}$ and $Shares\ Outstanding_{i,t}$ are the aggregate ETF flow and shares outstanding at time t, respectively. Additionally, we collect data on the daily buy value, sell value (both in dollars), and the dollar value-weighted percent effective spread from the NYSE Trade and Quote (TAQ) Millisecond Tools.[5] The daily dollar trading volume is calculated as the sum of the daily buy value and sell values. Using the 'Daily TAQ CRSP Link' provided by Wharton Research Data Services, we merge the Bloomberg data with TAQ data by matching Bloomberg's CUSIP numbers with TAQ's identifier, 'sym_root.' To calculate risk-adjusted NAV returns, we collect daily, weekly, and monthly data on the risk-free rate, market excess return, SMB, and HML factors from Ken French's Data Library.[6] We exclude all observations with missing values. Given that data on odd-lot trading, which significantly impacts market characteristics (O'Hara, Yao, and Ye 2014), only started being recorded in the TAQ database on December 9, 2013, we collect data from January 1, 2014, to December 31, 2023. This period offers a rich dataset, covering significant market events such as the end of quantitative easing in the U.S., China's economic slowdown, the COVID-19 pandemic, and the Russo-Ukrainian War. Notably, our sample period does not significantly overlap with those of prior studies on ETF flows,[7] and it is particularly relevant due to the rapid expansion in the number and size of ETFs following the COVID-19 pandemic.[8]

---

[5] These variables are constructed and provided by the TAQ database using the Lee and Ready (1991) algorithm. More detailed information can be found in the TAQ Daily Manual.

[6] Detailed information is described in https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

[7] Sample period of Ben-David, Franzoni, and Moussawi (2018) spans from 2000 to 2015. Sample period of Brown, Davies, and Ringgenberg (2021) spans from 2007 to 2016. Sample period of Xu, Yin, and Zhao (2022) spans from 2001 to 2016.

[8] According to a report by the Statista Research Department in 2024, the number of ETFs listed in the U.S. increased from 902 in 2010 to 2,235 in 2020, marking a rise of 1,333 over the decade. In contrast, the number grew by 1,008 in just three years, reaching 3,243 in 2023.

To construct the final sample for our empirical analysis, we apply the data filtering procedure outlined by Xu, Yin, and Zhao (2022). First, we exclude all ETFs with a survival period of less than 2.5 years (504 trading days). Second, we remove flow data from the first six months following each ETF's inception to avoid unstable flow patterns during the early life cycle (Broman and Shum, 2018). Lastly, we only include ETFs where non-zero fund flows account for more than 20% of their trading history. As a result, our final sample consists of 336 actively traded, non-leveraged passive equity ETFs listed on the U.S. market.

3.2 Flow decomposition

In Section 2, we introduce a method for separating fundamental flow using NAV returns. In this section, we explain how to empirically apply that method. To estimate daily fundamental flow and daily non-fundamental flow, we employ the following regression model.

$$AF_{i,t} = \beta_{i,0} + \beta_{i,1} r_{i,t} + \varepsilon_{i,t}. \tag{11}$$

In equation (11), $r_{i,t}$ represents the NAV returns at time t and the constant term accounts for fixed fund flows that occur in the ETF market independently of demand shocks.[9] To avoid look-ahead bias, we apply a 1-year rolling window approach for each ETF, estimating the $\beta_{i,1}$ coefficient at each time point. The estimated coefficient multiplied by NAV returns defines the fundamental flow. Non-fundamental flow is then defined as the aggregate flow minus both the fundamental flow and the constant term.

Table 1 reports the summary statistics and pairwise correlations for aggregate flow,

[9] We also conducted the decomposition using a model without a constant term. The resulting fundamental flow and non-fundamental flow are found to have correlations of 0.999 and 0.999, respectively, with those estimated from the model with a constant term. This indicates that the inclusion or exclusion of the constant term in the regression model has negligible impact on the decomposed flows.

fundamental flow, and non-fundamental flow. The results indicate that a large portion of the aggregate flow consists of non-fundamental flow, with a high correlation between the two flows. This finding aligns with our expectation that fundamental flow, which is driven by fundamental demand shocks, represents a relatively smaller portion of the total flow compared to non-fundamental flow. It is important to note that fundamental flow, reflects only the visible part of the fundamental demand shock, much like the 'tip of the iceberg'. Therefore, the small proportion of fundamental flow in aggregate flow does not imply that the impact of fundamental demand shocks is small. Rather, it suggests that we should pay attention to even small variations in fundamental flows. Moreover, the correlation between fundamental flow and non-fundamental flow is -0.053, which is extremely low, indicating that the two decomposed flows exhibit the theoretical characteristics we described. This further underscores the importance of separating these flows to more accurately analyze the term structures of demand shocks.

[Table 1 here]

Figure 2 displays the cross-sectional averages of the daily estimated fundamental flow, non-fundamental flow, and $\beta_{i,1}$ (hereafter referred to as sensitivity to NAV returns) for each ETF. The lower part of Figure 2 shows that the cross-sectional average of the estimated coefficients is positive, indicating that ETFs tend to exhibit a stronger response to fundamental demand shocks. This heightened sensitivity, similar to the larger response to non-fundamental demand shocks, likely arises from the greater liquidity of ETFs compared to their underlying assets. Furthermore, the sensitivity appears to decrease following the COVID-19 pandemic.

[Figure 2 here]

We also investigate whether this trend is evident in the leading ETFs that track various major indices. Figure 3 displays the daily fundamental flow, non-fundamental flow, and sensitivity for the top four ETFs (excluding sector ETFs) in our sample, ranked by daily dollar trading volume.[10] Notably, despite differences in the underlying indices tracked by these ETFs, the sensitivity trend is consistent with the pattern observed in Figure 2. From these results, we infer that the sharp increase in the number of ETFs following the COVID-19 pandemic has contributed to a decrease in the sensitivity of individual ETFs. This can be interpreted as the effect of multiple ETFs emerging to track similar underlying asset markets, causing the over- or under-reaction to fundamental demand shocks to be spread across several ETFs.

[Figure 3 here]

4. Empirical Results

4.1 Panel regression

To investigate the return predictability of fundamental and non-fundamental flows, we examine the relationship between decomposed ETF flows at daily, weekly, and monthly levels and future NAV returns using panel regression analysis. We estimate the following panel regression models.

$$r_{i,t+h} = \beta_0 + \beta_1 FF_{i,t} + \beta_2 X'_{i,t} + \alpha_i + \delta_t + \varepsilon_{i,t+h}, \tag{12}$$

$$r_{i,t+h} = \beta_0 + \sum_{d=2}^{p} \beta_{1,d} QUANTILE_{i,d,t} + \beta_2 X'_{i,t} + \alpha_i + \delta_t + \varepsilon_{i,t+h}, \tag{13}$$

$$r_{i,t+h} = \sum_{k=t+1}^{t+h} r_{i,k}, \tag{14}$$

where $r_{i,t}$ denotes the NAV returns of ETF i at time t, and $FF_{i,t}$ represents the flow

[10] Among the top six ETFs by trading volume, the Financial Select Sector SPDR Fund (ETF Ticker: XLF) and the VanEck Gold Miners ETF (ETF Ticker: GDX) are sector ETFs.

variable, indicating either fundamental flow or non-fundamental flow.[11] $QUANTILE_{i,d,t}$ is a decile (or quintile) dummy variable, which equals 1 if ETF i falls into the d-th decile (or quintile) group based on the fund flow at time t; otherwise, it equals 0. p is 10 for decile dummies and 5 for quintile dummies. Hereafter, we will refer to each quantile group simply as 'Decile d' for deciles and 'Quintile d' for quintiles. Since we obtained similar results using quintile dummies, the main manuscript reports the results only for decile dummies.[12]

We derive results at both the ETF level and the portfolio level using equations (12) and (13).[13] $X'_{i,t}$ represents a set of control variables, which include two lags of NAV returns, logarithmic dollar trading volume, logarithmic AUM, and dollar value-weighted percent effective spread. Our models include both time-fixed effects and ETF-fixed effects. When constructing weekly variables from daily data, we define the week as starting on Saturday and ending on Friday, consistent with the definition used in Ken French's Data Library, rather than based on trading days. This approach ensures consistency in capturing information on a weekly basis. For the weekly and monthly variables, we calculate the flow variable as the sum of daily values, while the control variables (excluding the two lags of NAV returns) are computed as averages. We analyze various horizons, from $h = 1$ to $h = 12$, and to account for

[11] Although not included in the main text, we also conduct panel regression analysis using both raw returns and risk factor-adjusted returns through the capital asset pricing model and the Fama-French three-factor model (Fama and French, 1993). We obtain similar results with the exception that fundamental flow exhibited long-horizon return predictability at the daily and weekly levels when using abnormal returns derived from the Fama-French three-factor model.

[12] The results for the quintile dummy can be found in the Online Appendix.

[13] Since the analysis of the long-short portfolio is deemed less reliable due to the inability to reflect the time-invariant characteristics of ETFs, we did not include it in the main text. However, we construct decile portfolios based on the previous period's ETF flow, going long on the portfolio with the highest inflow and short on the portfolio with the lowest inflow, creating value-weighted long-short portfolios at daily, weekly, and monthly frequencies for performance analysis. In the case of the non-fundamental flow based long-short portfolio, significantly negative Jensen's (1968) alpha and alpha from the Fama-French three-factor model (Fama and French, 1993) are observed across all horizons. For the fundamental flow based long-short portfolio, significantly negative Jensen's alpha and alpha from the Fama-French three-factor model are noted only in the daily portfolio. Detailed results can be found in the Online Appendix.

overlapping return periods, we use Driscoll and Kraay's (1998) standard errors with a lag corresponding to the return horizon. We standardize the control variables (excluding the two lags of NAV returns) and the flow variables by subtracting the sample mean and dividing by the sample standard deviation during the estimation period to estimate the panel regression model.

Table 2 shows the estimation results at the daily level. Panels A and C display the estimated coefficients for fundamental flow and non-fundamental flow, respectively, while Panels B and D present the estimated coefficients for the decile dummy variables based on fundamental flow and non-fundamental flow, respectively. According to the results in Panels A and C, fundamental flow exhibits significantly negative return predictability up to a four-day horizon, whereas non-fundamental flow shows significantly negative return predictability for all horizons beyond the one-day horizon. In both cases, the coefficients follow a monotonic trend: for fundamental flow, the absolute value of the coefficients begins to decrease after the three-day horizon, while non-fundamental flow reaches its largest negative coefficient at the 12-day horizon. In Panels B and D, the estimated coefficients for the decile dummy variables represent the differences from Decile 1. By examining the coefficient estimates for Decile 10, we can assess the differences between the highest inflow group and the lowest inflow group. In Panel B, Decile 10 exhibits a significantly larger negative return compared to Decile 1 for horizons up to two days (e.g., at $h = 2$, an annualized return of -2.73%). In Panel D, Decile 10 shows a significantly larger negative return compared to Decile 1 for all horizons beyond two days (e.g., at $h = 12$, an annualized return of -2.06%).[14]

---

[14] In the analysis using quintile dummy variables, the estimated coefficient for Quintile 5 shows significance for fundamental flow up to the three-day horizon, while for non-fundamental flow, significance is observed for all horizons beyond the three-

[Table 2 here]

Table 2 provides empirical evidence that fundamental flow is associated with short-horizon negative return predictability, while non-fundamental flow exhibits long-horizon negative return predictability. Notably, despite the substantial influence of noise beyond demand shocks at the daily level, non-fundamental flow demonstrates significant price reversals and long-horizon negative return predictability over approximately seven days. This suggests that even at the daily level, ETF flows effectively reflect information about demand shocks and reveal the daily structure of non-fundamental demand shocks.

Table 3 presents the estimation results at the weekly level. Panels A and B show that weekly fundamental flow does not exhibit significant return predictability at any horizon. This lack of predictability suggests that the fundamental mispricing caused by fundamental demand shocks dissipates within a week, preventing weekly fundamental flow from demonstrating long-term predictive power. In contrast, Panel C shows that non-fundamental flow exhibits significantly negative return predictability for all horizons up to nine weeks, except for the four-week and eight-week horizons. Additionally, Panel D reveals that Decile 10 exhibits significantly larger negative returns compared to Decile 1 at nearly all horizons, except for the three-week horizon (e.g., at $h = 12$, an annualized return of -1.89%).[15]

[Table 3 here]

Table 4 reports the estimation results at the monthly level. Panels A and B show that, similar to the findings at the weekly level in Table 3, fundamental flow does not exhibit any

day horizon.

[15] In the results for the quintile dummy, Quintile 5 exhibits significantly larger negative returns compared to Quintile 1 across all horizons.

significant predictive power. In contrast, Panel C shows that non-fundamental flow demonstrates significantly negative return predictability for all horizons beyond the three-month horizon. Additionally, Panel D reveals that for all horizons after the two-month horizon, Decile 10 exhibits significantly larger negative returns compared to Decile 1 (e.g., at $h = 12$, annualized return of -2.10%).[16] The results for non-fundamental flow in Table 4 indicate that monthly non-fundamental demand shocks lead to price reversals lasting nearly a year or longer. Furthermore, the coefficients follow a monotonic trend, consistent with the expected behavior of price reversals due to fundamental mispricing. These findings suggest that fundamental mispricing arising from non-fundamental demand shocks in underlying asset markets undergoes a slow and continuous convergence process. Ultimately, non-fundamental flow effectively captures information about non-fundamental demand shocks in underlying asset markets, illustrating the gradual convergence of fundamental mispricing.

[Table 4 here]

Theoretical and empirical results from our study provide the following key implications: Aggregate ETF flow can be decomposed into two orthogonal components, fundamental flow and non-fundamental flow, each with distinct return predictability. This decomposition is critical because aggregate flow includes fundamental flow, which lacks long-horizon return predictability. As a result, using aggregate flow as a signal for non-fundamental demand shocks misrepresents the true term structure of non-fundamental demand shocks (in the absence of fundamental demand shocks).

To examine whether aggregate flow accurately captures the long-term convergence of

[16] The results for the quintile dummy indicate that for all horizons beyond the four-month horizon, Quintile 5 shows significantly larger negative returns compared to Quintile 1.

fundamental mispricing driven by non-fundamental demand shocks, we estimate coefficients for aggregate flow using our panel regression model and compare its return predictability to that of the decomposed flows. Figure 4 displays the estimated coefficients and 90% confidence intervals from equation (12) for aggregate flow, fundamental flow, and non-fundamental flow at daily, weekly, and monthly levels.[17] The first row shows the results for daily fund flows, the second row for weekly flows, and the third row for monthly flows. Each row is arranged from left to right, showing the results for aggregate flow, fundamental flow, and non-fundamental flow.

[Figure 4 here]

According to the results in the first row of Figure 4, at the daily level, there is no significant difference in the return predictability of future NAV returns between aggregate flow and non-fundamental flow. However, in the second row, at the weekly level, aggregate flow shows significant predictability only at the one-week and two-week horizons, with smaller coefficient magnitudes compared to non-fundamental flow. This difference becomes more pronounced over longer horizons, as the magnitude of coefficients for weekly aggregate flow gradually decreases, while that for weekly non-fundamental flow increases monotonically. The third row, showing results at the monthly level, highlights a stark contrast between aggregate flow and non-fundamental flow. Monthly aggregate flow exhibits significantly negative return predictability only between the four-month and six-month horizons, with coefficient magnitudes starting to shrink from the five-month horizon onward. In contrast, monthly non-fundamental flow exhibits significantly negative return predictability for all horizons beyond

[17] We display the figures instead of tables for a clear and visually striking comparison. Detailed tabular results for the figures can be found in the Online Appendix.

the three-month horizon, with coefficients remaining stable up to the 12-month horizon. The overall difference in coefficient magnitude is striking: at the 12-month horizon, the predicted future NAV return difference between aggregate flow and non-fundamental flow, when each increases by one standard deviation, is 15.6 basis points. Moreover, the 90% confidence intervals for long-horizon non-fundamental flow are noticeably narrower than those for long-horizon aggregate flow.

We also compare the coefficients of the quantile dummies. For monthly aggregate flow, fundamental flow, and non-fundamental flow, we construct quintile and decile dummies, respectively, and estimate equation (13) to determine the coefficients for Quintile 5 and Decile 10.[18] The estimation results are displayed in Figure 5. The first row of Figure 5 reports the coefficients for Quintile 5, while the second row shows the coefficients for Decile 10, with error bars representing the 90% confidence intervals. Figure 5 clearly illustrates the significant differences in future return predictability between aggregate flow and non-fundamental flow. In the first row, for quintile dummies based on aggregate flow, Quintile 1 and Quintile 5 exhibit significant negative return differences at long horizons beyond the four-month mark. However, at the 12-month horizon, the coefficient for aggregate flow is -0.871 (t-statistic is -1.742), compared to -1.404 (t-statistic is -3.240) for non-fundamental flow, highlighting a substantial difference between the two flows. This difference becomes even more pronounced in the Decile 10 comparison. For aggregate flow, there is almost no significant return difference between Decile 1 and Decile 10, whereas non-fundamental flow continues to exhibit a monotonic trend, with a clearer return difference beyond the two-month horizon. At the 12-month horizon, the

[18] For the daily and weekly levels, constructing quintile and decile portfolios based on aggregate flow results in overlapping ETFs at the boundary thresholds, making it challenging to properly form portfolio dummy variables. Therefore, the analysis is conducted only at the monthly level.

coefficient for Decile 10 based on aggregate flow is -1.083 (*t*-statistic is -1.366), while for non-fundamental flow, it is -2.097 (*t*-statistic is -3.227).

[Figure 5 here]

If fund flows effectively capture information about non-fundamental demand shocks, the process of fundamental mispricing convergence should be reflected in return predictability. However, our empirical findings suggest that aggregate flow does not exhibit the typical price reversal pattern associated with non-fundamental demand shocks, as the significance of the quantile dummy coefficients diminishes and no clear monotonic trend emerges. In contrast, the non-fundamental flow extracted through our decomposition process effectively reflects the term structure of non-fundamental demand shocks, as demonstrated by the results in Figures 4 and 5.

We also investigate whether the return predictability of ETF flows varies over time, considering the significant changes in the information about fundamental demand shocks reflected in ETF flows around 2019 (see Figures 2 and 3), and the sharp increase in both the number and size of ETFs. Figure 6 displays the daily total AUM and the average illiquidity of our ETF sample, measured by the dollar-weighted percent effective spread, over the entire period. Since there is an inverse relationship between the illiquidity measure and liquidity, a lower value of the illiquidity measure indicates higher liquidity. Figure 6 shows a steady increase in ETF AUM and liquidity, with a particularly notable shift around 2019. The ability of ETFs to effectively reflect non-fundamental demand shocks in underlying asset markets is largely attributed to their high liquidity compared to underlying asset markets. If ETF liquidity is low, it might hinder the quick and accurate reflection of non-fundamental demand shocks, making the signaling power of ETF flows less clear. This suggests that the clarity of the information contained in ETF flows may differ between the past and present. Moreover, mutual fund flows, which serve as substitutes for ETFs, also reflect non-fundamental demand (Edelen, 1999; Barber, Odean, and Zheng, 2005; Frazzini and Lamont, 2008). When mutual funds were still dominant, the information reflected in ETF flows may have been more dispersed. However, as more funds have shifted from mutual funds to ETFs,[19] and with the growing number of ETFs, their interactions, and higher liquidity, ETF flows today are likely better at capturing demand shocks in underlying asset markets. Consequently, we expect that the return reversals observed through ETF flows to become more pronounced over time, which we further explore through a subperiod analysis.

[Figure 6 here]

Figure 7 presents the estimation results from equation (12), comparing the period before 2019 (before January 1, 2019) with the period after 2019 (starting January 1, 2019). Since the results for daily, weekly, and monthly flow show similar trends, we report only the monthly flow results. The results are organized from top to bottom, showing aggregate flow, fundamental flow, and non-fundamental flow, with the left side representing the period before 2019 ('Pre-2019') and the right side representing the period after 2019 ('Post-2019').

[Figure 7 here]

The findings indicate that, before 2019, most flows do not exhibit significant return predictability. However, in the period after 2019, non-fundamental flow shows significantly negative return predictability across all horizons. For aggregate flow, there is no significant predictability beyond the four- and five-month horizons in the post-2019 period, and the

[19] See, https://insight.factset.com/etf-trends-familiar-direction-but-bigger.

magnitude of the coefficients is also smaller compared to non-fundamental flow, returning to near zero at longer horizons. From Figure 7, we observe a marked difference in return predictability between aggregate flow and non-fundamental flow. The term structure displayed by aggregate flow does not significantly capture the characteristics of return reversals caused by fundamental mispricing. In contrast, non-fundamental flow exhibits significant changes in both the significance and magnitude of the estimated coefficients around 2019, indicating that increased accessibility to ETFs allows non-fundamental demand shocks in underlying asset markets to be more effectively reflected in ETF flows.

4.2 Flow commonality

In panel regression analysis, we examine the future asset return predictability of fundamental and non-fundamental flows, comparing these with aggregate flow and analyzing how their behavior differs between the past and present. Our empirical results indicate that, on average, non-fundamental flow exhibits significantly negative return predictability over long horizons, while fundamental flow exhibits insignificant return predictability. Notably, at the monthly frequency, price reversals caused by non-fundamental demand shocks last for a year or more. We suggest that non-fundamental flows more effectively capture information about non-fundamental demand shocks in underlying asset markets compared to aggregate flows. Building on this, it is essential to address several key questions: Is there commonality in the flows of individual ETFs tracking the same index? At what frequency should flows be observed to accurately capture real demand shocks? And do fund flows in leading ETFs exhibit a high correlation with those in other ETFs? Answering these questions is crucial for understanding the value of ETF flows as an observational tool for demand shocks.

In the past, the limited number of actively traded ETFs tracking the same index restricted analysis of flow commonality. However, today, numerous ETFs track several well-known major market indices, providing an appropriate environment to analyze the flow dynamics among ETFs that track the same index. To address the questions above, we collect additional data on global ETFs listed on exchanges worldwide that track five major indices: S&P 500, NASDAQ 100, EUROSTOXX 50, FTSE 100, and DAX using Bloomberg.[20] We gather daily NAV and shares outstanding for these global ETFs over our sample period and apply the same methodology from Section 3 to estimate aggregate, fundamental, and non-fundamental flows.

Figure 8 visually represents the annual number of global ETFs tracking each index before applying the ETF filtering process from Section 3. All five indices are tracked by numerous ETFs, with the number of ETFs tracking the S&P 500 and NASDAQ 100 indices increasing sharply after 2019. This surge likely reflects the strong global performance of the U.S. market, driving higher demand for ETFs tracking these indices. After filtering, the final number of ETFs used for analysis is 34 for the S&P 500 index, 13 for the NASDAQ 100 index, 10 for the EUROSTOXX 50 index, five for the FTSE 100 index, and seven for the DAX index. These include well-known ETFs such as the SPDR S&P 500 Trust ETF, Vanguard S&P 500 ETF, and iShares Core S&P 500 ETF for the S&P 500 index, and the Invesco QQQ Trust for the NASDAQ 100 index.

[Figure 8 here]

[20] We use the underlying index tickers provided by Bloomberg to select global ETFs tracking each index. For the S&P 500 index, we use 'SPX,' 'SPXT,' 'SPTR,' and 'SPTR500N'; for the NASDAQ 100 index, 'NDX' and 'XNDX'; for the EUROSTOXX 50 index, 'SX5T' and 'SX5E'; for the FTSE 100 index, 'UKX,' 'TUKXG,' and 'UKXNUK'; and for the DAX index, 'DAX' and 'DAXNR.'

To estimate the commonality in ETF flows, we use pairwise correlations a regression-based commonality estimation method widely used in finance. Following Chordia, Roll, and Subrahmanyam (2000), Karolyi, Lee, and Van Dijk (2012), Dang, Moshirian, and Zhang (2015), and Brockman, Chung, and Snow (2023), we estimate the following regression model for each ETF and use the R-squared as a measure of flow commonality.

$$FF_{i,t} = c_0 + c_1 FF_{m,t-1} + c_2 FF_{m,t} + c_3 FF_{m,t+1} + \varepsilon_t, \tag{15}$$

where $FF_{i,t}$ represents the ETF flow (aggregate flow, fundamental flow, or non-fundamental flow) for ETF i at time t, $FF_{m,t}$ is the market ETF flow at time t. The model includes the lagged, concurrent, and lead terms of the market ETF flow. For each underlying index, the market ETF flow for ETF i is calculated as the equal-weighted average of ETF flows, excluding ETF i. We estimate the pairwise correlations and R-squared values from equation (15) using daily, weekly, and monthly flows of the ETFs that track each underlying index.

Table 5 reports the results on the commonality of aggregate flow. Panels A, B, and C correspond to the results for daily, weekly, and monthly flows, respectively. Table 5 shows that both the average R-squared and the average absolute correlation increase from daily flows to monthly flows, suggesting that flow commonality strengthens as the observation frequency decreases. The lower commonality in daily ETF flows can be attributed to APs' arbitrage decisions, which are heavily influenced by individual ETF characteristics, such as liquidity, at the daily level. When ETF flows are aggregated over longer time frames, this limit to arbitrage effect diminishes, allowing information from underlying asset markets to be more effectively reflected in ETF flows. Notably, this pattern is consistent across all underlying indices.

[Table 5 here]

Table 6 reports the results on the commonality of fundamental flow. Similar to aggregate flow, the results show that as we move to monthly flows, both the average R-squared values and the average absolute correlations increase. However, despite the increase in commonality at the monthly level, the number of significantly correlated pairs decreases. For fundamental flow, high levels of commonality are evident even at the daily level, and the significance of pairwise correlations remains strong across all underlying indices. This high commonality in fundamental flow is driven by ETFs' relative sensitivity to fundamental demand shocks, as the underlying indices are unified. The key takeaway here is that the time-varying sensitivity of individual ETFs moves in a highly similar manner.

[Table 6 here]

Table 7 reports the results on the commonality of non-fundamental flow. Although aggregate and non-fundamental flows are highly correlated, they differ in terms of commonality. This difference likely arises because aggregate flow contains information on fundamental flow, which exhibits different commonality characteristics from non-fundamental flow. In the case of the S&P 500 and NASDAQ 100 indices, which are tracked by many ETFs, the average R-squared values and average absolute correlations for weekly and monthly flows are higher for non-fundamental flow than for aggregate flow with more significantly correlated pairs. For the S&P 500 index in particular, while the number of significantly correlated pairs for aggregate flow remains fairly consistent across daily, weekly, and monthly flows (as shown in Table 5), this number increases noticeably for non-fundamental flow when moving from daily to monthly frequencies. Since non-fundamental flow is influenced by both APs' arbitrage decisions and the specific non-fundamental demand shocks affecting individual ETFs, which are impacted by limits to arbitrage, non-fundamental flows should show stronger common movements over longer periods, assuming they effectively reflect non-fundamental demand shocks effectively.

[Table 7 here]

Our commonality analysis suggests that decomposed non-fundamental flow more accurately captures information about non-fundamental demand shocks compared to aggregate flow. Observing non-fundamental flow at the monthly level provides more accurate insights into non-fundamental demand shocks in underlying asset market than daily observations.

To further investigate whether leading ETFs play a dominant role in ETF flow dynamics, we examine the correlation between the decomposed flows of the most popular ETFs tracking the S&P 500 index and NASDAQ 100 index—namely, the SPDR S&P 500 Trust ETF and the Invesco QQQ Trust—and the decomposed flows of other ETFs. Figure 9 displays the results for SPDR S&P 500 Trust ETF's fundamental and non-fundamental flows, followed by Invesco QQQ Trust's fundamental and non-fundamental flows, with red indicating negative correlations and blue indicating positive correlations.

[Figure 9 here]

As expected, fundamental flows often show high correlations due to their sensitivity to fundamental demand shocks. However, non-fundamental flows do not exhibit noticeable correlations. This suggests that leading ETFs do not play a dominant role in ETF flow dynamics, which aligns with Nguyen and Rakowski's (2023) findings that major U.S. mutual funds do not dominate flow commonality.

5. Conclusion

ETF flows contain valuable information about both fundamental and non-fundamental demand shocks in underlying asset markets. In this study, we make the first attempt to decompose ETF flows into two orthogonal components to separately observe the effects of these demand shocks: fundamental flow and non-fundamental flow. We provide both a theoretical framework and empirical evidence supporting the use of decomposed ETF flows as an observational tool for analyzing fundamental and non-fundamental demand shocks. Through this decomposition, we address key research gaps in the term structure of demand shocks and the speed of price reversals, as highlighted by Brown, Davies, and Ringgenberg (2021). Our theoretical model suggests that fundamental and non-fundamental flows exhibit distinct return predictability, necessitating separate consideration. Empirically, we find that daily fundamental flow shows short-horizon negative return predictability, while weekly and monthly fundamental flows show no significant return predictability. In contrast, non-fundamental flow demonstrates significant long-horizon negative return predictability at daily, weekly, and monthly frequencies. Our findings reveal that fundamental mispricing caused by non-fundamental demand shocks tends to revert over an extended period. Specifically, the results from monthly non-fundamental flow suggest that price reversals driven by non-fundamental demand shocks occur over a long horizon, typically one year or more, with the speed of reversal gradually decreasing. This price reversal process is more significant and prolonged than what is observed through aggregate flow, indicating that non-fundamental flow effectively reflects pure information about non-fundamental demand shocks in underlying asset markets.

We also analyze the commonality of ETF flows tracking the same index and find that both fundamental and non-fundamental flows exhibit greater commonality at weekly and monthly levels compared to daily flows. Leading ETFs play a less dominant role in this commonality. We suggest that monthly flows, compared to daily flows, are less affected by noise from individual ETF characteristics and APs' arbitrage decisions. This allows low-frequency flows to more effectively capture information from demand shocks.

Overall, our study emphasizes that ETF flows can serve as a valuable tool for detecting demand shocks in underlying asset markets, particularly those that are challenging to observe through other means. We hope future research will explore more efficient methods of disentangling the various pieces of information embedded within ETF flows.

Supplementary material

Supplementary data are available.

References

Barber, B.M., Odean, T., Zheng, L., 2005. Out of sight, out of mind: The effects of expenses on mutual fund flows. J. Bus. 78, 2095–2120. https://doi.org/10.1086/497042

Bassiouny, A., Tooma, E., 2021. Intraday indirect arbitrage between European index ETFs. Int. Rev. Financ. Anal. 75, 101737. https://doi.org/10.1016/j.irfa.2021.101737

Ben-David, I., Franzoni, F., Moussawi, R., 2018. Do ETFs increase volatility? J. Finance 73, 2471–2535. https://doi.org/10.1111/jofi.12727

Ben-Rephael, A., Kandel, S., Wohl, A., 2012. Measuring investor sentiment with mutual fund flows. J. Financ. Econ. 104, 363–382. https://doi.org/10.1016/j.jfineco.2010.08.018

Berk, J.B., Green, R.C., 2004. Mutual fund flows and performance in rational markets. J. Polit. Econ. 112, 1269–1295. https://doi.org/10.1086/424739

Bhattacharya, A., O'Hara, M., 2017. Can ETFs increase market fragility? Effect of information linkages in ETF markets. Working Paper, Cornell University. https://doi.org/10.2139/ssrn.2740699

Box, T., Davis, R., Evans, R., Lynch, A., 2021. Intraday arbitrage between ETFs and their underlying portfolios. J. Financ. Econ. 141, 1078–1095. https://doi.org/10.1016/j.jfineco.2021.04.023

Brockman, P., Chung, D.Y., Snow, N.M., 2023. Search-based peer groups and commonality in liquidity. Rev. Financ. 27, 33–77. https://doi.org/10.1093/rof/rfab033

Broman, M.S., Shum, P., 2018. Relative liquidity, fund flows and short-term demand: Evidence from exchange-traded funds. Financ. Rev. 53, 87–115. https://doi.org/10.1111/fire.12159

Brown, D.C., Davies, S.W., Ringgenberg, M.C., 2021. ETF arbitrage, non-fundamental demand, and return predictability. Rev. Financ. 25, 937–972. https://doi.org/10.1093/rof/rfaa027

Chen, H.-L., Jegadeesh, N., Wermers, R., 2000. The value of active mutual fund management: An examination of the stockholdings and trades of fund managers. J. Financ. Quant. Anal. 35, 343. https://doi.org/10.2307/2676208

Chordia, T., Roll, R., Subrahmanyam, A., 2000. Commonality in liquidity. J. Financ. Econ. 56, 3–28. https://doi.org/10.1016/S0304-405X(99)00057-4

Dang, T.L., Moshirian, F., Zhang, B., 2015. Commonality in news around the world. J. Financ. Econ. 116, 82–110. https://doi.org/10.1016/j.jfineco.2014.11.007

Driscoll, J.C., Kraay, A.C., 1998. Consistent covariance matrix estimation with spatially dependent panel data. Rev. Econ. Stat. 80, 549–560. https://doi.org/10.1162/003465398557825

Edelen, R.M., 1999. Investor flows and the assessed performance of open-end mutual funds. J. Financ. Econ. 53, 439–466. https://doi.org/10.1016/S0304-405X(99)00028-8

Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. J. Financ. Econ. 33, 3–56. https://doi.org/10.1016/0304-405X(93)90023-5

Frazzini, A., Lamont, O.A., 2008. Dumb money: Mutual fund flows and the cross-section of stock returns. J. Financ. Econ. 88, 299–322. https://doi.org/10.1016/j.jfineco.2007.07.001

Israeli, D., Lee, C.M.C., Sridharan, S.A., 2017. Is there a dark side to exchange traded funds? An information perspective. Rev. Account. Stud. 22, 1048–1083. https://doi.org/10.1007/s11142-017-9400-8

Jensen, M.C., 1968. The performance of mutual funds in the period 1945-1964. J. Finance 23, 389. https://doi.org/10.2307/2325404

Kamstra, M.J., Kramer, L.A., Levi, M.D., Wermers, R., 2017. Seasonal asset allocation: Evidence from mutual fund flows. J. Financ. Quant. Anal. 52, 71–109. https://doi.org/10.1017/S002210901600082X

Karolyi, G.A., Lee, K.-H., van Dijk, M.A., 2012. Understanding commonality in liquidity around the world. J. Financ. Econ. 105, 82–112. https://doi.org/10.1016/j.jfineco.2011.12.008

Lee, C.M.C., Ready, M.J., 1991. Inferring trade direction from intraday data. J. Finance 46, 733–746. https://doi.org/10.1111/j.1540-6261.1991.tb02683.x

Lou, D., 2012. A flow-based explanation for return predictability. Rev. Financ. Stud. 25, 3457–3489. https://doi.org/10.1093/rfs/hhs103

Nguyen, A.T., Rakowski, D.A., 2022. Commonality in mutual fund flows and global market integration. Working Paper. https://doi.org/10.2139/ssrn.3884108

O'Hara, M., Yao, C., Ye, M., 2014. What's not there: Odd lots and market data. J. Finance 69, 2199–2236. https://doi.org/10.1111/jofi.12185

Shiller, R.J., 2003. From efficient markets theory to behavioral finance. J. Econ. Perspect. 17, 83–104. https://doi.org/10.1257/089533003321164967

Wermers, R., 2000. Mutual fund performance: An empirical decomposition into stock-picking talent, style, transactions costs, and expenses. J. Finance 55, 1655–1695. https://doi.org/10.1111/0022-1082.00263

Xu, L., Yin, X., Zhao, J., 2022. Are the flows of exchange-traded funds informative? Financ. Manag. 51, 1165–1200. https://doi.org/10.1111/fima.12396

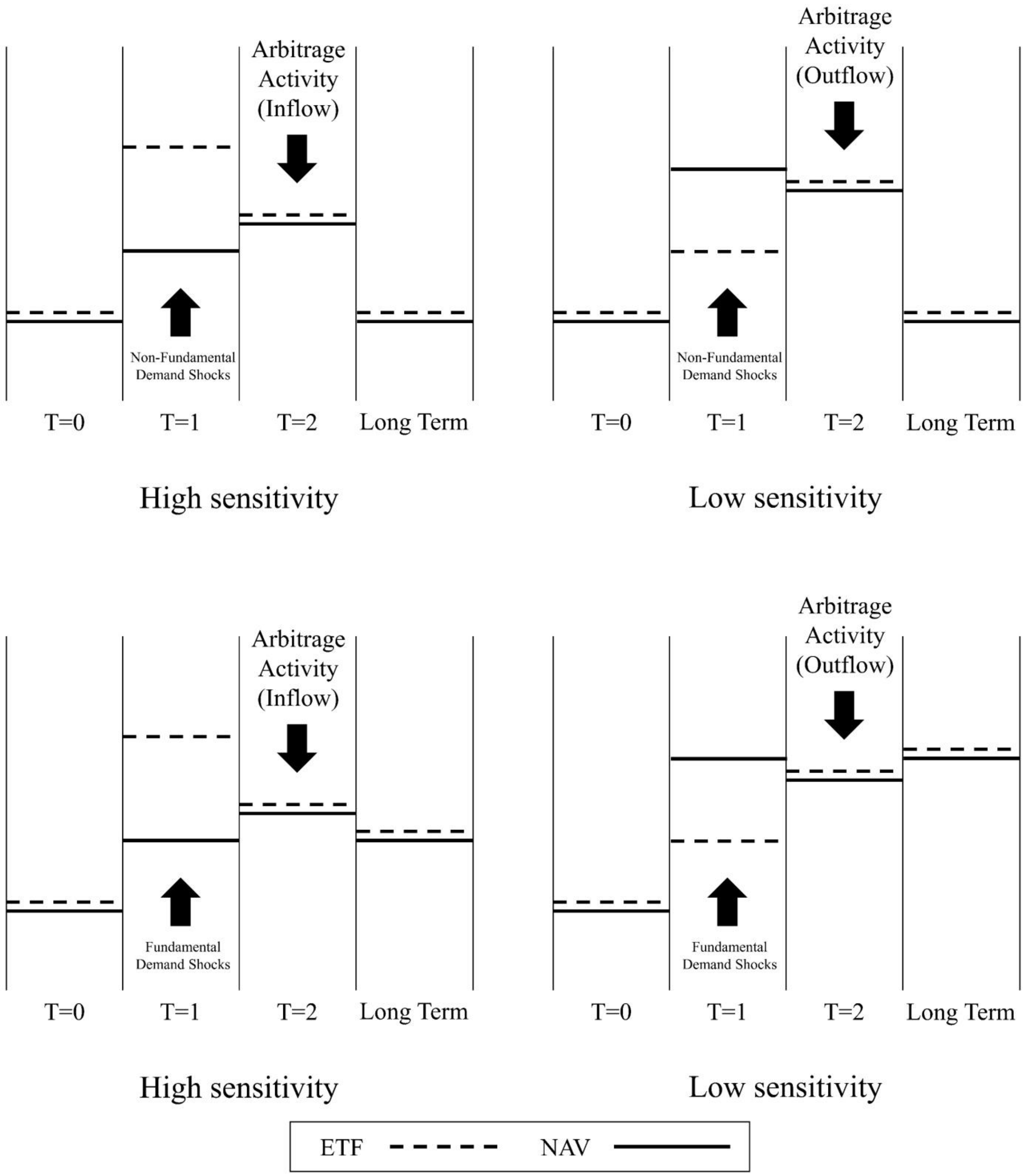

Figure 1. Mechanism of ETF flow

This figure visualizes four scenarios of our four-period theoretical model based on the sensitivity of the ETF to the demand shocks. High (low) sensitivity means that the ETF exhibits greater (smaller) volatility compared to the NAV. The above presents the case where a non-fundamental demand shock occurs, and the below presents the case where a fundamental demand shock occurs. The dashed line and the solid line indicate ETF price and NAV, respectively. At $T = 0$, both the ETF and NAV are aligned with the initial fundamental value. At $T = 1$, both non-fundamental demand shock and fundamental demand shock are realized. At $T = 2$, the arbitrage activity by APs occurs due to the mispricing between the ETF price and NAV. At $T = Long\ Term$, the NAV converges to a new fundamental value.

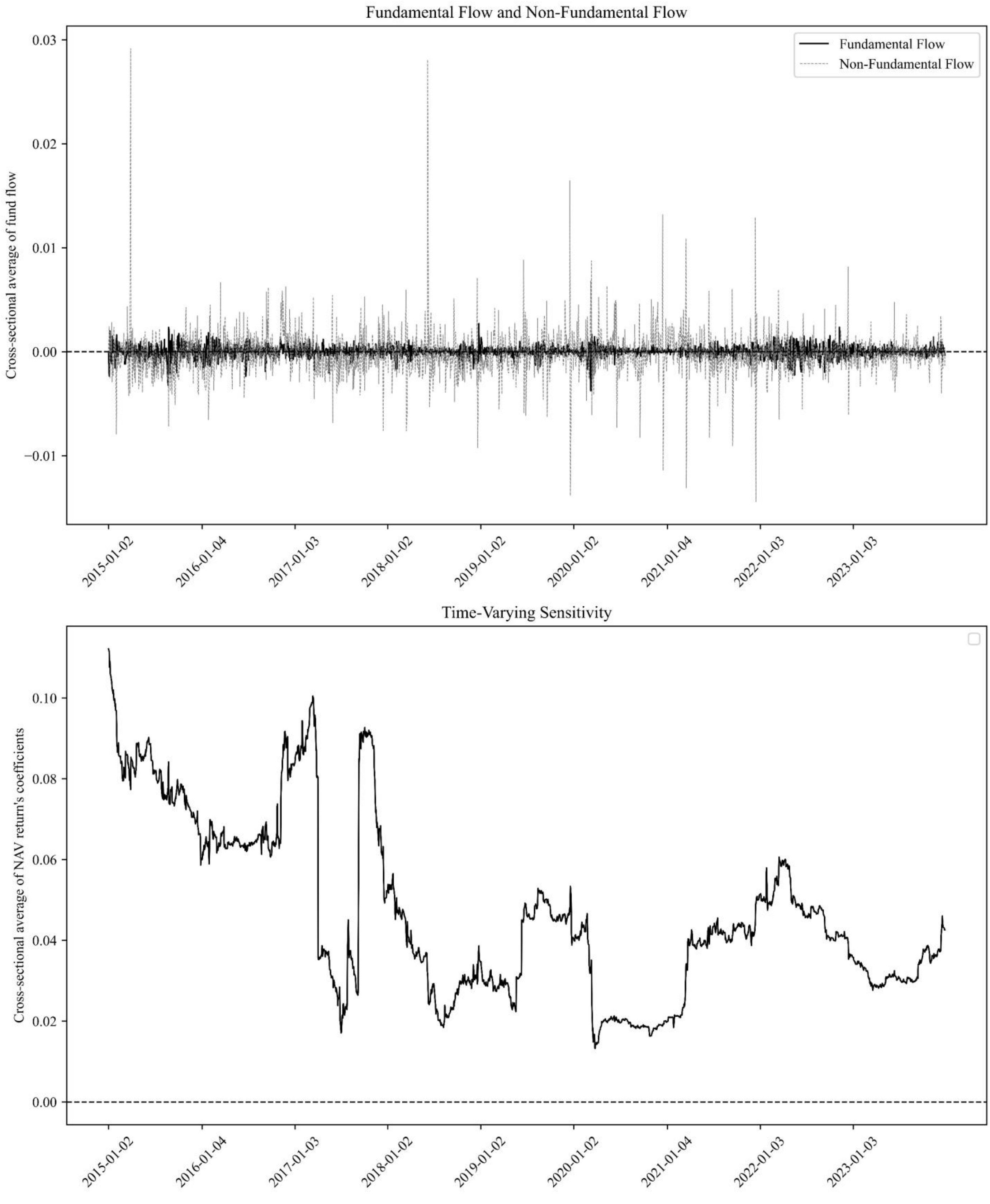

Figure 2. Time-varying fundamental flow, non-fundamental flow, and sensitivity

The top of this figure displays the cross-sectional average of the daily fundamental flow and daily non-fundamental flow. The black solid line represents fundamental flow, while the gray dashed line represents non-fundamental flow. The bottom of this figure presents the cross-sectional average of NAV return's coefficients, estimated by regressing aggregate flow on NAV returns using a one-year rolling window approach. The sample period spans from 2015 to 2023.

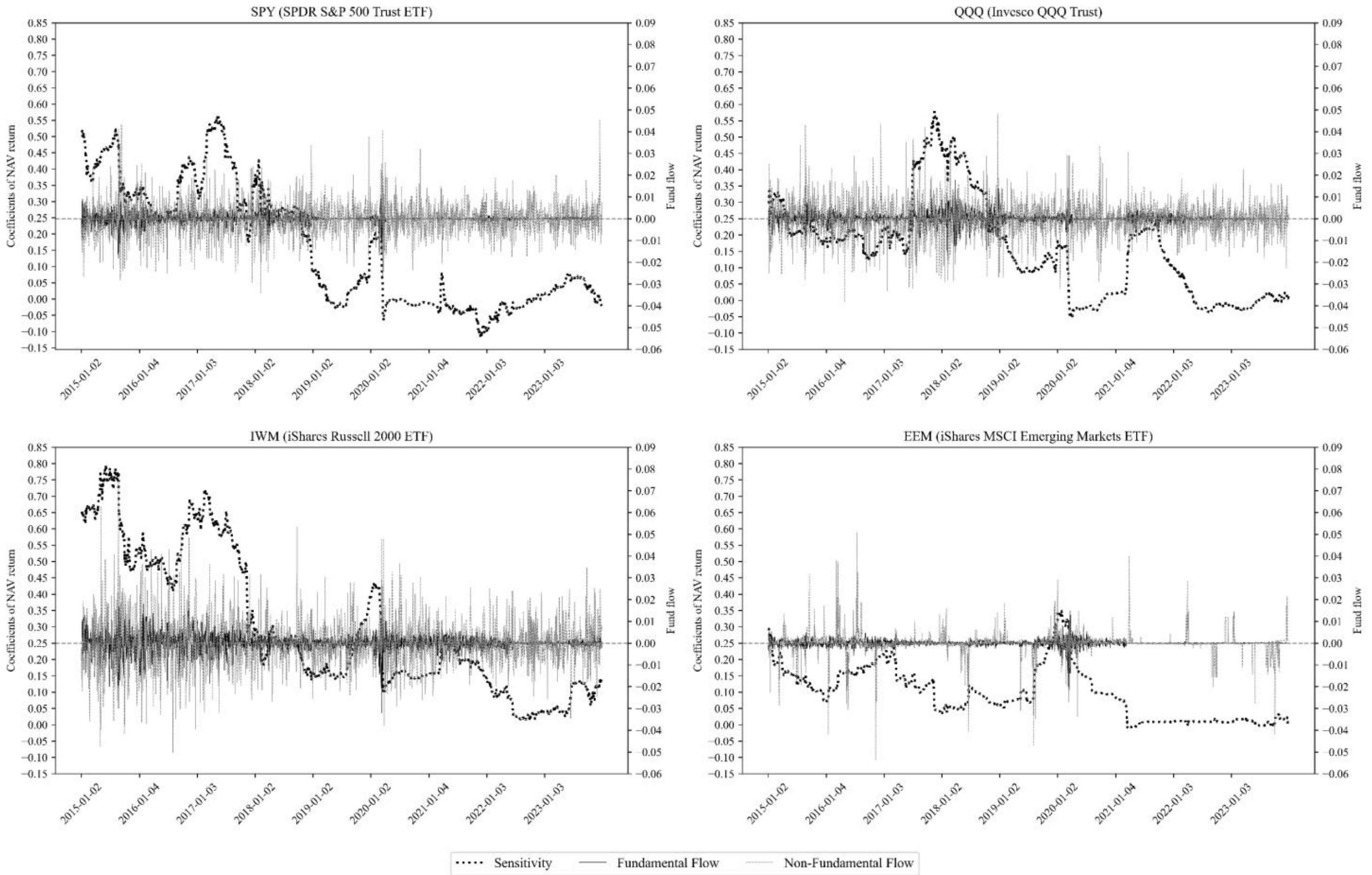

Figure 3. Time-varying fundamental flow, non-fundamental flow, and sensitivity of leading ETFs

This figure displays the daily fundamental flow, non-fundamental flow, and sensitivity of the top four ETFs (excluding sector ETFs) within our sample, ranked by daily dollar trading volume, over the period from 2015 to 2023. The black dotted line, black solid line, and gray dashed line represent sensitivity, fundamental flow, and non-fundamental flow, respectively.

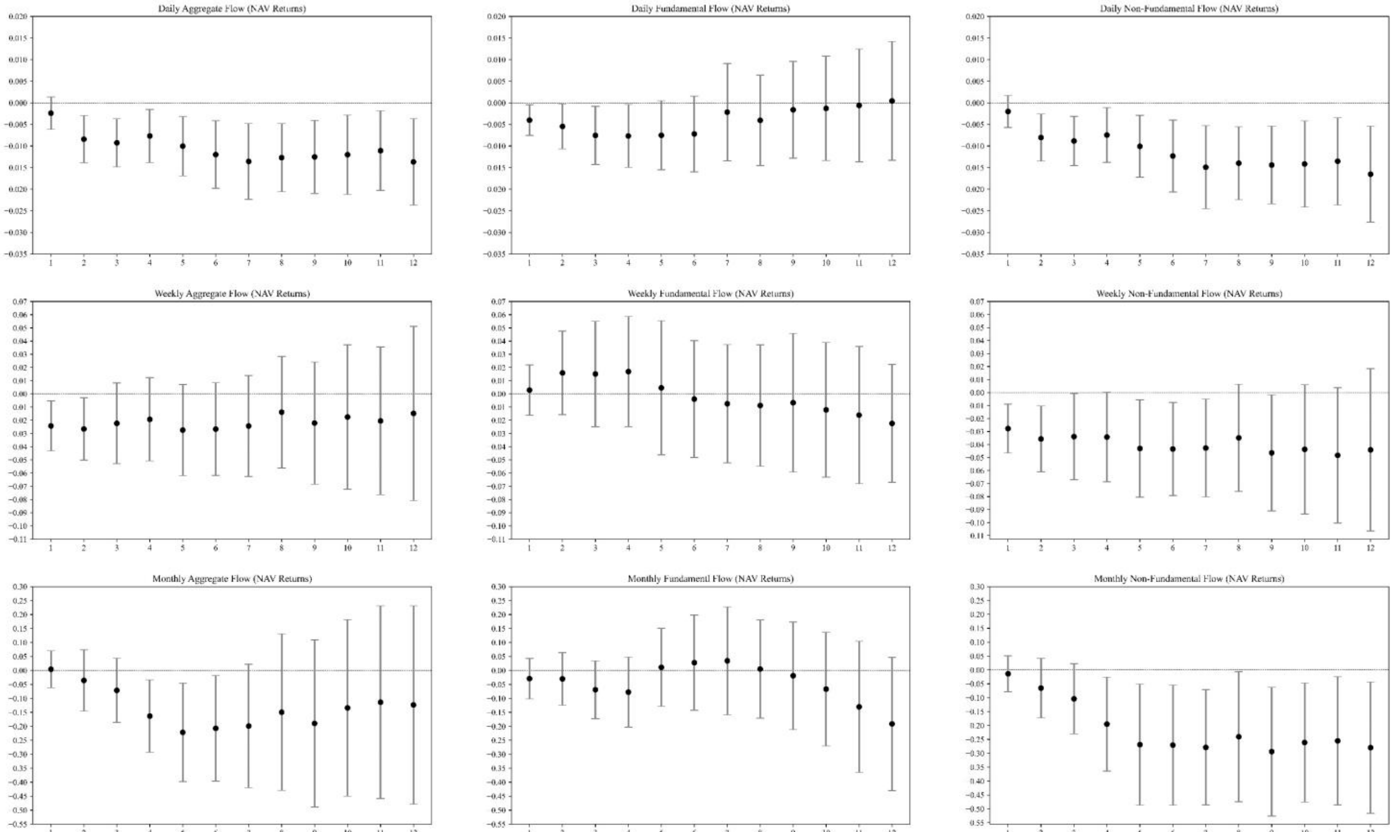

Figure 4. Comparison of return predictability among aggregate flow, fundamental flow, and non-fundamental flow

This figure plots the estimated coefficients and 90% confidence intervals from equation (12) for aggregate flow, fundamental flow, and non-fundamental flow at daily, weekly, and monthly levels. The first row displays results for the daily fund flows, the second row for the weekly fund flows, and the third row for the monthly fund flows, with the results arranged from left to right for aggregate flow, fundamental flow, and non-fundamental flow. 90% confidence intervals are estimated from Driscoll and Kraay's (1998) standard errors using a lag that corresponds to the return horizon.

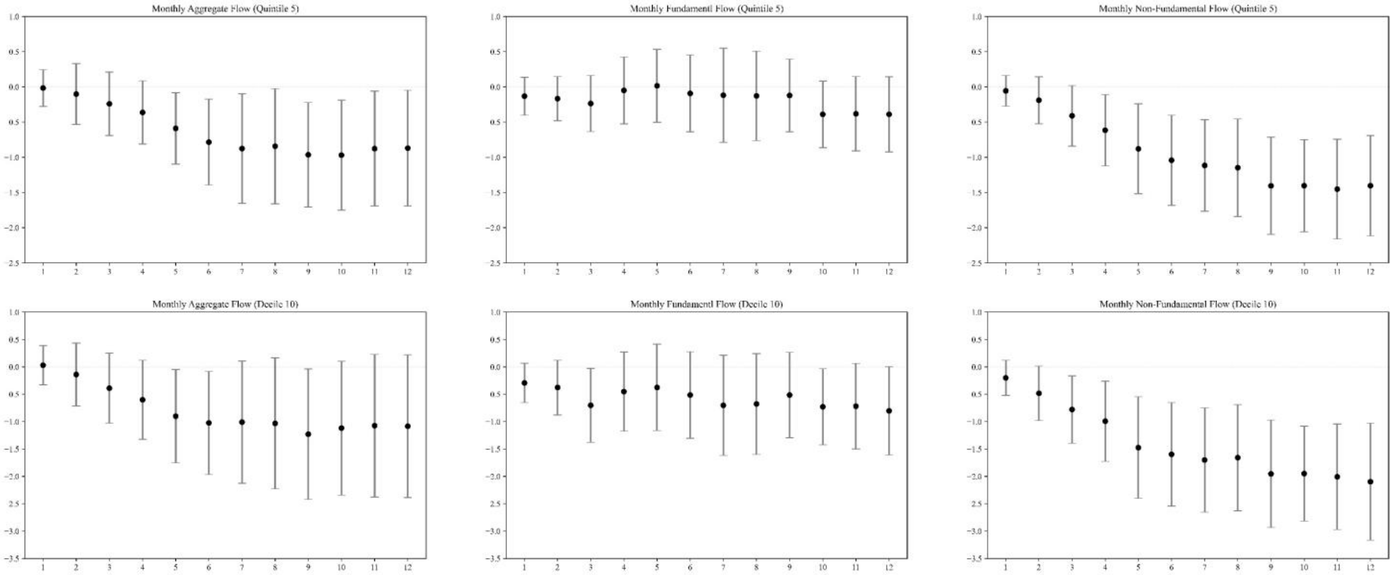

Figure 5. Portfolio-level comparison of return predictability among aggregate flow, fundamental flow, and non-fundamental flow

We construct quintile and decile dummies for monthly aggregate flow, monthly fundamental flow, and monthly non-fundamental flow, and estimate equation (13). The first row of the figure plots the coefficients for Quintile 5, while the second row plots the coefficients for Decile 10, with the error bars representing the 90% confidence intervals. 90% confidence intervals are estimated from Driscoll and Kraay's (1998) standard errors using a lag that corresponds to the return horizon. The left column of the figure presents the results for quantile dummies based on aggregate flow, the middle column of the figure presents the results for quantile dummies based on fundamental flow, and the right column of the figure presents the results for quantile dummies based on non-fundamental flow.

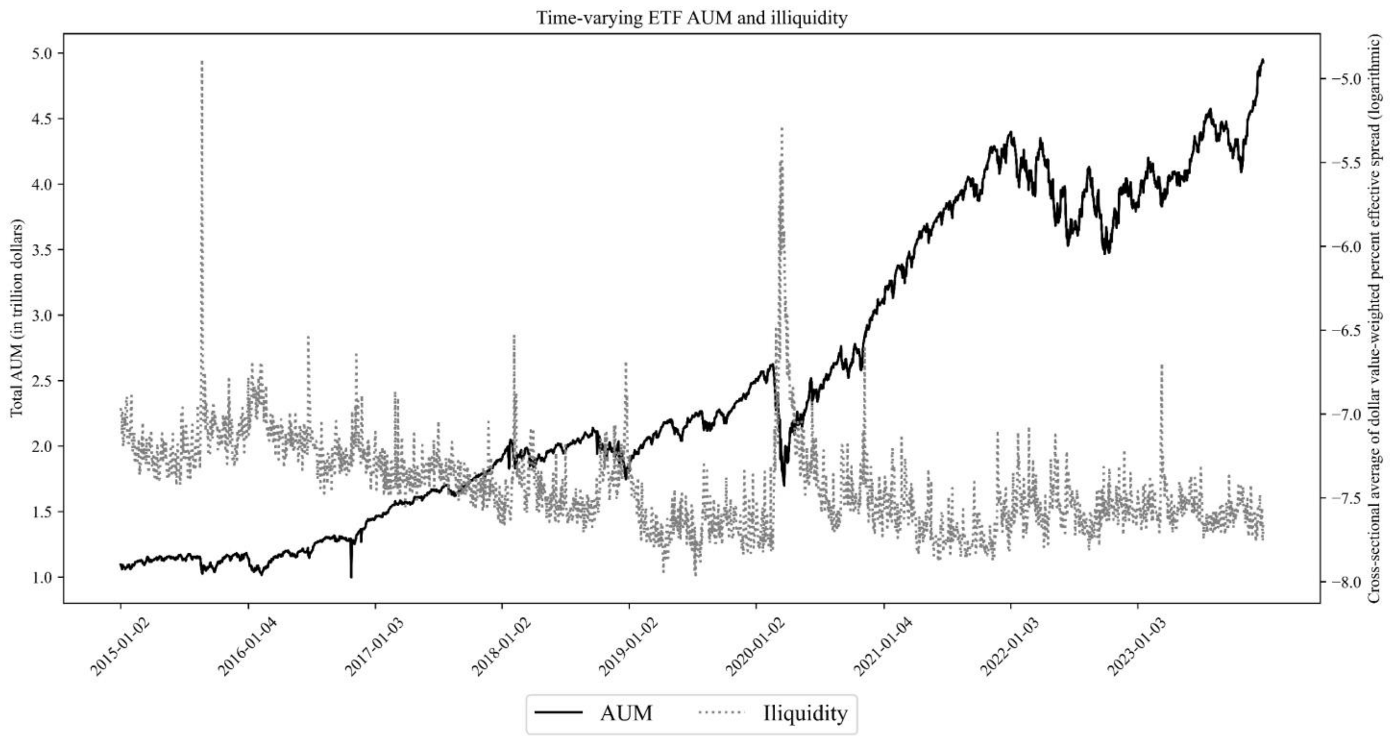

Figure 6. The trend of ETF AUM and illiquidity

This figure displays the daily total AUM and the daily cross-sectional average of illiquidity, measured by the dollar-weighted percent effective spread, within our sample over the period from 2015 to 2023. The black solid line represents AUM, while the gray dotted line represents illiquidity.

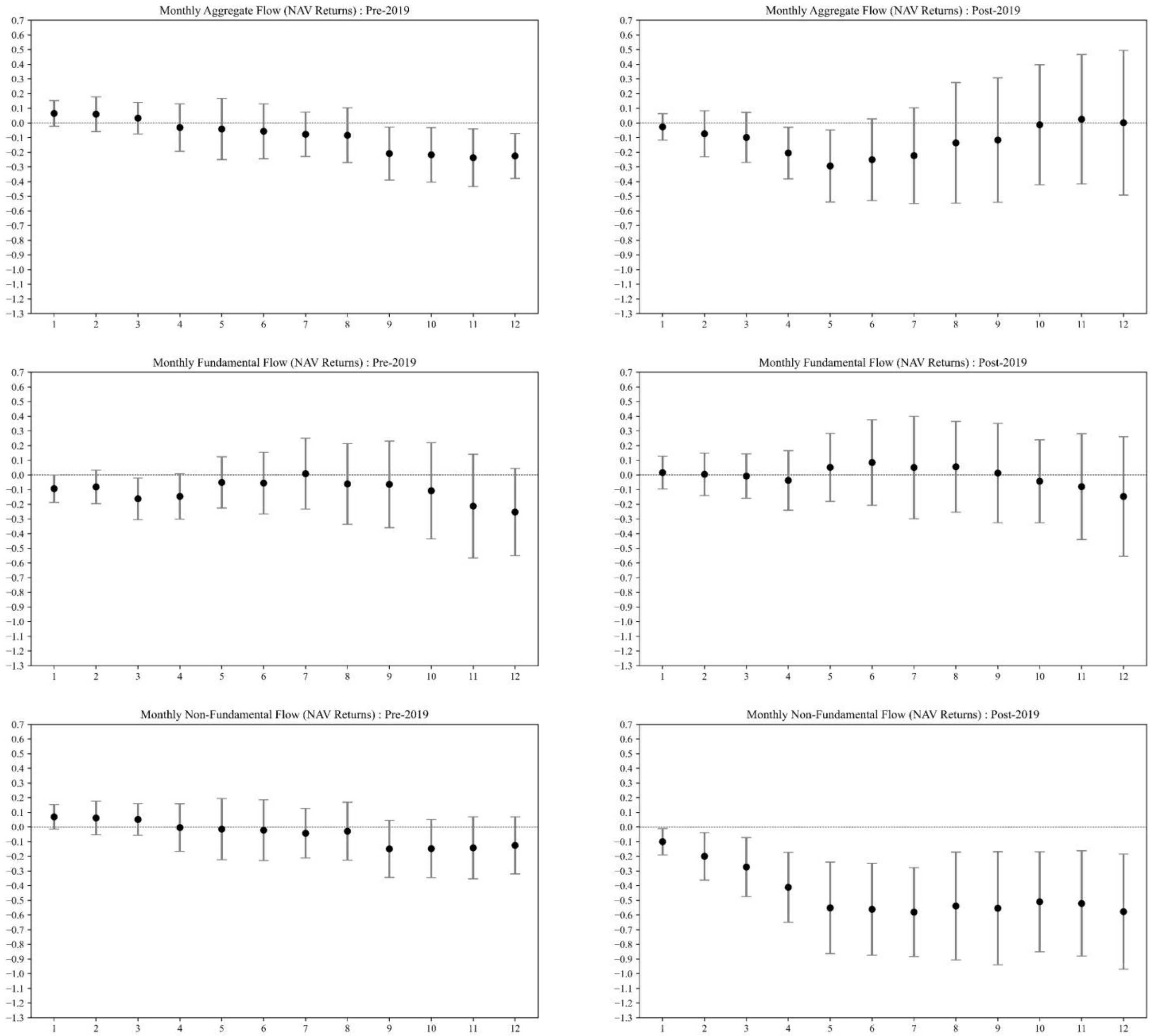

Figure 7. Pre-post comparison of return predictability among aggregate flow, fundamental flow, and non-fundamental flow

This figure plots the estimated coefficients and 90% confidence intervals from equation (12) for monthly aggregate flow, monthly fundamental flow, and monthly non-fundamental flow. Arranged from top to bottom, the figure displays results for monthly aggregate flow, followed by monthly fundamental flow, and then monthly non-fundamental flow. The left side of the figure presents the estimation period before 2019 ('Pre-2019'), while the right side presents the results for the estimation period after 2019 ('Post-2019'). 90% confidence intervals are estimated from Driscoll and Kraay's (1998) standard errors using a lag that corresponds to the return horizon.

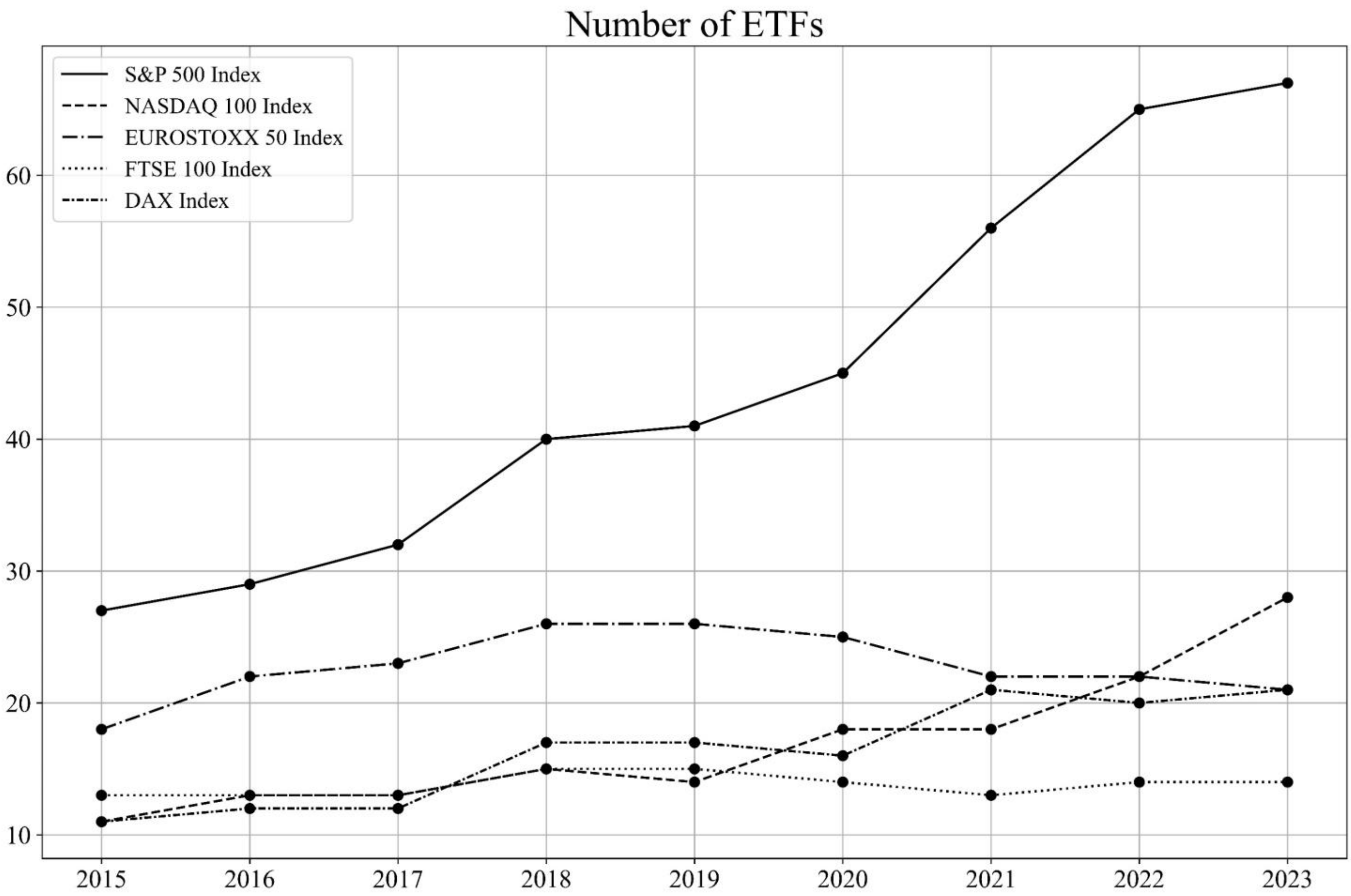

Figure 8. Annual number of global ETFs tracking the same index

This figure represents the annual number of global ETFs tracking each index—S&P 500 index, NASDAQ 100 index, EUROSTOXX 50 index, FTSE 100 index, and DAX index—before applying the ETF filtering procedure.

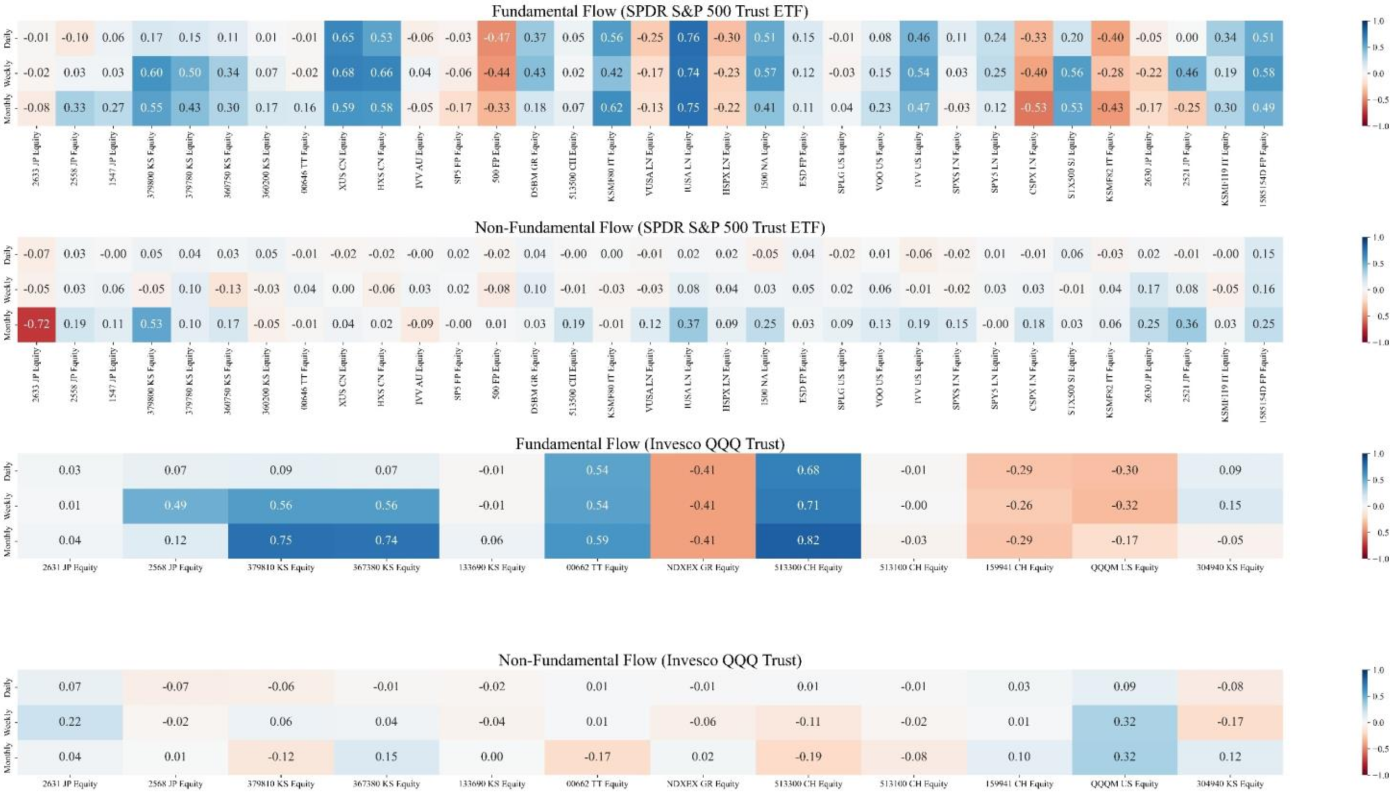

Figure 9. Flow correlation between leading ETFs and other ETFs

The top of this figure displays the correlation of fundamental and non-fundamental flows of SPDR S&P 500 Trust ETF with those of other ETFs tracking the S&P 500 index. The bottom of this figure displays the correlation of fundamental and non-fundamental flows of Invesco QQQ Trust with those of other ETFs tracking the NASDAQ 100 index. Each row in the heatmaps represents daily, weekly, and monthly flows, respectively. Red indicates a negative correlation, while blue indicates a positive correlation.

Table 1. Summary of ETF flows

This table summarizes the information about daily aggregate flow, fundamental flow, and non-fundamental flow. Panel A displays the number of observations ('*Number of Obs*'), average ('*Avg*'), standard deviation ('*Std*'), 25th percentile ('*Q25*'), median ('*Median*'), and 75th percentile ('*Q75*') of daily fund flows. Panel B displays the pairwise correlation between ETF flows.

Panel A. Summary statistics

	Aggregate Flow	*Fundamental Flow*	*Non-Fundamental Flow*
Number of Obs	645,775	645,775	645,775
Avg	0.061%	0.001%	-0.015%
Std	2.369%	0.358%	2.367%
Q25	0.000%	-0.019%	-0.170%
Median	0.000%	0.000%	-0.027%
Q75	0.034%	0.021%	0.102%

Panel B. Pairwise correlations

	Aggregate Flow	*Fundamental Flow*	*Non-Fundamental Flow*
Aggregate Flow	1.000		
Fundamental Flow	0.097	1.000	
Non-Fundamental Flow	0.983	-0.053	1.000

Table 2. Return predictability of daily ETF flows

This table reports the estimated coefficients of daily fund flow variables. We estimate the following panel regression models for horizons from $h = 1$ to $h = 12$:

$$r_{i,t+h} = \beta_0 + \beta_1 FF_{i,t} + \beta_2 X'_{i,t} + \alpha_i + \delta_t + \varepsilon_{i,t+h},$$

$$r_{i,t+h} = \beta_0 + \sum_{d=2}^{p} \beta_{1,d} QUANTILE_{i,d,t} + \beta_2 X'_{i,t} + \alpha_i + \delta_t + \varepsilon_{i,t+h},$$

$$r_{i,t+h} = \sum_{k=t+1}^{t+h} r_{i,k},$$

where $r_{i,t}$ denotes the NAV returns of ETF i at time t, and $FF_{i,t}$ represents the flow variable, indicating either fundamental flow or non-fundamental flow. $QUANTILE_{i,d,t}$ is a decile dummy variable (p is 10), which equals 1 if ETF i falls into the d-th decile group based on the fund flow at time t; otherwise, it equals 0. $X'_{i,t}$ represents a set of control variables, which include two lags of NAV returns, logarithmic dollar trading volume, logarithmic AUM, and dollar value-weighted percent effective spread. The models include both time-fixed effects and ETF-fixed effects. We standardize the control variables (excluding the two lags of NAV returns) and flow variables by subtracting the sample mean and dividing by the sample standard deviation during the estimation period to estimate the panel regression model. Panels A and C display the estimated coefficients for daily fundamental flow and non-fundamental flow ('*FF*'). Panels B and D present the estimated coefficients for the decile dummy variables based on daily fundamental flow and non-fundamental flow ('*Decile d*'). ***, **, and * denote significance levels at 1%, 5%, and 10% based on Driscoll and Kraay's (1998) standard errors using a lag that corresponds to the return horizon, respectively.

	h=1	*h=2*	*h=3*	*h=4*	*h=5*	*h=6*	*h=7*	*h=8*	*h=9*	*h=10*	*h=11*	*h=12*
Panel A. Daily fundamental flows												
FF	-0.004*	-0.005*	-0.008*	-0.008*	-0.008	-0.007	-0.002	-0.004	-0.002	-0.001	-0.001	0.000
	(-1.87)	(-1.73)	(-1.84)	(-1.73)	(-1.55)	(-1.36)	(-0.32)	(-0.64)	(-0.24)	(-0.17)	(-0.08)	(0.05)
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.001	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.003	0.003	0.003
Panel B. Daily fundamental flow decile dummies												
Decile 2	-0.017**	-0.021**	-0.019	-0.030**	-0.033**	-0.040**	-0.042**	-0.044**	-0.049**	-0.050*	-0.047*	-0.045
	(-2.40)	(-2.01)	(-1.46)	(-2.04)	(-1.96)	(-2.17)	(-2.04)	(-1.97)	(-2.05)	(-1.93)	(-1.72)	(-1.56)
Decile 3	-0.008	-0.006	-0.002	-0.008	-0.004	-0.008	-0.009	-0.016	-0.017	-0.021	-0.018	-0.020
	(-1.02)	(-0.51)	(-0.15)	(-0.48)	(-0.20)	(-0.37)	(-0.37)	(-0.60)	(-0.58)	(-0.65)	(-0.52)	(-0.53)

Decile 4	-0.010	-0.008	-0.004	-0.016	-0.018	-0.026	-0.016	-0.033	-0.038	-0.040	-0.041	-0.038
	(-1.08)	(-0.63)	(-0.26)	(-0.84)	(-0.79)	(-1.03)	(-0.58)	(-1.07)	(-1.11)	(-1.10)	(-1.07)	(-0.93)
Decile 5	-0.016*	-0.021	-0.012	-0.024	-0.022	-0.024	-0.019	-0.033	-0.040	-0.043	-0.047	-0.044
	(-1.66)	(-1.46)	(-0.64)	(-1.13)	(-0.86)	(-0.85)	(-0.62)	(-0.98)	(-1.06)	(-1.09)	(-1.13)	(-0.99)
Decile 6	-0.022**	-0.029*	-0.028	-0.045**	-0.046*	-0.052*	-0.036	-0.052	-0.058	-0.058	-0.061	-0.066
	(-2.20)	(-1.86)	(-1.47)	(-2.00)	(-1.78)	(-1.78)	(-1.13)	(-1.47)	(-1.49)	(-1.42)	(-1.38)	(-1.41)
Decile 7	-0.019*	-0.023	-0.024	-0.035*	-0.037	-0.042	-0.031	-0.044	-0.047	-0.045	-0.043	-0.045
	(-1.87)	(-1.58)	(-1.34)	(-1.66)	(-1.51)	(-1.54)	(-1.03)	(-1.32)	(-1.31)	(-1.20)	(-1.06)	(-1.05)
Decile 8	-0.017*	-0.021	-0.019	-0.032*	-0.028	-0.032	-0.023	-0.038	-0.041	-0.039	-0.038	-0.043
	(-1.83)	(-1.56)	(-1.18)	(-1.71)	(-1.26)	(-1.28)	(-0.82)	(-1.28)	(-1.27)	(-1.15)	(-1.05)	(-1.12)
Decile 9	-0.019**	-0.027**	-0.031**	-0.039**	-0.041**	-0.038*	-0.033	-0.043*	-0.046*	-0.048*	-0.046	-0.053*
	(-2.29)	(-2.15)	(-2.00)	(-2.21)	(-2.10)	(-1.76)	(-1.39)	(-1.71)	(-1.70)	(-1.73)	(-1.59)	(-1.75)
Decile 10	-0.018**	-0.022*	-0.025	-0.018	-0.020	-0.017	-0.006	-0.007	-0.006	-0.010	0.003	-0.007
	(-2.05)	(-1.73)	(-1.58)	(-1.01)	(-0.98)	(-0.82)	(-0.27)	(-0.28)	(-0.23)	(-0.36)	(0.10)	(-0.25)
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.001	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.003	0.003	0.003

Table 2. (Continued)

	h=1	*h=2*	*h=3*	*h=4*	*h=5*	*h=6*	*h=7*	*h=8*	*h=9*	*h=10*	*h=11*	*h=12*
Panel C. Daily non-fundamental flows												
FF	-0.002	-0.008**	-0.009**	-0.007*	-0.010**	-0.012**	-0.015**	-0.014***	-0.014***	-0.014**	-0.014**	-0.017**
	(-0.89)	(-2.43)	(-2.55)	(-1.93)	(-2.32)	(-2.43)	(-2.55)	(-2.72)	(-2.63)	(-2.33)	(-2.21)	(-2.45)
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.001	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.003	0.003	0.003
Panel D. Daily non-fundamental flow based decile dummies												
Decile 2	-0.007	-0.019**	-0.017	-0.023	-0.020	-0.020	-0.027	-0.030	-0.037	-0.045*	-0.050*	-0.057*
	(-1.25)	(-2.18)	(-1.49)	(-1.76)	(-1.37)	(-1.25)	(-1.50)	(-1.49)	(-1.64)	(-1.85)	(-1.87)	(-1.93)
Decile 3	-0.006	-0.017*	-0.018	-0.024*	-0.025	-0.032*	-0.040**	-0.043*	-0.050*	-0.055**	-0.059**	-0.066**
	(-0.92)	(-1.81)	(-1.55)	(-1.71)	(-1.56)	(-1.82)	(-2.02)	(-1.89)	(-1.95)	(-1.99)	(-1.97)	(-2.04)
Decile 4	-0.005	-0.019**	-0.022*	-0.025*	-0.017	-0.020	-0.028	-0.026	-0.028	-0.030	-0.034	-0.036
	(-0.84)	(-2.01)	(-1.77)	(-1.73)	(-1.03)	(-1.04)	(-1.34)	(-1.12)	(-1.09)	(-1.07)	(-1.13)	(-1.11)
Decile 5	-0.006	-0.016*	-0.022*	-0.026*	-0.026	-0.028	-0.032	-0.032	-0.043	-0.047	-0.054	-0.054
	(-0.96)	(-1.70)	(-1.75)	(-1.74)	(-1.54)	(-1.45)	(-1.48)	(-1.34)	(-1.54)	(-1.56)	(-1.63)	(-1.51)
Decile 6	-0.008	-0.021**	-0.026**	-0.028*	-0.026	-0.036*	-0.038*	-0.036	-0.046*	-0.056*	-0.062**	-0.064*
	(-1.31)	(-2.22)	(-2.07)	(-1.85)	(-1.51)	(-1.87)	(-1.77)	(-1.53)	(-1.76)	(-1.92)	(-1.96)	(-1.88)
Decile 7	0.003	-0.005	-0.013	-0.020	-0.024	-0.023	-0.033	-0.030	-0.035	-0.041	-0.047	-0.050
	(0.44)	(-0.52)	(-1.01)	(-1.32)	(-1.32)	(-1.12)	(-1.43)	(-1.16)	(-1.23)	(-1.33)	(-1.38)	(-1.38)
Decile 8	-0.001	-0.007	-0.015	-0.021	-0.023	-0.031	-0.045*	-0.047*	-0.056*	-0.067**	-0.072**	-0.080**
	(-0.13)	(-0.67)	(-1.09)	(-1.22)	(-1.13)	(-1.41)	(-1.82)	(-1.72)	(-1.89)	(-2.02)	(-2.05)	(-2.11)
Decile 9	-0.004	-0.017	-0.023	-0.039**	-0.048**	-0.050**	-0.066***	-0.068	-0.078**	-0.089**	-0.098**	-0.103**
	(-0.58)	(-1.61)	(-1.58)	(-2.30)	(-2.42)	(-2.21)	(-2.63)	(-2.43)	(-2.47)	(-2.56)	(-2.57)	(-2.47)
Decile 10	-0.003	-0.019	-0.028*	-0.032*	-0.043**	-0.054**	-0.059**	-0.059**	-0.076**	-0.084**	-0.091**	-0.099**
	(-0.45)	(-1.62)	(-1.81)	(-1.71)	(-2.04)	(-2.26)	(-2.23)	(-2.00)	(-2.27)	(-2.27)	(-2.33)	(-2.33)
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.001	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.003	0.003	0.003	0.003

Table 3. Return predictability of weekly ETF flows

This table reports the estimated coefficients of weekly fund flow variables. We estimate the following panel regression models for horizons from $h = 1$ to $h = 12$:

$$r_{i,t+h} = \beta_0 + \beta_1 FF_{i,t} + \beta_2 X'_{i,t} + \alpha_i + \delta_t + \varepsilon_{i,t+h},$$

$$r_{i,t+h} = \beta_0 + \sum_{d=2}^{p} \beta_{1,d} QUANTILE_{i,d,t} + \beta_2 X'_{i,t} + \alpha_i + \delta_t + \varepsilon_{i,t+h},$$

$$r_{i,t+h} = \sum_{k=t+1}^{t+h} r_{i,k},$$

where $r_{i,t}$ denotes the NAV returns of ETF i at time t, and $FF_{i,t}$ represents the flow variable, indicating either fundamental flow or non-fundamental flow. $QUANTILE_{i,d,t}$ is a decile dummy variable (p is 10), which equals 1 if ETF i falls into the d-th decile group based on the fund flow at time t; otherwise, it equals 0. $X'_{i,t}$ represents a set of control variables, which include two lags of NAV returns, logarithmic dollar trading volume, logarithmic AUM, and dollar value-weighted percent effective spread. The models include both time-fixed effects and ETF-fixed effects. We standardize the control variables (excluding the two lags of NAV returns) and flow variables by subtracting the sample mean and dividing by the sample standard deviation during the estimation period to estimate the panel regression model. Panels A and C display the estimated coefficients for weekly fundamental flow and non-fundamental flow ('*FF*'). Panels B and D present the estimated coefficients for the decile dummy variables based on weekly fundamental flow and non-fundamental flow ('*Decile d*'). ***, **, and * denote significance levels at 1%, 5%, and 10% based on Driscoll and Kraay's (1998) standard errors using a lag that corresponds to the return horizon, respectively.

	h=1	*h=2*	*h=3*	*h=4*	*h=5*	*h=6*	*h=7*	*h=8*	*h=9*	*h=10*	*h=11*	*h=12*
Panel A. Weekly fundamental flows												
FF	0.003	0.016	0.015	0.017	0.005	-0.004	-0.007	-0.009	-0.007	-0.012	-0.016	-0.022
	(0.24)	(0.83)	(0.62)	(0.67)	(0.15)	(-0.15)	(-0.27)	(-0.32)	(-0.21)	(-0.39)	(-0.51)	(-0.82)
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.002	0.003	0.004	0.005	0.007	0.008	0.008	0.010	0.011	0.012	0.014	0.014
Panel B. Weekly fundamental flow based decile dummies												
Decile 2	-0.014	-0.043	-0.048	-0.096	-0.087	-0.095	-0.115	-0.148	-0.126	-0.092	-0.104	-0.159

	(-0.46)	(-0.91)	(-0.81)	(-1.43)	(-1.05)	(-0.99)	(-1.05)	(-1.28)	(-0.97)	(-0.67)	(-0.72)	(-1.04)
Decile 3	-0.010	0.004	-0.029	-0.074	-0.045	-0.045	-0.074	-0.077	-0.066	-0.063	-0.113	-0.134
	(-0.26)	(0.07)	(-0.42)	(-0.94)	(-0.46)	(-0.40)	(-0.58)	(-0.56)	(-0.42)	(-0.38)	(-0.64)	(-0.72)
Decile 4	-0.037	-0.043	-0.052	-0.137	-0.118	-0.124	-0.156	-0.152	-0.147	-0.101	-0.115	-0.153
	(-0.94)	(-0.76)	(-0.71)	(-1.64)	(-1.11)	(-1.02)	(-1.18)	(-1.07)	(-0.92)	(-0.60)	(-0.63)	(-0.77)
Decile 5	-0.018	-0.030	-0.088	-0.152	-0.132	-0.168	-0.176	-0.210	-0.202	-0.184	-0.185	-0.243
	(-0.42)	(-0.45)	(-1.08)	(-1.63)	(-1.13)	(-1.23)	(-1.21)	(-1.29)	(-1.11)	(-0.94)	(-0.91)	(-1.11)
Decile 6	-0.038	-0.035	-0.078	-0.154	-0.144	-0.147	-0.172	-0.173	-0.163	-0.134	-0.124	-0.157
	(-0.83)	(-0.52)	(-0.91)	(-1.48)	(-1.12)	(-1.05)	(-1.16)	(-1.06)	(-0.88)	(-0.68)	(-0.59)	(-0.70)
Decile 7	-0.024	0.005	-0.020	-0.082	-0.098	-0.128	-0.161	-0.163	-0.173	-0.151	-0.182	-0.232
	(-0.56)	(0.08)	(-0.25)	(-0.91)	(-0.87)	(-1.03)	(-1.19)	(-1.10)	(-1.00)	(-0.83)	(-0.95)	(-1.13)
Decile 8	0.001	0.025	0.007	-0.067	-0.084	-0.114	-0.124	-0.134	-0.138	-0.121	-0.114	-0.157
	(0.02)	(0.40)	(0.09)	(-0.79)	(-0.82)	(-0.97)	(-0.95)	(-0.93)	(-0.86)	(-0.70)	(-0.63)	(-0.85)
Decile 9	-0.048	-0.069	-0.050	-0.133^{*}	-0.178^{*}	-0.185^{*}	-0.223^{**}	-0.225^{*}	-0.233^{*}	-0.239^{*}	-0.277^{*}	-0.318^{**}
	(-1.25)	(-1.25)	(-0.72)	(-1.77)	(-1.86)	(-1.76)	(-1.97)	(-1.80)	(-1.72)	(-1.67)	(-1.87)	(-2.15)
Decile 10	0.052	0.067	0.061	0.061	-0.028	-0.043	-0.027	-0.017	-0.029	-0.043	-0.105	-0.130
	(1.21)	(1.04)	(0.81)	(0.68)	(-0.27)	(-0.36)	(-0.21)	(-0.12)	(-0.18)	(-0.26)	(-0.63)	(-0.73)
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.002	0.003	0.004	0.005	0.007	0.007	0.008	0.010	0.011	0.012	0.014	0.014

Table 3. (Continued)

	h=1	*h=2*	*h=3*	*h=4*	*h=5*	*h=6*	*h=7*	*h=8*	*h=9*	*h=10*	*h=11*	*h=12*
Panel C. Weekly non-fundamental flows												
FF	-0.028**	-0.036**	-0.034*	-0.034	-0.043*	-0.043**	-0.043*	-0.035	-0.046*	-0.044	-0.048	-0.044
	(-2.43)	(-2.31)	(-1.68)	(-1.64)	(-1.89)	(-1.99)	(-1.87)	(-1.39)	(-1.71)	(-1.44)	(-1.53)	(-1.16)
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.002	0.003	0.004	0.005	0.007	0.008	0.009	0.010	0.011	0.012	0.014	0.014
Panel D. Weekly non-fundamental flow based decile dummies												
Decile 2	-0.044	-0.059	-0.065	-0.120*	-0.135*	-0.171**	-0.267***	-0.293***	-0.307***	-0.323***	-0.317***	-0.364***
	(-1.45)	(-1.49)	(-1.25)	(-1.76)	(-1.81)	(-2.24)	(-3.10)	(-3.16)	(-3.04)	(-3.12)	(-2.82)	(-3.00)
Decile 3	-0.065**	-0.120***	-0.137**	-0.180**	-0.231**	-0.306***	-0.370***	-0.387***	-0.431***	-0.477***	-0.506***	-0.567***
	(-2.19)	(-2.60)	(-2.14)	(-2.30)	(-2.56)	(-3.30)	(-3.60)	(-3.57)	(-3.63)	(-3.72)	(-3.59)	(-3.62)
Decile 4	-0.049	-0.076	-0.099	-0.134*	-0.168*	-0.227**	-0.289***	-0.289***	-0.318***	-0.319***	-0.318**	-0.372***
	(-1.56)	(-1.63)	(-1.60)	(-1.70)	(-1.89)	(-2.44)	(-2.88)	(-2.83)	(-2.98)	(-2.69)	(-2.50)	(-2.69)
Decile 5	-0.049	-0.044	-0.065	-0.106	-0.162*	-0.200**	-0.252**	-0.268**	-0.288**	-0.315**	-0.360**	-0.426**
	(-1.62)	(-0.98)	(-1.09)	(-1.37)	(-1.69)	(-1.99)	(-2.26)	(-2.27)	(-2.35)	(-2.36)	(-2.37)	(-2.58)
Decile 6	-0.054*	-0.067	-0.090	-0.131	-0.155	-0.180*	-0.243**	-0.235*	-0.253**	-0.275**	-0.300**	-0.390**
	(-1.66)	(-1.41)	(-1.42)	(-1.59)	(-1.61)	(-1.71)	(-2.07)	(-1.95)	(-2.03)	(-2.02)	(-2.04)	(-2.38)
Decile 7	-0.063**	-0.093*	-0.113*	-0.167**	-0.211**	-0.251**	-0.292**	-0.291**	-0.326**	-0.344**	-0.388**	-0.463***
	(-2.02)	(-1.93)	(-1.73)	(-2.02)	(-2.18)	(-2.39)	(-2.53)	(-2.39)	(-2.46)	(-2.33)	(-2.40)	(-2.66)
Decile 8	-0.077**	-0.099**	-0.122*	-0.170*	-0.219**	-0.221*	-0.282**	-0.294**	-0.300**	-0.345**	-0.402**	-0.478***
	(-2.37)	(-1.97)	(-1.78)	(-1.88)	(-2.09)	(-1.92)	(-2.23)	(-2.17)	(-2.07)	(-2.19)	(-2.41)	(-2.62)
Decile 9	-0.066*	-0.097*	-0.151**	-0.210**	-0.255**	-0.314***	-0.382***	-0.374***	-0.401***	-0.426***	-0.469***	-0.565***
	(-1.81)	(-1.83)	(-2.04)	(-2.17)	(-2.31)	(-2.71)	(-3.04)	(-2.84)	(-2.98)	(-2.93)	(-2.99)	(-3.33)
Decile 10	-0.089**	-0.111*	-0.132	-0.190*	-0.221*	-0.243*	-0.325**	-0.314**	-0.355**	-0.374**	-0.399**	-0.453**
	(-2.17)	(-1.75)	(-1.57)	(-1.94)	(-1.92)	(-1.90)	(-2.33)	(-2.06)	(-2.18)	(-2.14)	(-2.13)	(-2.20)
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.002	0.003	0.004	0.005	0.007	0.008	0.009	0.010	0.011	0.012	0.014	0.015

Table 4. Return predictability of monthly ETF flows

This table reports the estimated coefficients of monthly fund flow variables. We estimate the following panel regression models for horizons from $h = 1$ to $h = 12$:

$$r_{i,t+h} = \beta_0 + \beta_1 FF_{i,t} + \beta_2 X'_{i,t} + \alpha_i + \delta_t + \varepsilon_{i,t+h},$$

$$r_{i,t+h} = \beta_0 + \sum_{d=2}^{p} \beta_{1,d} QUANTILE_{i,d,t} + \beta_2 X'_{i,t} + \alpha_i + \delta_t + \varepsilon_{i,t+h},$$

$$r_{i,t+h} = \sum_{k=t+1}^{t+h} r_{i,k},$$

where $r_{i,t}$ denotes the NAV returns of ETF i at time t, and $FF_{i,t}$ represents the flow variable, indicating either fundamental flow or non-fundamental flow. $QUANTILE_{i,d,t}$ is a decile dummy variable (p is 10), which equals 1 if ETF i falls into the d-th decile group based on the fund flow at time t; otherwise, it equals 0. $X'_{i,t}$ represents a set of control variables, which include two lags of NAV returns, logarithmic dollar trading volume, logarithmic AUM, and dollar value-weighted percent effective spread. The models include both time-fixed effects and ETF-fixed effects. We standardize the control variables (excluding the two lags of NAV returns) and flow variables by subtracting the sample mean and dividing by the sample standard deviation during the estimation period to estimate the panel regression model. Panels A and C display the estimated coefficients for monthly fundamental flow and non-fundamental flow ('*FF*'). Panels B and D present the estimated coefficients for the decile dummy variables based on monthly fundamental flow and non-fundamental flow ('*Decile d*'). ***, **, and * denote significance levels at 1%, 5%, and 10% based on Driscoll and Kraay's (1998) standard errors using a lag that corresponds to the return horizon, respectively.

	h=1	*h*=2	*h*=3	*h*=4	*h*=5	*h*=6	*h*=7	*h*=8	*h*=9	*h*=10	*h*=11	*h*=12
Panel A. Monthly fundamental flows												
FF	-0.029	-0.030	-0.069	-0.078	0.011	0.028	0.034	0.005	-0.019	-0.067	-0.130	-0.192
	(-0.66)	(-0.53)	(-1.10)	(-1.02)	(0.13)	(0.27)	(0.29)	(0.05)	(-0.16)	(-0.54)	(-0.91)	(-1.32)
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.006	0.011	0.017	0.023	0.028	0.033	0.038	0.043	0.048	0.054	0.060	0.065

Panel B. Monthly fundamental flow based decile dummies

Decile 2	-0.261**	-0.312	-0.580***	-0.802***	-0.730***	-0.863***	-1.000***	-1.105***	-0.943**	-0.936**	-0.962*	-1.112*
	(-1.97)	(-1.43)	(-2.96)	(-4.30)	(-3.25)	(-3.27)	(-3.06)	(-2.84)	(-2.08)	(-2.07)	(-1.82)	(-1.85)
Decile 3	-0.210	-0.108	-0.079	-0.199	0.000	-0.100	-0.113	-0.169	-0.152	-0.327	-0.510	-0.694
	(-1.46)	(-0.49)	(-0.35)	(-0.80)	(-0.00)	(-0.25)	(-0.23)	(-0.30)	(-0.25)	(-0.52)	(-0.73)	(-0.94)
Decile 4	-0.178	-0.269	-0.327	-0.463	-0.078	-0.121	-0.050	-0.198	-0.111	-0.171	-0.390	-0.583
	(-1.09)	(-1.08)	(-1.43)	(-1.82)	(-0.23)	(-0.29)	(-0.09)	(-0.32)	(-0.16)	(-0.24)	(-0.50)	(-0.78)
Decile 5	-0.204	-0.186	-0.256	-0.316	0.039	0.035	-0.010	-0.033	-0.242	-0.398	-0.646	-0.854
	(-1.12)	(-0.65)	(-1.03)	(-1.16)	(0.12)	(0.11)	(-0.02)	(-0.07)	(-0.46)	(-0.70)	(-1.01)	(-1.26)
Decile 6	-0.286*	-0.315	-0.356	-0.473**	-0.163	-0.205	-0.083	-0.011	-0.061	-0.210	-0.413	-0.564
	(-1.84)	(-1.30)	(-1.60)	(-2.05)	(-0.69)	(-0.73)	(-0.22)	(-0.03)	(-0.12)	(-0.41)	(-0.70)	(-0.96)
Decile 7	-0.355**	-0.429*	-0.448*	-0.547**	-0.251	-0.388	-0.383	-0.384	-0.247	-0.339	-0.327	-0.566
	(-2.20)	(-1.71)	(-1.92)	(-2.07)	(-0.80)	(-1.06)	(-0.81)	(-0.73)	(-0.42)	(-0.55)	(-0.47)	(-0.76)
Decile 8	-0.270*	-0.272	-0.220	-0.394	-0.195	-0.298	-0.198	-0.296	-0.265	-0.459	-0.691	-0.772
	(-1.73)	(-1.11)	(-0.89)	(-1.38)	(-0.58)	(-0.75)	(-0.36)	(-0.50)	(-0.42)	(-0.75)	(-0.95)	(-1.10)
Decile 9	-0.229	-0.257	-0.326	-0.422	-0.281	-0.458	-0.468	-0.598	-0.593	-0.896*	-0.922	-0.995
	(-1.22)	(-1.03)	(-1.46)	(-1.44)	(-0.84)	(-1.16)	(-0.91)	(-1.07)	(-1.05)	(-1.73)	(-1.59)	(-1.61)
Decile 10	-0.292	-0.377	-0.703*	-0.451	-0.375	-0.513	-0.702	-0.678	-0.514	-0.728*	-0.719	-0.803
	(-1.34)	(-1.24)	(-1.71)	(-1.03)	(-0.78)	(-1.07)	(-1.26)	(-1.21)	(-1.08)	(-1.72)	(-1.51)	(-1.64)
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.007	0.013	0.020	0.026	0.031	0.037	0.041	0.046	0.050	0.055	0.060	0.065

Table 4. (Continued)

	h=1	*h=2*	*h=3*	*h=4*	*h=5*	*h=6*	*h=7*	*h=8*	*h=9*	*h=10*	*h=11*	*h=12*
Panel C. Monthly non-fundamental flows												
FF	-0.014	-0.066	-0.104	-0.195*	-0.269**	-0.270**	-0.279**	-0.241*	-0.294**	-0.261**	-0.255*	-0.280*
	(-0.35)	(-1.00)	(-1.36)	(-1.90)	(-2.03)	(-2.06)	(-2.21)	(-1.69)	(-2.09)	(-2.01)	(-1.82)	(-1.95)
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.006	0.011	0.017	0.023	0.029	0.034	0.038	0.043	0.049	0.054	0.060	0.065
Panel D. Monthly non-fundamental flow based decile dummies												
Decile 2	-0.337**	-0.593**	-0.570*	-0.745*	-0.998**	-0.789	-0.736	-0.685	-0.749	-0.893*	-0.958*	-1.044*
	(-2.49)	(-2.56)	(-1.91)	(-1.93)	(-2.02)	(-1.53)	(-1.30)	(-1.22)	(-1.33)	(-1.75)	(-1.77)	(-1.80)
Decile 3	-0.229*	-0.480**	-0.502	-0.672	-0.966*	-0.731	-0.617	-0.692	-0.835	-0.850	-0.976*	-0.851
	(-1.76)	(-2.13)	(-1.60)	(-1.55)	(-1.82)	(-1.37)	(-1.05)	(-1.26)	(-1.51)	(-1.63)	(-1.76)	(-1.42)
Decile 4	-0.176	-0.540**	-0.711**	-1.006**	-1.332**	-1.340**	-1.321**	-1.255*	-1.435**	-1.460**	-1.561**	-1.384*
	(-1.12)	(-2.20)	(-2.34)	(-2.33)	(-2.40)	(-2.28)	(-2.09)	(-1.90)	(-2.22)	(-2.21)	(-2.33)	(-1.95)
Decile 5	-0.197	-0.467*	-0.571	-0.775*	-1.123**	-1.012	-1.034	-1.162	-1.348*	-1.328*	-1.425*	-1.455*
	(-1.19)	(-1.80)	(-1.64)	(-1.70)	(-1.97)	(-1.63)	(-1.44)	(-1.55)	(-1.82)	(-1.75)	(-1.78)	(-1.68)
Decile 6	-0.220	-0.378	-0.481	-0.749	-1.121**	-1.149*	-1.104	-1.103	-1.323*	-1.293*	-1.241*	-1.206
	(-1.37)	(-1.35)	(-1.33)	(-1.57)	(-1.97)	(-1.81)	(-1.57)	(-1.50)	(-1.74)	(-1.74)	(-1.70)	(-1.54)
Decile 7	-0.143	-0.323	-0.494	-0.762*	-1.083**	-1.023*	-1.021	-1.145*	-1.322**	-1.322**	-1.281*	-1.218*
	(-0.88)	(-1.20)	(-1.60)	(-1.75)	(-2.09)	(-1.84)	(-1.58)	(-1.70)	(-2.01)	(-2.02)	(-1.85)	(-1.65)
Decile 8	-0.181	-0.305	-0.540	-0.904**	-1.206**	-1.146*	-1.110	-1.140	-1.383*	-1.491**	-1.567**	-1.368*
	(-1.11)	(-1.13)	(-1.59)	(-2.00)	(-2.12)	(-1.86)	(-1.57)	(-1.53)	(-1.89)	(-2.05)	(-2.08)	(-1.78)
Decile 9	-0.226	-0.457*	-0.583	-0.947**	-1.232**	-1.222**	-1.201**	-1.254**	-1.529***	-1.684***	-1.759***	-1.653***
	(-1.40)	(-1.66)	(-1.64)	(-2.11)	(-2.23)	(-2.12)	(-1.99)	(-2.03)	(-2.65)	(-2.89)	(-2.90)	(-2.70)
Decile 10	-0.200	-0.483	-0.778**	-0.992**	-1.474***	-1.596***	-1.701***	-1.658***	-1.953***	-1.949***	-2.007***	-2.097***
	(-1.02)	(-1.60)	(-2.08)	(-2.23)	(-2.61)	(-2.77)	(-2.94)	(-2.81)	(-3.28)	(-3.70)	(-3.42)	(-3.23)
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.007	0.013	0.020	0.026	0.033	0.038	0.042	0.046	0.052	0.056	0.061	0.066

Table 5. Flow commonality of aggregate flows

This table reports the flow commonality estimated by R-squared from a regression model and pairwise correlation of aggregate flows in ETFs tracking the same index. For each ETF, we estimate the R-squared using the following regression model:

$$FF_{i,t} = c_0 + c_1 FF_{m,t-1} + c_2 FF_{m,t} + c_3 FF_{m,t+1} + \varepsilon_t,$$

where $FF_{i,t}$ represents aggregate flow in ETF i at time t, $FF_{m,t}$ is market ETF flow at time t. For each underlying index, when calculating the market ETF flow for ETF i, we use the equal-weighted average of the ETF flows, excluding ETF i. Panels A, B, and C report the results of daily, weekly, and monthly aggregate flow, respectively. For each underlying index (S&P 500 index, NASDAQ 100 index, EUROSTOXX 50 index, FTSE 100 index, and DAX index), each row presents the number of ETFs (‘*N of ETFs*’), the average of R-squared (‘*Avg of R^2*’), the median of R-squared (‘*Median of R^2*’), the standard deviation of R-squared (‘*Std of R^2*’), the number of pairs (‘*N of Pairs*’), the number of positively correlated pairs (‘*N of (+)*’), the number of negatively correlated pairs (‘*N of (-)*’), and the average of absolute pairwise correlations (‘*Avg of Abs*’). The number of significant pairwise correlations is reported in parentheses.

	S&P 500	*NASDAQ 100*	*EUROSTOXX 50*	*FTSE 100*	*DAX*
Panel A. Daily aggregate flows					
N of ETFs	34	13	10	5	7
Avg of R^2	0.60%	0.50%	0.40%	0.20%	1.20%
Median of R^2	0.50%	0.10%	0.40%	0.20%	0.40%
Std of R^2	0.60%	1.20%	0.30%	0.20%	1.50%
N of Pairs	546	78	45	10	21
N of (+)	299 (56)	50 (10)	35 (9)	6 (4)	14 (6)
N of (-)	247 (31)	28 (2)	10 (1)	4 (0)	7 (0)
Avg of Abs	0.036	0.047	0.028	0.035	0.044
Panel B. Weekly aggregate flows					
N of ETFs	34	13	10	5	7
Avg of R^2	2.80%	1.60%	3.30%	1.60%	4.60%
Median of R^2	1.90%	0.60%	2.90%	1.40%	2.30%
Std of R^2	3.30%	2.10%	2.40%	0.60%	4.60%
N of Pairs	546	78	45	10	21
N of (+)	327 (61)	46 (9)	32 (12)	10 (2)	17 (7)
N of (-)	219 (21)	32 (3)	13 (1)	0 (0)	4 (0)
Avg of Abs	0.075	0.093	0.067	0.064	0.104
Panel C. Monthly aggregate flows					
N of ETFs	34	13	10	5	7
Avg of R^2	8.90%	8.70%	9.40%	10.70%	12.90%
Median of R^2	5.40%	4.60%	9.70%	10.00%	12.10%
Std of R^2	12.70%	13.90%	4.50%	5.50%	9.60%
N of Pairs	546	78	45	10	21
N of (+)	335 (58)	52 (5)	33 (10)	10 (4)	15 (6)
N of (-)	211 (31)	26 (3)	12 (1)	0 (0)	6 (1)
Avg of Abs	0.160	0.185	0.158	0.174	0.185

Table 6. Flow commonality of fundamental flows

This table reports the flow commonality estimated by R-squared from a regression model and pairwise correlation of fundamental flows in ETFs tracking the same index. For each ETF, we estimate the R-squared using the following regression model:

$$FF_{i,t} = c_0 + c_1 FF_{m,t-1} + c_2 FF_{m,t} + c_3 FF_{m,t+1} + \varepsilon_t,$$

where $FF_{i,t}$ represents fundamental flow in ETF i at time t, $FF_{m,t}$ is market ETF flow at time t. For each underlying index, when calculating the market ETF flow for ETF i, we use the equal-weighted average of the ETF flows, excluding ETF i. Panels A, B, and C report the results of daily, weekly, and monthly fundamental flow, respectively. For each underlying index (S&P 500 index, NASDAQ 100 index, EUROSTOXX 50 index, FTSE 100 index, and DAX index), each row presents the number of ETFs ('*N of ETFs*'), the average of R-squared ('*Avg of R^2*'), the median of R-squared ('*Median of R^2*'), the standard deviation of R-squared ('*Std of R^2*'), the number of pairs ('*N of Pairs*'), the number of positively correlated pairs ('*N of (+)*'), the number of negatively correlated pairs ('*N of (-)*'), and the average of absolute pairwise correlations ('*Avg of Abs*'). The number of significant pairwise correlations is reported in parentheses.

	S&P 500	*NASDAQ 100*	*EUROSTOXX 50*	*FTSE 100*	*DAX*
Panel A. Daily fundamental flows					
N of ETFs	34	13	10	5	7
Avg of R^2	2.30%	14.60%	4.90%	2.20%	9.40%
Median of R^2	0.80%	7.50%	3.20%	1.80%	7.90%
Std of R^2	3.40%	23.00%	5.90%	2.00%	6.30%
N of Pairs	546	78	45	10	21
N of (+)	287 (217)	39 (20)	21 (19)	5 (4)	11 (9)
N of (-)	259 (170)	39 (20)	24 (22)	5 (5)	10 (8)
Avg of Abs	0.195	0.211	0.285	0.302	0.258
Panel B. Weekly fundamental flows					
N of ETFs	34	13	10	5	7
Avg of R^2	3.10%	14.90%	7.10%	5.60%	14.80%
Median of R^2	2.30%	3.90%	5.10%	5.10%	10.70%
Std of R^2	3.40%	25.30%	8.10%	4.00%	9.80%
N of Pairs	546	78	45	10	21
N of (+)	296 (221)	38 (31)	21 (16)	4 (4)	12 (8)
N of (-)	250 (170)	40 (26)	24 (20)	6 (6)	9 (7)
Avg of Abs	0.288	0.362	0.275	0.314	0.298
Panel C. Monthly fundamental flows					
N of ETFs	34	13	10	5	7
Avg of R^2	23.80%	16.70%	10.00%	13.20%	22.80%
Median of R^2	14.80%	6.50%	8.20%	16.80%	25.50%
Std of R^2	26.80%	24.80%	8.20%	9.80%	19.70%
N of Pairs	546	78	45	10	21
N of (+)	308 (161)	40 (23)	21 (9)	4 (3)	13 (10)
N of (-)	238 (94)	38 (21)	24 (15)	6 (3)	8 (5)
Avg of Abs	0.310	0.404	0.249	0.350	0.351

Table 7. Flow commonality of non-fundamental flows

This table reports the flow commonality estimated by R-squared from a regression model and pairwise correlation of non-fundamental flows in ETFs tracking the same index. For each ETF, we estimate the R-squared using the following regression model:

$$FF_{i,t} = c_0 + c_1 FF_{m,t-1} + c_2 FF_{m,t} + c_3 FF_{m,t+1} + \varepsilon_t,$$

where $FF_{i,t}$ represents non-fundamental flow in ETF i at time t, $FF_{m,t}$ is market ETF flow at time t. For each underlying index, when calculating the market ETF flow for ETF i, we use the equal-weighted average of the ETF flows, excluding ETF i. Panels A, B, and C report the results of daily, weekly, and monthly non-fundamental flow, respectively. For each underlying index (S&P 500 index, NASDAQ 100 index, EUROSTOXX 50 index, FTSE 100 index, and DAX index), each row presents the number of ETFs ('*N of ETFs*'), the average of R-squared ('*Avg of R^2*'), the median of R-squared ('*Median of R^2*'), the standard deviation of R-squared ('*Std of R^2*'), the number of pairs ('*N of Pairs*'), the number of positively correlated pairs ('*N of (+)*'), the number of negatively correlated pairs ('*N of (-)*'), and the average of absolute pairwise correlations ('*Avg of Abs*'). The number of significant pairwise correlations is reported in parentheses.

	S&P 500	*NASDAQ 100*	*EUROSTOXX 50*	*FTSE 100*	*DAX*
Panel A. Daily non-fundamental flows					
N of ETFs	34	13	10	5	7
Avg of R^2	0.60%	0.50%	0.30%	0.30%	1.20%
Median of R^2	0.30%	0.10%	0.30%	0.30%	0.40%
Std of R^2	0.60%	1.20%	0.20%	0.20%	1.50%
N of Pairs	546	78	45	10	21
N of (+)	301 (63)	48 (12)	33 (8)	5 (4)	16 (6)
N of (-)	245 (30)	30 (2)	12 (1)	5 (1)	5 (0)
Avg of Abs	0.037	0.045	0.028	0.042	0.044
Panel B. Weekly non-fundamental flows					
N of ETFs	34	13	10	5	7
Avg of R^2	3.10%	1.90%	2.80%	2.70%	4.60%
Median of R^2	1.80%	0.90%	2.00%	1.60%	1.80%
Std of R^2	3.30%	2.20%	2.50%	2.30%	4.80%
N of Pairs	546	78	45	10	21
N of (+)	347 (79)	46 (15)	31 (12)	6 (2)	17 (6)
N of (-)	199 (14)	32 (3)	14 (1)	4 (1)	4 (0)
Avg of Abs	0.080	0.107	0.064	0.081	0.110
Panel C. Monthly non-fundamental flows					
N of ETFs	34	13	10	5	7
Avg of R^2	11.50%	11.20%	7.70%	9.30%	12.40%
Median of R^2	6.40%	7.50%	7.00%	13.10%	8.90%
Std of R^2	13.10%	12.60%	4.30%	5.90%	10.00%
N of Pairs	546	78	45	10	21
N of (+)	354 (99)	49 (10)	31 (8)	7 (2)	12 (6)
N of (-)	192 (21)	29 (1)	14 (1)	3 (1)	9 (0)
Avg of Abs	0.180	0.196	0.152	0.168	0.184