Review of Finance, 2021, 937–972
doi: 10.1093/rof/rfaa027
Advance Access Publication Date: 7 October 2020

ETF Arbitrage, Non-Fundamental Demand, and Return Predictability*

David C. Brown,[1] Shaun William Davies[2], and Matthew C. Ringgenberg[3]

[1]University of Arizona, [2]University of Colorado Boulder, and [3]University of Utah

Abstract

Non-fundamental demand shocks have significant effects on asset prices, but observing these shocks is challenging. We use the exchange-traded fund (ETF) primary market to study non-fundamental demand. Unique to the ETF market, specialized arbitrageurs called authorized participants correct violations of the law of one price between an ETF and its underlying assets by creating or redeeming ETF shares. We show theoretically and empirically that creation and redemption activities (ETF flows) provide signals of non-fundamental demand shocks. A portfolio that is short high-flow ETFs and long low-flow ETFs earns excess returns of 1.1–2.0% per month, consistent with non-fundamental demand distorting asset prices away from fundamental values. Moreover, we show non-fundamental demand imposes non-trivial costs on investors, leading to underperformance.

Keywords: Exchange-traded funds (ETFs), ETF flows, Non-fundamental demand, return predictability

JEL classification: G12, G14

Received November 23, 2019; accepted September 22, 2020 by Editor Jules van Binsbergen.

* We are grateful for the helpful comments of Jules van Binsbergen (the Editor), and Itzhak Ben-David, Azi Ben-Rephael, Hank Bessembinder, Jesse Blocher, Scott Cederburg, Don Chance, Alex Chinco, Zhi Da, Caitlin Dannhauser, Joey Engelberg, Nicholas Hirschey, Gerard Hoberg, Doron Israeli, Rabih Moussawi, Stefan Nagel, Antti Petajisto, Sophie Shive, Rick Sias, Harry Turtle, Ed Van Wesep, several anonymous reviewers, and conference and seminar participants at Research Affiliates, Cubist Systematic Strategies, the 2016 Brevan Howard Centre for Financial Analysis' Finance Theory Group Conference, HKU, HKUST, University of Colorado, the European Winter Finance Conference 2017, the ASU Sonoran Winter Finance Conference 2017, the 2017 FSU SunTrust Beach Conference, the 2017 Front Range Finance Seminar, and the 2018 American Finance Association Annual Meeting. We also thank the Investment Company Institute for providing mutual fund flow data.

1. Introduction

Non-fundamental demand shocks cause asset prices to deviate from their fundamental values. In turn, this can lead to financial market externalities that cause distortions in real investment and output.[1] Despite the importance of non-fundamental demand shocks for financial markets and the real economy, measuring them remains a challenge. Identifying non-fundamental demand shocks from stock price changes is difficult because fundamentals values are unobservable. Similarly, identifying non-fundamental demand from trading volume is difficult because volume does not reveal the underlying motivations for trading. Even mutual fund flows are confounded by information about fund manager skill, making it difficult to disentangle non-fundamental demand from demand for specific fund managers.[2]

In this article, we show theoretically and empirically that exchange-traded funds (ETFs) offer a unique setting to identify and study non-fundamental demand. We develop a model of the ETF market to demonstrate that ETF flows are symptomatic of shocks that dislocate prices from fundamental values. The intuition is simple: while fundamental value is inherently unobservable, ETFs and their underlying assets share the same fundamental value. Thus, violations of the law of one price between the two signal that at least one of them was affected by non-fundamental demand (e.g., sentiment, noise, etc.). Unique to the ETF market, specialized arbitrageurs called authorized participants correct violations of the law of one price by creating or redeeming ETF shares. Because these creations or redemptions by authorized participants (i.e., ETF flows) are publicly observable, they provide a novel signal of the occurrence of non-fundamental demand.[3]

We confirm the model's predictions and show empirically that ETFs provide valuable conditioning information to identify fundamental mispricing. That is, ETF flows predict future asset returns in both an ETF's shares and its underlying assets, consistent with non-fundamental demand generating mispricings that subsequently reverse. Given that most ETFs are passively managed, are traded by a rich cross-section of investors, and encompass nearly every asset class, our findings show ETF flows provide arguably the cleanest and most extensive measure of non-fundamental demand shocks to date. Moreover, our analysis allows us to explore the implications of non-fundamental demand; we find it imposes economically significant costs on investors. For example, in State Street's S&P 500 ETF (SPY), the largest ETF in the world, investors bear additional indirect costs that are an order of magnitude larger than the ETF's explicit management fee.

1 For example, several papers on mutual fund fire sales argue that non-fundamental demand shocks cause fund managers to liquidate holdings. This causes other assets to decline in value, thereby decreasing the value of firms and causing reductions in aggregate investment and output (Lorenzoni, 2008; Shleifer and Vishny, 2011).

2 Mutual fund flows have been shown to contain information about investor demands and/or sentiment (Ippolito, 1992; Sirri and Tufano, 1998; Cooper, Gulen, and Rau, 2005; Lou, 2012; Kamstra *et al.*, 2017). Theory suggests that mutual fund flows are also contaminated by information about fund manager skill (Berk and Green, 2004). Warther (1995) and Ben-Rephael, Kandel, and Wohl (2012) examine flows aggregated across mutual funds to measure aggregate investor sentiment allowing them to avoid the confounding influence of fund manager skill.

3 Importantly, ETF flows are primary market trades that are distinct from the secondary market trades of ETF investors. While ETF investors may trade for fundamental reasons, we focus only on primary market trades that occur because of violations of the law of one price.

We start by developing a parsimonious model of the ETF market based on its prominent features: (i) both ETF shares and ETF underlying assets are traded in secondary markets, (ii) unlike open-ended mutual funds, the prices of ETF shares may differ from the prices of the underlying assets, (iii) unlike closed-end funds, authorized participants restore relative price efficiency via primary market activities, and (iv) unlike most open- and closed-end funds that are actively managed, ETFs are passively managed.[4] Authorized participants create new ETF shares by delivering the underlying assets if the ETF shares are trading at a premium or, conversely, authorized participants redeem ETF shares for the underlying assets if the ETF shares are trading at a discount. We solve for authorized participants' equilibrium arbitrage activity (i.e., ETF flows), the equilibrium ETF price, and the equilibrium underlying assets' prices. We show that ETF flows are signals of unobservable non-fundamental demand shocks. Put differently, ETF flows can identify unobservable fundamental mispricing.

Figure 1 illustrates the mechanism. At $t = 0$, both the ETF and the underlying assets share the same price. At $t = 1$, latent non-fundamental demand shocks arrive and push the price of the ETF above the underlying net asset value (NAV). These demand shocks could differentially affect the ETF and the underlying assets, either because the demand shock that hits the ETF is larger than the shock to the underlying assets or because the ETF is relatively more sensitive to demand shocks. This leads to an ETF premium (i.e., a relative mispricing). At $t = 2$, authorized participants buy shares in the underlying assets and sell shares in the ETF to correct the relative mispricing. They close their trades by creating new shares in the ETF, generating observable ETF flows that reveal the non-fundamental demand shocks. Importantly, while arbitrageurs trade to correct relative mispricing, their trades do not correct fundamental mispricing. Thus, in the long term, the prices of both the ETF and the underlying assets exhibit return predictability as the fundamental mispricing corrects.

While the model succinctly shows that ETF flows signal non-fundamental demand, the result is not driven by the setup or assumptions. Trading by authorized participants only occurs if there is a profitable arbitrage trade, and this must indicate that there was excess demand for either the ETF shares or the underlying assets. Since the ETF shares are a claim to the underlying assets, excess demand for either the shares or the underlying assets must be due to a non-fundamental force. Of course, it is possible that the non-fundamental force is correlated with fundamental demand, for example, over- or under-reaction to news about future cash flows or discount rates. In this case, what matters is the size of the fundamental shock relative to the non-fundamental component. To use an analogy, ETF flows are akin to observing ice in the ocean; while they are observable, the extent to which they signal large or small fundamental mispricing beneath the surface is unclear. On the one hand, the non-fundamental component signaled by ETF flows may be (i) highly correlated with a fundamental demand shock and (ii) relatively small as compared to the fundamental demand shock. In such a case, return reversals will likely be small, short-lived, and hard to

4 Closed-end funds do not have authorized participants who enforce the law of one price between the fund shares and the underlying assets. As a result, the funds can trade at a premium or, more often, at a significant discount (relative to the underlying). Previous studies have examined these large and persistent mispricings and offer a wide range of explanations: managerial skill and fees (Berk and Stanton, 2007), liquidity (Cherkes, Sagi, and Stanton, 2008), costly arbitrage capital (Pontiff, 1996), and investor sentiment (Lee, Shleifer, and Thaler, 1991; Baker and Wurgler, 2006).

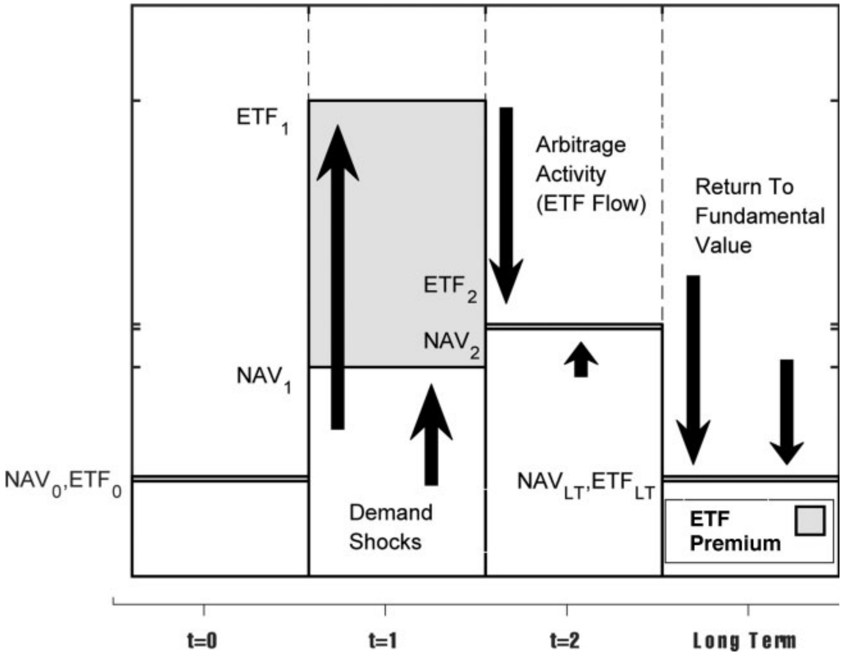


Figure 1. ETF flows and ETF premia changes as signals of non-fundamental demand shocks. At $t = 0$, no mispricing exists between the ETF share price (ETF_0) and the ETF NAV (NAV_0) and both are priced at their shared fundamental value. At $t = 1$, imbalanced non-fundamental demand shocks generate a mispricing by pushing the ETF share price and the ETF NAV away from their shared value. At $t = 2$, arbitrageurs restore relative price efficiency, putting upward price pressure on the ETF NAV and downward price pressure on the ETF share price. In the long term, the ETF and NAV prices revert back to their shared fundamental value.

detect.[5] That is, there is not much beneath the surface. On the other hand, the non-fundamental component signaled by ETF flows may be (i) not strongly related to fundamentals and (ii) economically significant. In this case, return predictability is likely to be large. That is, ETF flows are the tip of the iceberg. Ultimately, it is an empirical question of whether ETF flows signal large price distortions. We find that they do.

We test the model's predictions using a sample of ETFs over the period 2007–16. We find strong evidence that ETF flows predict future asset returns, consistent with ETF flows' identifying non-fundamental demand. We start our empirical analysis with univariate portfolio sorts examining ETF flows. Each month, we form long-short portfolios based on the prior month's ETF flows; our long-short portfolio buys ETFs with the smallest flows and sells short ETFs with the largest flows. Consistent with our model, we find strong evidence that ETF flows contain information about future asset returns. At a 1-month horizon, our long-short portfolio earns returns of 1.1–2.0% per month. Moreover, return predictability

5 For example, Marshall, Nguyen, and Visaltanachoti (2013), Madhavan and Sobczyk (2016), and Ben-David, Franzoni, and Moussawi (2018) provide evidence consistent with ETF flows' speeding price discovery, particularly when the ETF is more liquid than the underlying assets. If the non-fundamental shocks signaled by ETF flows are highly correlated with fundamental price discovery, then measurable return predictability is unlikely.

remains at the 3- and 6-month horizons indicating that ETF flows signal non-fundamental demand shocks that lead to fundamental mispricing that corrects over time. Because univariate sorts do not control for fund- and time-level heterogeneity, we next turn to a multivariate regression approach. We again find that ETF flows contain information about future asset returns in the cross-section. ETFs in the top decile of flows underperform ETFs in the bottom decile by a statistically significant 1.8% over the next month. This result holds at 3- and 6-month horizons and remains statistically and economically significant when we examine risk-adjusted returns. Consistent with our model's predictions, we also find that return predictability is strongest in leveraged ETFs and high activity ETFs (defined as those with the most active primary markets), which are characterized by volatile flows. Moreover, the results are not driven by extreme returns in a few sample months. Trading strategies produce annualized Sharpe ratios of 0.60 and 0.99 at the 1- and 6-month horizons.

Our results document a strong negative relation between ETF flows and subsequent returns. This suggests that non-fundamental demand imposes non-trivial costs on investors. To examine the implications of our results for ETF investor profitability, we consider two performance measures: the internal rate of return (IRR) and the share-growth-adjusted return (SGAR). The IRR measures the returns of a representative investor that holds all shares in an ETF and is responsible for all of the ETF's inflows and outflows. The SGAR has an asset allocation interpretation: it considers an investor who rebalances monthly across a risk-free asset and the ETF based on the last month's flows. Both measures speak to the relation between flows and returns, but they differ in important ways. The IRR is agnostic to timing, accounting for the relation between flows and returns at all horizons. This flexibility comes at a cost; the measure cannot distinguish between whether flows lead returns or returns lead flows (Hayley, 2014). The SGAR is intentionally less flexible; by construction, it only examines the relation between last month's flows and the current month's returns. Both measures provide the same conclusion: non-fundamental demand distorts prices and imposes non-trivial costs on ETF investors, leading to underperformance.

We contribute to the literature on the information in investor flows. First, our model is simple and it provides clear intuition for the ETF market. In the same spirit that Berk and Green (2004) show that mutual fund flows reflect demand for specific managers, our model shows that ETF flows signal non-fundamental demand. As such, we provide a simple workhorse model that may be built upon and utilized in future ETF studies. Second, as noted in Daniel, Hirshleifer, and Subrahmanyam (2001), " … expected returns also depend on current mispricing, so returns can be predicted better by conditioning on proxies for misvaluation." This article does this—we provide a novel measure of misvaluation. Many papers find that primary and secondary market flows contain information about investor demands. For example, several papers find that flows into, and out of, mutual funds and hedge funds contain information about future asset returns (Ippolito, 1992; Sirri and Tufano, 1998; Cooper, Gulen, and Rau, 2005; Frazzini and Lamont, 2008; Ben-David, Franzoni, and Moussawi, 2012; Ben-Rephael, Kandel, and Wohl, 2012; Lou, 2012).[6] Our results are related to, but distinct from, these existing findings for three reasons. First, our

6 There is also a large literature that documents a relation between primary market issuances (and related corporate events) and future stocks returns (Ikenberry, Lakonishok, and Vermaelen, 1995; Loughran and Ritter, 1995; Loughran and Vijh, 1997; Baker and Wurgler, 2000; Daniel and Titman, 2006; Fama and French, 2008; Pontiff and Woodgate, 2008).

model shows that authorized participant trading indicates the occurrence of non-fundamental demand. This is in contrast to other measures of investor flows, like mutual fund flows, which could reflect learning about fund manager skill and thus do not necessarily reflect non-fundamental demand. Second, the signal of non-fundamental demand does not originate from one class of investors; ETFs are used by a broad cross-section of market participants that includes retail traders, large institutions, and hedge funds. Third, we empirically demonstrate that the information in ETF flows is distinct from the information in mutual fund flows: our results are robust to controlling for aggregate mutual fund flows (Ben-Rephael, Kandel, and Wohl, 2012) and we do not find return predictability in passively managed index mutual fund flows.[7] Thus, while a number of papers have used mutual fund flows to measure deviations from fundamental values in asset prices, this article shows that ETF flows provide a powerful and clean measure for use in future studies.

We also provide novel evidence that ETF investors systematically underperform due to non-fundamental demand and display poor timing. Several papers show that investors in other asset classes underperform because of timing effects.[8] In our setting, ETF investors are simply moving into or out of a passively managed basket of securities. As these flows originate from a diverse group of investors, it is perhaps surprising that these ETF investors collectively mistime their flows. However, these results are reconcilable with our model; ETF flows signal non-fundamental demand shocks that predictably reverse, regardless of the source. Thus, ETF flows aggregate non-fundamental demand shocks from different types of investors.

This article adds to the growing literature on the relation between ETFs and other market outcomes.[9] As non-fundamental demand shocks can cause both relative and absolute mispricing, our findings complement the results in Ben-David, Franzoni, and Moussawi (2018). Ben-David, Franzoni, and Moussawi (2018) shows that arbitrageur trades, which are acting to restore relative price efficiency, transmit volatility to the underlying assets via price pressure. They also show that arbitrage activity itself is related to underlying stock prices, which generates return predictability. However, they examine price pressure from arbitrage activity which subsequently reverses at short horizons. In contrast, we show that the need to restore relative price efficiency in the first place signals there was a non-fundamental demand shock that generated an absolute mispricing. Empirically, these mispricings are distinct from those studied in Ben-David, Franzoni, and Moussawi (2018) and subsequently reverse over horizons of 1–6 months.[10] Overall, our model provides a

7 See Tables IA7, IA14, IA22, and IA23 of the Online Appendix.

8 Sapp and Tiwari (2004), Friesen and Sapp (2007), and Hsu, Myers, and Whitby (2016) show that mutual fund investors chase past returns, leading to negative market-timing and Dichev (2007) finds that investors in equities systematically underperform once distributions and contributions are accounted for.

9 A number of papers study the direct effects of ETF flows on assets. Baltussen, van Bekkum, and Da (2019) and Da and Shive (2018) show that ETFs induce comovement between underlying assets, and Ben-David, Franzoni, and Moussawi (2018) and Krause, Ehsani, and Lien (2014) document volatility transmission from ETFs to the funds' underlying assets. For more related work, see Bessembinder (2015), Israeli, Lee, and Sridharan (2017), Jiang and Yan (2020), Staer (2016), Agarwal *et al.* (2019), Dannhauser (2017), Dannhauser and Hoseinzade (2017), Glosten, Nallareddy, and Zou (2020), Pan and Zeng (2019), Staer and Sottile (2018), and Dannhauser and Pontiff (2019).

10 See Section 8 of the Online Appendix for a detailed discussion.

microfoundation for using ETF flows as signals of non-fundamental demand shocks and our empirical findings confirm the model's predictions.

2. Model of ETF Trade

In this section, we provide a succinct model which shows that ETF flows must imply a non-fundamental distortion occurred and that ETF flows can signal a fundamental mispricing. Importantly, while the model is built from several assumptions which allow for richer empirical predictions, our insights are not driven by these assumptions. To show this, we begin with a simple thought experiment before developing the model. Consider two traded securities that share the same fundamental value with prices A and B, respectively. The law of one price says that A should equal B. If $A \neq B$, there is a relative mispricing, and $A - B$ reflects the excess demand for A. By taking the difference between A and B, their shared fundamental value cancels out and the residual must reflect mispricing from non-fundamental excess demand.[11] Thus, trades that exploit the mispricing (i.e., arbitrage trades) imply the realization of this non-fundamental excess demand. As the model will show, ETF flows always signal a non-fundamental force.

Consider a four-period model $t \in \{0, 1, 2, longterm\}$in which a passively managed ETF provides exposure to a benchmark index (e.g., the S&P 500). In period $t = 0$, initial prices for the ETF shares and the ETF underlying assets are set. At $t = 1$, both fundamental news and non-fundamental distortions are realized, potentially giving rise to a relative mispricing.[12] At $t = 2$, arbitrageurs called authorized participants to trade against the mispricing and equilibrium prices for the shares and the underlying assets are determined. The equilibrium prices are not necessarily equal to fundamental value. At $t = longterm$, prices return to fundamental value. We elaborate below.

In each period t, there is a q_t-length measure of ETF shares and the market value of each share is denoted p_t. The underlying assets backing the ETF shares mirror the benchmark index. Each unit of the ETF's underlying assets has an unobservable fundamental value Ω_t and a tradable value π_t (i.e., NAV). For simplicity, the number of units of the underlying asset held by the ETF is also q_t so that the NAV per ETF share is π_t. In each period, the ETF premium is the difference in p_t and π_t,

$$\psi_t \equiv p_t - \pi_t. \tag{1}$$

A non-zero value of ψ_t represents a violation of the law of one price. A non-zero value of ψ_t is a true-arbitrage opportunity as there is a long-short trade to exploit the mispricing.

Define the ETF fundamental mispricing as the difference between the ETF share price and its unobservable fundamental value,

$$\varphi_t = p_t - \Omega_t, \tag{2}$$

and define the NAV fundamental mispricing as the difference between the NAV per share and its unobservable fundamental value,

11 Note that mispricing between A and B could be correlated with fundamental demand. In Section 7 of the Online Appendix, we provide additional discussion and we address several misconceptions regarding the ETF market mechanism.

12 Our use of the term "non-fundamental" includes beliefs that are uncorrelated with fundamental news and also over- and under-reaction to fundamental news.

$$\alpha_t = \pi_t - \Omega_t. \tag{3}$$

Non-zero values of φ_t and α_t reflect fundamental mispricing, sometimes referred to as a risky-arbitrage opportunity. However, the mispricing is not observable because Ω_t is latent. Even if one has a signal regarding the values of φ_t and α_t, there is not a risk-free trading strategy that can capture the mispricing.

In addition to the ETF shares and the underlying assets trading in a secondary market, there also exists a primary market for the ETF shares. The primary market is constituted by $N \geq 1$authorized participants and the ETF sponsor (e.g., BlackRock). In response to a relative mispricing ($\psi_t \neq 0$), each of the N authorized participants may deliver the underlying assets in exchange for new ETF shares or may deliver existing ETF shares in exchange for the underlying assets.[13] We denote each authorized participant i's arbitrage demand (share creations or share redemptions) at time t as $\delta_{t,i}$ and the aggregate demand of all authorized participants at time t as $\Delta_t = \sum_{i=1}^{N} \delta_{t,i}$. Δ_t may be positive in value (ETF shares are created in net) or Δ_t may be negative in value (ETF shares are redeemed in net). Thus, Δ_t measures investor flows in and out of the underlying assets through the ETF; we refer to Δ_t as the ETF flow hereafter. Because the ETF flow only occurs at $t = 1$, we drop the subscript t.

The market price for ETF shares is determined by the intersection of ETF share supply and aggregate ETF investor demand. Investors' collective demand is downward-sloped,

$$p_t = \Omega_t + \beta - \eta q_t + \epsilon_t^{etf}, \tag{4}$$

in which Ω_t is the latent fundamental value, $\beta \geq 0$ is a constant so that the initial shares outstanding are strictly positive and $\eta \geq 0$ proxies for investors' sensitivity to the measure of shares. A downward-sloped demand curve is micro-founded on investor risk aversion and lower values of η imply less price impact from ETF flows. The variable ϵ_t^{etf} is a non-fundamental component to ETF investor demand drawn from a mean zero distribution with variance σ_e^2. We are agnostic to the source of the non-fundamental demand component; it could be due to noise/liquidity trading (Black, 1986; De Long et al., 1990) or investor sentiment (Lee, Shleifer, and Thaler, 1991; Baker and Wurgler, 2006, 2007; Frazzini and Lamont, 2008). Or, the non-fundamental demand component could be due to characteristics of trading in the ETF that differ from trading in the underlying assets. For example, trading costs often render the ETF cheaper than buying the underlying assets individually. Alternatively, the non-fundamental demand component could be over- or under-reaction to fundamental news. Importantly, the non-fundamental component is any tension that distorts the ETF price away from fundamentals. From here forward, we refer to the non-fundamental component as a non-fundamental demand shock without loss of generality.

The NAV is given by,

$$\pi_t = \Omega_t + \epsilon_t^{nav} + \lambda\Delta, \tag{5}$$

in which Ω_t is the latent fundamental value and ϵ_t^{nav} is an aggregate non-fundamental shock to the underlying asset demand drawn from a mean zero distribution with variance σ_n^2. Similar to ϵ_t^{etf}, ϵ_t^{nav} is any non-fundamental tension that distorts the NAV. Additionally, the trading activity of authorized participants may create price pressure on the underlying asset via $\lambda \geq 0$. For example, if the price of the underlying assets are determined by a market

13 Without loss of generality, for derivatives based ETFs (e.g., leveraged ETFs), the authorized participants may deliver cash in the amount of NAV for an ETF share, and vice versa.

maker that does not observe the ETF flow, she may interpret trades as signals of the assets' fundamental value as in Kyle (1985). Alternatively, λ may be determined by investor risk aversion as a positive value of λ implies that investor demand for the underlying assets is also downward-sloped.

The ETF shares' and ETF underlying assets' shared latent fundamental value Ω_t is given by,

$$\Omega_t = \Omega_{t-1} + \omega_t, \tag{6}$$

in which ω_t is fundamental news (i.e., a fundamental shock) distributed according to a zero-mean distribution.

Importantly, we allow non-fundamental demand shocks to hit both the ETF shares and the underlying assets. For example, our model can handle the possibility that the ETF shares attract sentiment trades and also that the underlying assets are illiquid. In this example, there are non-fundamental forces at work in both the ETF shares and the underlying assets; the ETF share prices may be pushed around by sentiment trades while the underlying asset prices are stale. As we show shortly, equilibrium ETF flows are linear in the difference of these two non-fundamental demand shocks. In other words, what is important is the net non-fundamental demand between the ETF shares and the underlying assets. We denote the correlation between ϵ_t^{etf} and ϵ_t^{nav} as $\rho = \frac{\mathrm{Cov}(\epsilon^{etf}, \epsilon^{nav})}{\sigma_e \sigma_n}$. Furthermore, $\epsilon_0^{etf} = 0$ and $\epsilon_0^{nav} = 0$ and, as such, we drop the subscript t hereafter. This implies that $\pi_0 = \Omega_0$. The initial measure of the ETF's shares q_0 is set so that an ETF premium does not initially exist and the ETF is fairly priced with respect to fundamental value. $q_0 \equiv \frac{\beta}{\eta}$ solves,

$$\pi_0 = \beta + \Omega_0 - \eta q_0. \tag{7}$$

Within period $t = 1$: (i) demand shocks ϵ^{etf}, ϵ^{nav}, and ω_t are realized, and (ii) investor demands shift for the ETF shares and the underlying assets, giving rise to an interim ETF premium ψ_1. At $t = 2$: (i) authorized participants buy the less expensive asset and sell the more expensive asset, generating price pressure on both the ETF shares and the underlying assets, (ii) authorized participants create or redeem shares to complete the arbitrage trade, and (iii) the equilibrium prices for the ETF shares and the ETF underlying assets are established. In practice, the time between $t = 1$ and $t = 2$ may take place over an instant as arbitrageurs quickly restore relative price efficiency.

To provide the most natural and intuitive explanation of our model, we focus on the special case in which $\rho = 1$, $\sigma_e \neq \sigma_n$, and in the limit $N \to \infty$. Assuming $\rho = 1$ implies that the non-fundamental demand shocks to the ETF shares and to the underlying assets are perfectly correlated, and assuming $\sigma_e \neq \sigma_n$ implies that a non-fundamental demand shock gives rise to a relative mispricing. Considering the limiting case in which $N \to \infty$ focuses on perfect competition among authorized participants. A general solution to the model is located in the Appendix.

Remark 1. The ETF primary market mechanism's ability to restore relative price efficiency requires, $\eta > 0$, $\lambda > 0$, or both.

ETF design, both in our model and in practice, is predicated on the requirement that relative price efficiency is restored by affecting the supply of ETF shares. Remark 1 highlights that affecting the supply of ETF shares has a price effect if, and only if, either the demand for the ETF shares is downward sloping, the demand for the underlying assets is downward sloping, or both.

After the demand shocks are realized, each authorized participant i chooses a length $\delta_i \in \Re$ of shares to create or redeem in the ETF to exploit the true-arbitrage opportunity. Each authorized participant solves the following optimization,

$$\max_{\delta \in \Re} \delta_i(p_2(\delta_i|\delta_{-i}) - \pi_2(\delta_i|\delta_{-i})), \tag{8}$$

in which δ_{-i} is the creation/redemption activity of the other authorized participants.

Lemma 1. *The aggregate ETF flow is,*

$$\lim_{N\to\infty} \Delta^* = \frac{\epsilon^{etf} - \epsilon^{nav}}{\lambda + \eta}. \tag{9}$$

According to Lemma 1, the equilibrium ETF flow does not contain the fundamental shock ω_1 and is linear in the difference of the two non-fundamental demand shocks.[14] As such, there is a natural economic interpretation of the ETF flow outlined in Equation (9): ETF flows occur when there is net excess demand in either the ETF shares or the ETF underlying assets. When the net excess demand favors the ETF shares, there is a positive flow (i.e., share creations). When the net excess demand favors the ETF underlying assets, there is a negative flow (i.e, share redemptions).

Remark 2. ETF flows are symptomatic of non-fundamental demand distortions.

Equation (9) and Remark 2 motivate our empirical analyses by showing that ETF flows are sufficient to identify that at least one non-fundamental demand shock occurred (either in the ETF share demand, the ETF underlying asset demand, or in both). A non-fundamental shock could be attributed to many sources: under- or over-reaction to fundamental news, market frictions, liquidity shocks, and investor sentiment shocks.[15] Notably, while the demand for the ETF shares and the demand for the underlying assets both contain the fundamental component, the fundamental component does not directly show up in the relative mispricing.[16] As a consequence, ETF flows provide a clean signal that a non-fundamental demand shock occurred.

Lemma 2. *The variance of the ETF flow is,*

$$\lim_{N\to\infty} \mathrm{Var}(\Delta^*) = \frac{\sigma_e^2 + \sigma_n^2 - 2\rho\sigma_e\sigma_n}{(\lambda+\eta)^2}, \tag{10}$$

and $\mathrm{Var}(\Delta^*)$ *is increasing in* σ_e^2 *and* σ_n^2 *and decreasing in* $\rho\sigma_e\sigma_n$.

Remark 3. ETFs characterized by larger exposure to non-fundamental demand shocks (large values of σ_e^2, σ_n^2, or both) and larger differences in sensitivities to non-fundamental demand shocks (large differences in σ_e and σ_n) have more volatile flows.

14 Proofs of Lemma 1, Lemma 2, and Proposition 1 are in the Appendix.

15 See Hirshleifer (2001) and Barberis and Thaler (2003) for surveys of the behavioral finance literature.

16 While the fundamental component does not directly affect ETF flows, the non-fundamental component can be correlated with the fundamental component (e.g., under- or over-reaction). See Section 7 of the Online Appendix for additional discussion.

Remark 3 implies that ETFs with volatile flows are most symptomatic of non-fundamental demand shocks, particularly if the investor clienteles that trade the ETF shares and the ETF underlying assets are sufficiently different (i.e., large differences in σ_e and σ_n).

Having outlined the model and having demonstrated the mechanism, it is worthwhile to place our model in the fund flows literature. The most natural benchmark model is that of Berk and Green (2004).[17] Berk and Green (2004) provides a rational explanation for the "fund-flow anomaly."[18] Specifically, in the model, fund managers have the unobservable skills to earn abnormal returns, funds experience diminishing returns to fund size, and investors with perfectly elastic capital learn about managerial skill via realized performance. As such, after performance is revealed, investors update their beliefs about each fund manager's skill and reallocate capital such that any expected abnormal returns are competed away. Thus, in Berk and Green (2004), mutual fund flows reflect the evolution of investors' beliefs regarding managerial skill.

Our model is distinct from that of Berk and Green (2004) for one primary reason: ETFs are passively managed. As such, realized performance is unrelated to managerial skill and cannot reflect the evolution of rational beliefs. However, ETF flows in our model do share one important similarity with the mutual fund flows in Berk and Green (2004); they both reflect competition among rational agents. In Berk and Green (2004), mutual fund flows reflect rational investors competing away expected abnormal returns arising from managerial skill. In our model, ETF flows reflect rational authorized participants competing away relative mispricing arising from non-fundamental demand.

Using the equilibrium ETF flow Δ^*, we solve for the equilibrium $t=2$ prices and the corresponding equilibrium ETF premium ψ_2, ETF fundamental mispricing φ_2, and the NAV fundamental mispricing α_2. Hereafter, we drop the time subscript 2 and add a superscript of $*$ to highlight that the ETF premium, ETF fundamental mispricing, and the NAV fundamental mispricing are equilibrium outcomes.

Proposition 1. *The $t=2$ equilibrium ETF premium is given by,*

$$\lim_{N\to\infty} \psi^* = \lim_{N\to\infty} \left(1 - \frac{N}{N+1}\right)(\epsilon^{etf} - \epsilon^{nav}) = 0. \tag{11}$$

The $t=2$ equilibrium ETF fundamental mispricing and NAV fundamental mispricing are given by,

$$\lim_{N\to\infty} \varphi^* = \lim_{N\to\infty} \alpha^* = \epsilon^{etf}\frac{\lambda}{\lambda+\eta} + \epsilon^{nav}\frac{\eta}{\lambda+\eta}. \tag{12}$$

While we focus on the limiting case in which $N\to\infty$, Equation (11) shows that, for a finite N, the equilibrium ETF premium is linear in the difference of the two non-fundamental demand shocks (i.e., $\epsilon^{etf} - \epsilon^{nav}$), similar to the ETF flow.

17 See also the models of Pástor and Stambaugh (2012), Brown and Wu (2016), and Brown and Davies (2017).

18 The "fund-flow anomaly" is empirical evidence that shows mutual fund investors chase realized performance (i.e., inflows after positive relative performance and outflows after negative relative performance), but this return-chasing behavior is unrelated to the fund's subsequent performance. See Ippolito (1992), Chevalier and Ellison (1997), and Sirri and Tufano (1998).

Remark 4. In the absence of perfect competition, ETF premia changes are symptomatic of non-fundamental demand distortions.

Remark 4 mirrors Remark 2: observed ETF premia changes are symptomatic of non-fundamental demand shocks if N is finite. However, ψ^* is decreasing in N and goes to zero in the limit, while Δ^* goes to a finite quantity in the limit. Thus, if ETF flows and ETF premia are measured with error, the signal-to-noise ratio in using Δ^* to signal non-fundamental demand is increasing in N and the signal-to-noise ratio in using ψ^* to signal non-fundamental demand is decreasing in N. In our empirical analysis, we attempt to identify fundamental mispricing using both ETF flows and ETF premia changes. We find that ETF flows are strong signals of non-fundamental demand, while ETF premia changes are not, consistent with the authorized participant market being highly competitive.

Remark 5. If $\lambda > 0$, then the non-fundamental demand shock ϵ^{etf} may be transmitted to the underlying assets via the ETF primary market mechanism. If $\eta > 0$, then the non-fundamental demand shock ϵ^{nav} may be transmitted to the ETF shares via the ETF primary market mechanism.

Remark 5 implies two potential sources of price distortions for the ETF share price and the ETF underlying asset prices. The first source is straightforward, that is, the latent demand shocks themselves: ϵ^{etf} distorts the ETF share price and ϵ^{nav} distorts the underlying asset prices. The second source, which is more subtle, is a transmission mechanism via the primary market: arbitrageurs trade against relative mispricing until it dissipates, implying that arbitrage trades have price impact. Thus, if a price distortion occurs on one side of the ETF basket (in either the ETF shares or the underlying assets), subsequent arbitrage trades transmit the price distortion to the other side of the basket. In Section 8 of the Online Appendix, we provide a test for identifying these two different sources and we show empirically that both sources distort ETF share prices and ETF underlying asset prices.

Proposition 1 also shows that the equilibrium ETF fundamental mispricing and the equilibrium NAV fundamental mispricing are a weighted average of the two non-fundamental demand shocks in which the weight on the ETF demand shock is $\frac{\lambda}{\lambda+\eta}$ and the weight on the NAV demand shock is $\frac{\eta}{\lambda+\eta}$. When η and λ are equal, the fundamental mispricing is a simple average of ϵ^{etf} and ϵ^{nav}.

At $t = longterm$, we assume the ETF share price and underlying asset prices converge to their latent fundamental value. That is, while short-run responses to fundamental mispricing are inelastic, long-run responses are highly elastic (e.g., due to slow-moving capital as in Duffie, 2010).

Remark 6. ETF flows almost always signal a fundamental mispricing that subsequently reverses.

Remark 6 highlights the primary motivation for our empirical analysis. While relative pricing is efficient ($\psi^* = 0$), ETF flows reveal fundamental mispricing ($\varphi^* = \alpha^* \neq 0$) that reverses in time (the long-term price reversion is $-\varphi^* = -\alpha^*$).[19] In practice, this reversion may be fast if traders are willing and able to exploit the profitable trade quickly. Conversely, the reversion may be slow if traders are unwilling or unable to exploit the trade

19 Note, as can be seen in Equation (12), it is possible that the fundamental mispricing is zero even if there is a non-fundamental demand shock, for example, when ϵ^{etf} and ϵ^{nav} exactly offset.

due to market frictions or risks. The long-term reversion may also be obscured by fundamental price movements that are correlated with ETF flows. So while the reversion would still exist, its magnitude may be small and difficult to detect. Whether or not the fundamental mispricing is observable is an empirical question and the subject of our study.

Finally, how ETF flows are related to future returns depends on whether ETF shares or the underlying assets are more sensitive to non-fundamental demand shocks. While our model is agnostic regarding the direction of the relation, the sign of the relation is given by the slope coefficient from a regression of long-term returns on ETF flows (assuming $\rho = 1$),

$$\frac{Cov(-\varphi^*, \Delta^*)}{Var(\Delta^*)} = \frac{Cov(-(\lambda\epsilon^{etf} + \eta\epsilon^{nav}), \epsilon^{etf} - \epsilon^{nav})}{Var(\epsilon^{etf} - \epsilon^{nav})} \tag{13}$$

$$= \frac{(\sigma_n - \sigma_e)(\lambda\sigma_e + \eta\sigma_n)}{Var(\epsilon^{etf} - \epsilon^{nav})}, \tag{14}$$

with the sign on the coefficient being determined by $\sigma_n - \sigma_e$.

Remark 7. If ETF share demand is relatively more sensitive to non-fundamental demand shocks as compared to the ETF underlying asset demand, that is, $\sigma_e > \sigma_n$, then ETF flows negatively predict returns. If $\sigma_n > \sigma_e$, then ETF flows positively predict returns.

Remark 8. Because $\varphi^* = \alpha^*$, return predictability of ETF share returns and NAV returns should be qualitatively the same.

Remarks 7 and 8 provide two insights to guide the subsequent empirical analysis. According to Remark 7, the sign on the coefficient from return predictability regressions provides insights into whether the ETF shares or the underlying assets are more sensitive to non-fundamental demand shocks. In our empirical tests, we consistently find a statistically significant negative coefficient, suggesting that ETF share demand is relatively more sensitive to non-fundamental demand shocks. According to Remark 8, our results should be approximately the same using ETF share returns or NAV returns as the dependent variable, which we find to be the case.

As discussed at the beginning of this section, the insight that ETF flows must imply a non-fundamental distortion is not driven by the setup or assumptions. However, there are several assumptions and modeling conveniences used which limit what we can learn about the ETF market. Notably, (i) we assume that the opportunity cost of arbitrage capital is zero, (ii) we do not model the authorized participants' trades in a dynamic framework, and (iii) we ignore the possibility that restoring relative price efficiency in one ETF may generate mispricing in a different ETF. Considering the possibility that arbitrage capital is costly implies that some arbitrage opportunities will not be large enough to attract authorized participants. Moreover, if the cost of arbitrage capital is stochastic, ETF flows and changes in ETF premia may reflect innovations to this cost.[20] In regards to our model's timing, we implicitly assume that authorized participants treat a relative mispricing as a one-shot game. In reality, authorized participants may trade against relative pricing while taking into consideration the possibility of future trades and the future actions of their competitors. Finally, we assume that authorized participants focus on a single arbitrage

20 Importantly, in our empirical analysis, we address that arbitrage activity is an equilibrium outcome and it reflects, among other things, the cost of arbitrage capital and macroeconomic conditions. Specifically, we include both date fixed effects and fund-level controls in our panel regressions.

Table I. Yearly ETF Sample

Notes: ETFs are included in the $50M+ sample from the first month in which end-of-month market capitalization exceeds $50 million. $50M+ ETFs are considered mature from the first month in which creation or redemption activity was reported on at least 50% of trading days. Leveraged ETFs include long (2× and 3×) and short (–1×, –2×, and –3×) funds. We require at least 12 months of data for an ETF to be included in our sample. Number and market capitalization are measured at the end of each calendar year. Market capitalization is reported in billions.

Year	All ETFs		$50M+ ETFs		Mature ETFs		Mature unleveraged		Mature leveraged	
	Number	Market cap.	Number	Market cap.	Number	Market cap.	Number	Market cap.	Number	Market cap.
2007	559	$605	370	$603	124	$516	119	$511	5	$5
2008	667	$532	445	$530	178	$468	159	$451	19	$17
2009	769	$774	524	$769	226	$687	189	$662	37	$25
2010	927	$993	635	$988	270	$891	225	$865	45	$26
2011	1,119	$1,044	731	$1,039	331	$956	278	$929	53	$26
2012	1,205	$1,341	807	$1,336	360	$1,223	301	$1,200	59	$23
2013	1,299	$1,682	910	$1,677	408	$1,529	344	$1,501	64	$28
2014	1,409	$1,976	1,029	$1,971	439	$1,785	370	$1,757	69	$28
2015	1,576	$2,108	1,113	$2,103	516	$1,914	436	$1,885	80	$29
2016	1,515	$2,526	1,101	$2,513	518	$2,275	438	$2,246	80	$29

opportunity. In reality, there may be dozens of arbitrage opportunities at a given moment and an authorized participant's optimal trading is more complicated than what we study. Thus, while our model is simple and insightful, the model has limitations that merit future research.

3. Data

We study the ability of ETF flows and ETF premia changes to predict subsequent returns using data from several sources. From Bloomberg, we get daily data on ETF share prices, NAVs, shares outstanding, and trading volumes.[21] Each date, we calculate ETF premia (discounts) as the difference between each ETF's price and its NAV and report the value as a percentage of NAV. We merge these data with information from the Center for Research in Security Prices (CRSP), including Lipper Codes and ETF returns, as well as holdings data for many of our sample ETFs. To calculate risk-adjusted measures of returns, we add information on the Fama–French three-factor (Fama and French, 1993), Fama–French three-factor plus momentum (Carhart, 1997), and Fama–French five-factor (Fama and French, 2015) models from Kenneth French's website, as well as information on the Hou–Xue–Zhang four-factor model (Hou, Xue, and Zhang, 2015) provided by Lu Zhang.

Table I displays a time-series count of the number of ETFs in our sample. The ETF market has grown rapidly over the last decade, and by the end of 2016, our sample includes

21 A number of ETFs have anomalous data on prices and shares outstanding that appear to be incorrect. We clean the data by removing the anomalies that are not verifiable via other data sources. See Section 9 of the Online Appendix for more details on database construction and cleaning. Furthermore, Ben-David, Franzoni, and Moussawi (2018) suggest that Bloomberg provides the most accurate daily ETF data.

1,515 unique ETFs.[22] To mitigate the impact of illiquidity and possible non-synchronous prices due to infrequent trading, we limit our sample to ETFs with at least $50 million in assets. Of the total ETFs, 27% are excluded using the $50 million threshold, but they collectively account for less than 1% of market capitalization. To focus on ETFs with active primary markets, we also consider a sample of ETFs that are flagged as "mature" once they exceed the $50 million threshold and experience a month in which at least one-half of the trading days had some share creation/redemption activity. As shown in Table I, this filter removes approximately half of the remaining ETFs, but only reduces the total market capitalization by 10%.

In several of our tests, we split our mature sample into two subsets: unleveraged mature ETFs and leveraged mature ETFs. There are good reasons to split unleveraged and leveraged ETFs into subsets. First, leveraged ETFs are characterized by relatively more extreme flows as compared to unleveraged ETFs. For example, leveraged ETFs represent only 15.0% of the monthly ETF observations in our mature sample but they represent 33.2% of observations in the top and bottom deciles of our portfolio sorts. Second, leveraged ETFs are uniquely tailored for short-horizon traders who want to make magnified bets on the performance of a benchmark index. As such, leveraged ETFs are characterized by high share turnover and low institutional ownership relative to unleveraged ETFs.[23] Third, leveraged ETF shares are backed by derivative contracts (e.g., total return swaps) and share creations/redemptions are performed using cash. Conversely, almost all unleveraged ETF shares are backed by an underlying basket of securities that replicate the benchmark index (either through full replication or an optimized replication to avoid illiquid securities) and they use in-kind creations/redemptions. The final two columns of Table I provide the time-series counts of the unleveraged mature sample and the leveraged mature sample. Early in the sample, leveraged mature ETFs represent only a small percentage of the mature ETF sample, both in terms of count and total market capitalization. By the end of the sample, leveraged mature ETFs represent a larger percentage of the mature ETF sample count, but they remain only a small percentage of total market capitalization.

Table II displays summary statistics for six samples: the entire ETF sample, the sample of $50M+ ETFs, the mature sample of ETFs, the unleveraged mature and leveraged mature samples, and a sample of unleveraged ETFs with high primary market activity (which are introduced and studied later in the paper). In comparing the entire ETF sample to the $50M+ sample, the $50M+ sample is generally representative of the entire sample with $50M+ ETFs exhibiting better liquidity (tighter bid-ask spreads and greater short interest percentages) and experiencing more frequent and larger flows. The two samples are nearly identical in terms of the Lipper category percentages. In comparing the $50M+ ETFs to the mature ETFs, the mature sample ETFs are larger and generally experience more trading and better liquidity; mature ETFs have more shares outstanding, more turnover, and tighter bid-ask spreads. In terms of Lipper categories, the two samples are fairly similar, but mature ETFs tend to be more focused on equities and less focused on bonds and international assets.

Table II shows that the mature unleveraged ETFs and mature leveraged ETFs greatly differ. The leveraged sample exhibits considerably more trade volume, both in shares and as a percentage of shares outstanding, consistent with leveraged ETFs being used primarily by

22 While we have data on 1,707 ETFs at the end of 2016, we require at least 12 months of observations to be included in our sample. As a result, the ETFs introduced in 2016 are not in our sample.

23 See Davies (2020) for additional detail.

Table II. ETF sample characteristics

Notes: ETFs are included in the \$50M+ sample from the first month in which end-of-month market capitalization exceeds \$50 million. \$50M+ ETFs are considered mature from the first month in which creation or redemption activity was reported on at least 50% of trading days. Leveraged ETFs include long (2× and 3×) and short (–1×, –2×, and –3×) funds. Trading days are considered active if the number of shares outstanding changed, indicating either creation or redemption activity. Average creation/redemption activity is measured by aggregating the absolute value of daily percentage changes in shares outstanding at the monthly level. The Lipper category "Bonds" includes ETFs classified as "Mixed" or "Municipal," and the "Commodities" category includes ETFs classified as "Currency."

	All ETFs	\$50M+ ETFs	Mature ETFs	Mature Unleveraged	Mature Leveraged	High Activity Unleveraged
Average ETF Characteristics						
Shares outstanding (millions)	20.9	30.0	60.0	68.3	13.3	111.1
Average monthly volume (millions)	24	35	76	71	107	165
Average monthly volume (percentage of shares out)	82	94	158	79	602	149
ETF market capitalization (billions)	\$1.2	\$1.7	\$3.5	\$4.0	\$0.5	\$7.6
Bid-ask spread (%)	0.34	0.19	0.10	0.10	0.10	0.06
Short interest percentage (%)	5.7	6.7	11.6	11.9	9.8	25.6
Average premium (%)	0.10	0.09	0.09	0.11	−0.01	0.29
Percentage of active days (%)	15.9	21.8	36.9	37.5	34.0	64.4
Average creation/redemption activity (%)	14.4	15.2	19.5	16.6	35.8	31.2
Monthly observations	126,668	87,710	38,283	32,545	5,738	10,792
Lipper category percentages						
Broad equities (%)	33.8	33.6	33.2	25.7	76.0	32.7
Sector equities (%)	23.0	23.8	28.6	33.1	2.9	35.6
Bonds (%)	18.5	17.7	13.9	14.1	12.8	16.5
Commodities (%)	5.2	6.1	6.9	6.7	8.4	5.3
International (%)	18.7	18.7	17.4	20.5	0.0	9.9

short-horizon traders. Furthermore, while the percentages of days with share creation/redemption activity are approximately the same, the magnitude of flows is over twice as large in the leveraged sample.

Motivated by our theoretical model, we examine the relation between signals of non-fundamental demand shocks (i.e., ETF flows and ETF premia changes) and subsequent returns. We calculate ETF flows and ETF premia changes at a monthly frequency and we avoid higher frequency measures for several reasons. First, the accounting standards for share creation/redemption activity vary across ETFs—some funds use $T+1$ accounting (i.e., they register the share creation activity the day after it occurs), while other funds use T accounting. Moreover, these accounting standards have changed over time, and the change from $T+1$ to T accounting, or vice versa, is not public.[24] Second, there is evidence that authorized participants strategically delay creating or redeeming shares to take advantage of failure-to-deliver rules at clearinghouses. Evans et al. (2019) describe how authorized

24 See Staer (2016) for additional details.

participants can wait until $T+6$ to create new shares and avoid costs associated with short-selling. Third, arbitrage activity itself creates price pressure in the ETF and underlying assets.[25] To avoid the effects of price pressure from arbitrage activity, and instead emphasize longer-term return predictability due to non-fundamental demand shocks, we measure ETF flows at the monthly horizon.

4. Non-Fundamental Demand and Return Reversals

Our model shows that ETF flows and ETF premia changes signal relative non-fundamental demand shocks that generate mispricing between the ETF shares and the ETF underlying assets.[26] The model also shows that ETF flows and ETF premia changes provide conditioning information to identify fundamental mispricing. However, the model is agnostic regarding the magnitude of the fundamental mispricing and the horizon at which it reverses. In this section, we analyze the relation between lagged flows/premia changes and future returns in the cross-section. Under the null hypothesis, there should be no return predictability from either ETF flows or ETF premia changes. That is, any fundamental mispricing (i.e., a risky arbitrage) is either too small to be measured or is quickly exploited by market participants.

To measure ETF flows, we calculate creation and/or redemption activity in a given ETF. Formally, we define ETF flow as the percentage change in ETF shares outstanding for fund j at time t,

$$ETF\,Flow_{j,t} = \frac{SharesOutstanding_{j,t}}{SharesOutstanding_{j,t-1}} - 1. \tag{15}$$

To measure ETF premia changes for ETF j, we calculate the ETF premium at time t and subtract the ETF premium at time $t-1$,

$$ETF\,Prem\,Change_{j,t} = (p_t/\pi_t - 1) - (p_{t-1}/\pi_{t-1} - 1), \tag{16}$$

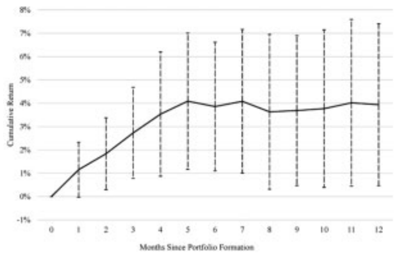
in which p_t is the ETF share price at time t and π_t is the NAV per share at time t.

We sort ETFs into portfolios based on either $ETFFlow_{j,t}$or $ETFPremChange_{j,t}$to test whether these signals of relative non-fundamental demand shocks are related to future ETF performance. In all of our portfolio sorts, we sort based on characteristics at a monthly level, preventing time trends and differences in sample size from driving our results. We measure ETF performance using ETF returns from the months following portfolio formation. For the sorts using ETF flows, ETFs with the most positive flows (i.e., share creations) in the past month are sorted into Decile 10, and ETFs with the most negative flows (i.e., share redemptions) are sorted into Decile 1.[27] For the sorts using ETF premia changes,

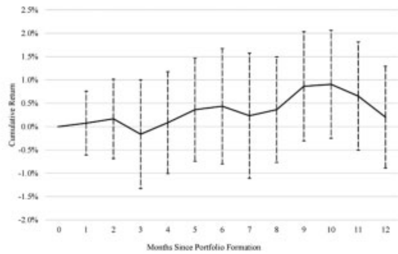
25 Ben-David, Franzoni, and Moussawi (2018) show that arbitrage activity itself leads to increased volatility in the underlying assets due to price pressure from ETF flows. See also Fulkerson and Jordan (2013). Similarly, Table IA21 of the Online Appendix shows that in our sample, large ETF premia changes lead to predictable ETF and NAV returns (in opposite directions) over the following day.

26 ETF premia changes only signal non-fundamental demand shocks if the market for authorized participants is not perfectly competitive.

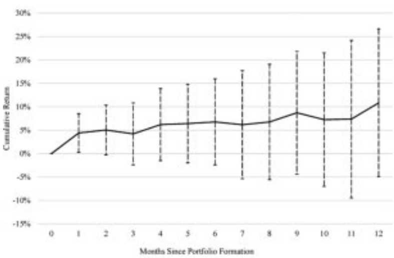
27 Because $ETFFlow_{j,t}$and $ETFPremChange_{j,t}$are both measured as percentage changes, it is possible that smaller ETFs are more likely to be sorted into the extreme deciles and thereby drive the results of our portfolio sort tests. Table IA12 of the Online Appendix shows that this is not the case by first sorting ETFs by size (market capitalization) before sorting into flow deciles. Large ETFs show strong return predictability, particularly for mature, unleveraged ETFs.



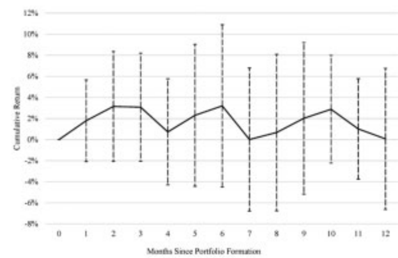
(a) Mature ETFs: $ETFFlow$



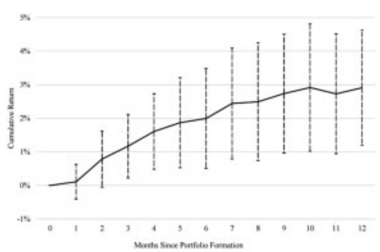
(b) Mature ETFs: $ETFPremChange$



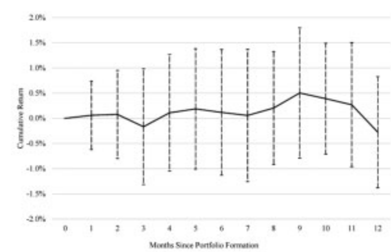
(c) Lev. ETFs: $ETFFlow$



(d) Lev. ETFs: $ETFPremChange$



(e) Unlev. ETFs: $ETFFlow$



(f) Unlev. ETFs: $ETFPremChange$

Figure 2. Returns to long-short portfolios based on ETF returns. Each month, ETFs are sorted based on the prior month's *ETFFlow* or *ETFPremChange*. ETFs with the highest redemption activity or most negative premium change are sorted into Decile 1, while ETFs with the highest creation activity or most positive premium change are sorted into Decile 10. Value-weighted portfolio returns are calculated by forming long-short portfolios, which are long Decile 1 ETFs and short Decile 10 ETFs. Portfolio returns are calculated using ETF returns for the 12 months following formation, and each horizon's returns are averaged across months. Error bars provide the 95% confidence interval for each average return. (a) Mature ETFs: *ETFFlow*. (b) Mature ETFs: *ETFPremChange*. (c) Leveraged ETFs: *ETFFlow*. (d) Leveraged ETFs: *ETFPremChange*. (e) Unleveraged ETFs: *ETFFlow*. (f) Unleveraged ETFs: *ETFPremChange.*

ETFs with the largest increase in average premia in the past month are sorted into Decile 10 and ETFs with the largest decrease in average premia are sorted into Decile 1.

Figure 2 plots the value-weighted raw ETF returns from a long-short trading strategy that goes long the ETFs in Decile 1 and goes short the ETFs in Decile 10. In each plot, the horizontal axis represents the number of months following the formation of the long-short portfolio and the vertical axis represents the portfolio's raw return. At each monthly data point, error bars are included to provide the 95% confidence interval. Figure 2(a) and (b) depicts the trading strategy raw returns for the mature ETF sample, based on ETF flows and based on ETF premia changes. Figure 2(c) and (d) uses the mature leveraged ETFs, and Figure 2(e) and (f) uses the mature unleveraged ETFs.

The plots illustrate four findings that guide our subsequent analysis. First, ETF flows appear to provide cross-sectional return predictability: in the mature ETF sample, the mature leveraged ETF sample, and the mature unleveraged ETF sample, raw returns appear to be positive and significantly different than zero (albeit, over different horizons). Second, ETF flows are negatively related to subsequent returns which suggest, according to Remark 7, that ETF shares are relatively more sensitive to non-fundamental demand shocks on average. Third, ETF premia changes do not appear to predict cross-sectional returns. While the returns from conditioning on ETF premia changes are generally positive, they are not statistically different from zero. As such, ETF flows are better signals for non-fundamental demand shocks than ETF premia changes. This is consistent with our model's prediction if the primary market for ETFs is highly competitive and premia are measured with noise. In what follows, we focus exclusively on ETF flows, but we include additional analysis in Table IA8 of the Online Appendix using ETF premia changes. Fourth, the return predictability horizons differ significantly between mature leveraged ETFs and mature unleveraged ETFs. Almost all returns in the trading strategy using mature leveraged ETFs are earned in the first month. Conversely, the returns in the trading strategy using mature unleveraged ETFs are primarily earned in months two through six.

Motivated by the plots in Figure 2, Panel A of Table III displays raw returns for portfolios formed using lagged ETF flows.[28] We construct equal-weighted and value-weighted (based on ETFs' assets under management) long-short portfolios and report portfolio returns at 1-, 3-, and 6-month horizons. Beginning with our broadest samples, the \$50M+ and mature ETF samples exhibit statistically significant long-short portfolio returns at the 1-, 3-, and 6-month horizons (to account for overlapping return periods, we use Newey–West standard errors with the lag equal to the return horizon in months). One-month long-short portfolio returns range from 1.1% to 2.0% monthly (which compounds to 14.5–27.2% annually), 3-month returns range from 2.3% to 2.7% quarterly (9.7–11.4% annually), and 6-month returns range from 3.6% to 3.9% semiannually (7.3–7.9% annually). Collectively, these results show that ETF flows predict future returns in the cross-section.

To examine the differences in return predictability horizons shown in Figure 2, we analyze mature leveraged and unleveraged ETFs separately. Beginning with the mature leveraged ETF results, 1-month long-short portfolio returns are statistically significant and range from 4.4% to 4.5% (67.8–68.7% annually).[29] However, over 3- and 6-month

28 In Section 1 of the Online Appendix, we replicate Table III using risk-adjusted returns from the Fama–French three-factor model (Fama and French, 1993), the Fama–French four-factor Carhart (1997) model, the Fama–French five-factor model (Fama and French, 2015), and the Hou–Xue–Zhang four-factor model (Hou, Xue, and Zhang, 2015). See Tables IA1–IA4 of the Online Appendix.

29 In Tables IA5 and IA6 of the Online Appendix, we show that our leveraged ETFs results are not driven by the optionality embedded in leveraged ETFs. Specifically, leveraged ETFs follow a dynamic trading strategy that synthesizes an Asian option, so their realized returns depend on the realized volatilities and realized returns of the underlying indexes. To ensure our results are not driven by this feature, we de-leverage these ETF returns. Our conclusions remain unchanged.

Table III. Portfolio sorts on ETF flow: ETF returns

Notes: Each month, ETFs are sorted based on the prior month's *ETFFlow*. ETFs with the highest redemption activity are sorted into Decile 1, while ETFs with the highest creation activity are sorted into Decile 10. Raw ETF return differences are tested by forming long-short portfolios, which are long high redemption ETFs and short high creation ETFs. In Panel B, ETFs are first sorted based on the prior month's primary market activity, which is measured as the number of days with changes in shares outstanding. ETFs are then sorted based on the prior month's *ETFFlow*. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively, using Newey–West standard errors with the lag equal to the return horizon in months.

	One month			Three months			Six months		
	Decile 1	Decile 10	Long-short	Decile 1	Decile 10	Long-short	Decile 1	Decile 10	Long-Short
Panel A: Baseline results									
$50M+ ETFs									
Equal-weighted	0.587*	−0.932**	1.519***	0.788	−1.546**	2.334***	1.423	−2.157**	3.580***
Value-weighted	0.738*	−0.395	1.134**	1.684**	−0.979	2.663***	2.794**	−0.991	3.785***
Mature ETFs									
Equal-weighted	0.704**	−1.321***	2.025***	0.695	−1.880**	2.575***	1.272	−2.519**	3.791***
Value-weighted	0.689*	−0.465	1.154*	1.686**	−1.044	2.730***	2.951**	−0.908	3.859***
Mature, leveraged ETFs									
Equal-weighted	1.072	−3.380***	4.452***	−0.972	−5.015**	4.043	−2.639	−8.024***	5.385
Value-weighted	1.310	−3.099**	4.409**	−0.820	−5.039**	4.219	−1.573	−8.343**	6.771
Mature, unleveraged ETFs									
Equal-weighted	0.238	0.063	0.175	0.959	0.523	0.435	2.183	1.009	1.175*
Value-weighted	0.441	0.331	0.110	1.636**	0.466	1.170**	3.022**	1.028	1.994***
Panel B: Mature, unleveraged ETFs									
Low activity ETFs									
Equal-weighted	0.257	0.844*	−0.586**	1.122	1.917**	−0.795*	2.555*	3.205**	−0.651
Value-weighted	0.378	0.880*	−0.502	1.441*	1.390	0.050	2.360*	2.663*	−0.303
Medium activity ETFs									
Equal-weighted	0.376	−0.007	0.383	0.892	0.421	0.470	1.734	1.001	0.733
Value-weighted	0.257	0.423	−0.165	1.009	0.471	0.538	1.699	0.586	1.113
High activity ETFs									
Equal-weighted	0.388	−0.025	0.413	1.066	−0.156	1.222**	2.602**	0.091	2.512***
Value-weighted	0.690	0.273	0.416	1.697**	0.157	1.539**	3.609***	0.277	3.332***

horizons, the leveraged ETF returns are flat relative to what is earned in the first month and are not significantly different from zero. Turning to the mature unleveraged ETF results, returns do not exhibit predictability at a 1-month horizon. However, at the 3-month horizon, the value-weighted, long-short portfolio earns statistically significant returns of 1.2% (4.8% annually). At the 6-month horizon, both the equal-weighted and value-weighted portfolios earn statistically significant returns of 1.2% (2.4% annually) and 2.0% (4.0% annually).

Our model provides two reasons why leveraged ETFs show strong return predictability. Remark 3 suggests that the best measures of non-fundamental demand shocks occur in ETFs in which (i) flows are volatile, and (ii) there are different trader clienteles transacting the ETF shares and the ETF underlying assets. Leveraged ETFs meet both conditions. First, Table II shows that leveraged ETFs have over twice as much creation/redemption activity as unleveraged ETFs. Second, while traders purchase leveraged ETF shares for short-horizon, magnified exposure to market benchmarks, the shares are backed by derivative securities (e.g., total return swaps) which are transacted for many other purposes such as risk management and hedging. Furthermore, any trader can purchase leveraged ETF shares without special authorizations (even in some retirement accounts like 401(k)s), while the derivative securities are traded almost exclusively among institutions. Thus, there is a wedge between the clienteles that trade leveraged ETF shares and that trade the underlying derivative securities. One clientele is largely retail and motivated by short-horizon, leveraged exposure and the other clientele is largely institutional and motivated by multiple purposes including hedging.

Remark 3 is not specific to leveraged ETFs. ETFs with active primary markets feature volatile flows and different clienteles owning the ETFs and their underlying assets. Accordingly, we perform an additional test on the mature unleveraged ETF sample in Panel B of Table III. The additional test provides robustness in showing that our return predictability results are not driven exclusively by leveraged ETFs. To focus on ETFs with active primary markets, we first sort the sample into terciles based on the number of days in the month for which there was any share creations/redemptions.[30] We refer to the top tercile as high activity ETFs. Within each tercile, we sort ETFs into deciles according to ETF flows. Panel B shows stronger return predictability for high activity ETFs as compared to the full sample of unleveraged ETFs in Panel A, both in terms of economic magnitudes and statistical significance. The high activity ETFs exhibit statistically significant return predictability at both 3- and 6-month horizons. For equal- and value-weighted portfolios, the 3-month long-short portfolio returns are 1.2% (5.0% annually) and 1.5% (6.3% annually), and the 6-month returns are 2.5% (5.1% annually) and 3.3% (6.8% annually). The analysis shows that, while the results are most pronounced in the mature leveraged ETF sample, the results extend to unleveraged ETFs as well.[31]

The horizons at which the returns persist are also of interest. The existing literature on investor flows typically finds evidence of long-horizon predictability. For example, Frazzini

30 In Table IA9 of the Online Appendix, we replicate this analysis by sorting on flow volatility instead of an activity. While the results are noisier, our conclusions remain unchanged.

31 Tables IA10 and IA11 of the Online Appendix analyze return predictability separately for broad equity, sector equity, bond, international, and commodity ETFs. The results show that ETF categories with more active primary and secondary markets show more return predictability, consistent with the results for leveraged ETFs and unleveraged, high activity ETFs.

and Lamont (2008) document evidence of return predictability at a horizon of several years. Instead, we find return predictability over the 1- to 6-month horizons. Table III shows that the non-fundamental demand shocks measured by flows in leveraged ETFs generate short-lived dislocations that correct over 1 month, while non-fundamental demand shocks measured by flows in unleveraged ETFs generate longer-lived dislocations that take several months to remedy. The difference in horizons for the return predictability could be due to several factors. For example, leveraged ETFs and unleveraged ETFs may attract different investor clienteles who are exposed to different types of non-fundamental demand. Alternatively, market participants may face different risks in exploiting the fundamental mispricing signaled by leveraged ETF flows as compared to the fundamental mispricing signaled by unleveraged ETF flows.

Return predictability from ETF flows is related to, but distinct from, the large literature showing that mutual fund flows contain information about future asset returns. For example, Ben-Rephael, Kandel, and Wohl (2012) show that flows aggregated across mutual funds can be used to measure non-fundamental demand. To test whether our findings are distinct from theirs, we orthogonalize *ETFFlow* to US aggregate net equity fund flows from Ben-Rephael, Kandel, and Wohl (2012). We document in Table IA7 of the Online Appendix that ETFlow is not simply measuring a known source of return predictability in a different way. The coefficient estimates and standard errors are consistent with those in Table III and lead to the same conclusions. In other words, the signal of non-fundamental demand from ETF flows is different from that found in aggregate mutual fund flows.

Remark 8 predicts that return predictability should be similar using either ETF returns or NAV returns. Table IV uses NAV returns and confirms this prediction by showing very similar results to the ETF returns in Table III.[32] In addition to validating our model, finding similar return predictability for ETF and NAV returns reinforces that the return predictability is not due to the arbitrage activity itself (which puts opposite pressure on the ETF and NAV). Rather, our findings are consistent with long-term price reversals following non-fundamental demand shocks that are signaled by arbitrage activity.

4.1 Panel Regressions

To examine the robustness of our portfolio sorts, we estimate the panel regression:

$$Performance_{j,t+h} = \beta FlowMeasure_{j,t} + \gamma_t + \delta V_{j,t} + \epsilon_{j,t+h}, \tag{17}$$

in which $Performance_{j,t+h}$ is a measure of ETF j's performance (i.e., raw return including distributions or abnormal returns using a risk model) at horizon h, h is either 1, 3, or 6 months, $FlowMeasure_{j,t}$ is either the set of indicator variables $Decile_{j,d,t}$ for whether ETF j is in decile portfolio d in period t (from sorting on $ETFFlow_{j,t}$) or the continuous measure $ETFFlow_{j,t}$, γ_t are date fixed effects, and $V_{j,t}$ are lagged fund controls including two lags of ETF returns, ETF premium, ETF share volume, and ETF market capitalization. To account for overlapping return periods, we calculate Driscoll and Kraay's (1998) standard errors with the lag equal to the return horizon in months.

We calculate abnormal returns by regressing monthly ETF returns on Fama–French three-factor plus momentum (Carhart, 1997) factor returns:

32 Moreover, Section 6 of the Online Appendix shows that our results extend to the individual stock level.

Table IV. Portfolio sorts on ETF flows: NAV returns.

Notes: Each month, ETFs are sorted based on the prior month's *ETFFlow*. ETFs with the highest redemption activity are sorted into Decile 1, while ETFs with the highest creation activity are sorted into Decile 10. Raw NAV return differences are tested by forming long-short portfolios, which are long high redemption ETFs and short high creation ETFs. In Panel B, ETFs are first sorted based on the prior month's primary market activity, which is measured as the number of days with changes in shares outstanding. ETFs are then sorted based on the prior month's *ETFFlow*. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively, using Newey–West standard errors with the lag equal to the return horizon in months.

Sample	One month			Three months			Six months		
	Decile 1	Decile 10	Long-short	Decile 1	Decile 10	Long-short	Decile 1	Decile 10	Long-short
Panel A: Baseline results									
$50M+ ETFs									
Equal-weighted	0.537*	−0.906**	1.444***	0.664	−1.501**	2.165***	1.240	−2.128**	3.368***
Value-weighted	0.718*	−0.411	1.129**	1.613**	−0.948	2.561***	2.695**	−0.976	3.671***
Mature ETFs									
Equal-weighted	0.663**	−1.308***	1.971***	0.564	−1.838**	2.403***	1.075	−2.527**	3.602***
Value-weighted	0.674	−0.479	1.153*	1.609**	−1.024	2.633***	2.847**	−0.913	3.760***
Mature, leveraged ETFs									
Equal-weighted	0.974	−3.330***	4.304**	−1.233	−5.004**	3.771	−2.957	−8.054***	5.097
Value-weighted	1.315	−2.988**	4.303**	−0.912	−4.939**	4.027	−1.740	−8.258**	6.518
Mature, unleveraged ETFs									
Equal-weighted	0.200	0.081	0.120	0.875	0.596	0.279	2.058	1.070	0.988
Value-weighted	0.415	0.305	0.110	1.590**	0.511	1.079**	2.953**	1.069	1.884**
Panel B: Mature, unleveraged ETFs									
Low activity									
Equal-weighted	0.199	0.823*	−0.625**	0.973	1.987**	−1.014**	2.384*	3.202**	−0.818
Value-weighted	0.350	0.862*	−0.512	1.332	1.525	−0.192	2.295*	2.666*	−0.371
Medium activity									
Equal-weighted	0.339	0.039	0.300	0.802	0.507	0.296	1.610	1.047	0.563
Value-weighted	0.230	0.460	−0.230	0.946	0.544	0.401	1.616	0.659	0.956
High activity									
Equal-weighted	0.352	0.013	0.339	1.007	−0.065	1.071**	2.500*	0.200	2.300***
Value-weighted	0.639	0.279	0.360	1.648**	0.206	1.442**	3.528***	0.332	3.196***

$$Ret_{j,t+h} = \alpha + \Gamma X_{t+h} + \eta_{j,t+h}, \tag{18}$$

in which X_{t+h} is a vector of factor returns for time t at horizon h and the residual, $\eta_{j,t+h}$, represents the abnormal return. Importantly, date fixed effects help differentiate our results from several existing measures that have been used as proxies for non-fundamental demand, including aggregate sentiment proxies (Baker and Wurgler, 2007) and trading frictions, like intermediary funding liquidity (He, Kelly, and Manela, 2017).[33] In addition, using two lags of ETF return controls for return predictability due to known sources of return-induced non-fundamental trades (i.e., extrapolation or contrarian trading). As such, our results are more likely to isolate noise that is distinct from other non-fundamental demand proxies.

Starting at the 1-month horizon, Column (1) of Table V shows that Decile 10 ETFs (those with the most inflows) underperform, in raw returns, Decile 1 ETFs (the omitted group and those with the most outflows) by 186 bps per month (24.8% annually), which is significant at a 1% p-value threshold. Column (2) shows that Decile 10 ETFs underperform, in abnormal returns, Decile 1 ETFs by 79 bps (9.9% annually), which is significant at a 1% p-value threshold. Column (3), which uses a continuous measure of ETF flows, shows a negative coefficient on $ETFFlow_{j,t}$, but it is not statistically significant. Comparing Columns (1) and (2) to Column (3) suggests a non-linear relation between lagged ETF flows and subsequent performance.

Turning to the 3-month horizon, Column (4) shows that Decile 10 ETFs underperform, in raw returns, Decile 1 ETFs by 210 bps per quarter (8.7% annually), which is significant at a 1% p-value threshold. Column (5) shows that Decile 10 ETFs underperform, in abnormal returns, Decile 1 ETFs by 76 bps per quarter (3.1% annually), which is significant at a 5% p-value threshold. Column (6) uses a continuous measure of ETF flows and, while the coefficient on $ETFFlow_{j,t}$is negative, it is not statistically significant.

Finally, Columns (7)–(9) consider performance measures at a 6-month horizon. Column (7) shows that Decile 10 ETFs underperform, in raw returns, Decile 1 ETFs by 375 bps per half-year (7.6% annually), which is significant at a 1% p-value threshold. Column (8) shows that Decile 10 ETFs underperform, in abnormal returns, Decile 1 ETFs by 158 bps per half-year (3.2% annually), which is significant at a 5% p-value threshold. Column (9) mirrors Columns (3) and (6); while the coefficient on $ETFFlow_{j,t}$is negative, it is not statistically significant. In Table IA13 of the Online Appendix, we show that our results are qualitatively similar using NAV returns in place of ETF returns, consistent with Remark 8.[34]

Table V confirms the long-short portfolio results in Table III and both tables document an economically meaningful and statistically significant relation between ETF flows and subsequent returns. The analysis is consistent with ETF flows' signaling non-fundamental

33 Using a panel regression with date fixed effects also alleviates potential concerns of Stambaugh's (1999) bias. Furthermore, ETF flows are neither highly persistent nor correlated with contemporaneous returns.

34 In Table IA14 of the Online Appendix, we show that these results hold after we orthogonalize *ETFFlow* to US aggregate net equity fund flows from Ben-Rephael, Kandel, and Wohl (2012). Tables IA15 and IA16 of the Online Appendix show that these results continue to hold after accounting for the optionality embedded in leveraged ETFs. Tables IA17–IA19 of the Online Appendix show that, for subsamples of leveraged and unleveraged ETFs, panel regression results are consistent with the portfolio sort results reported in Tables III and IV.

Table V. Panel regressions of ETF returns on ETF flows.

Notes: The table displays panel regressions of monthly ETF returns (in percent) on measures of past creation and redemption activity according to the model:

$$Performance_{j,t+h} = \beta FlowMeasure_{j,t} + \gamma_t + \delta V_{j,t} + \epsilon_{j,t+h},$$

in which $Performance_{j,t+h}$ is a measure of ETF *j*'s performance (i.e., raw return, including distributions, *Ret*, or abnormal returns using the Fama–French three-factor plus momentum (Carhart, 1997) factor returns, *AbnRet*) at horizon *h*, *h* is either 1, 3, or 6 months, $FlowMeasure_{j,t}$ is either the set of indicator variables $Decile_{j,d,t}$ for whether ETF *j* is in decile portfolio *d* in period *t* (from sorting on $ETFFlow_{j,t}$) or the continuous measure $ETFFlow_{j,t}$, γ_t are date fixed effects, and $V_{j,t}$ are lagged fund controls of ETF return (two lags), ETF premium, ETF share volume, and ETF market capitalization. *ETFFlow* and ETF characteristics are standardized by subtracting the sample mean and dividing by the sample standard deviation. *t*-statistics calculated using Driscoll–Kraay standard errors with the lag equal to the return horizon in months are shown below the estimates in parentheses. ***, **, * indicates statistical significance at the 1%, 5%, and 10% levels, respectively.

	One month			Three months			Six months		
	(1) *Ret*	(2) *AbnRet*	(3) *AbnRet*	(4) *Ret*	(5) *AbnRet*	(6) *AbnRet*	(7) *Ret*	(8) *AbnRet*	(9) *AbnRet*
Decile 2	−0.49*	−0.36**		0.09	−0.20		0.80	−0.22	
	(−1.76)	(−2.36)		(0.21)	(−0.67)		(1.16)	(−0.48)	
Decile 3	−0.41	−0.35**		0.42	−0.25		1.05	−0.54	
	(−1.40)	(−2.32)		(0.90)	(−0.84)	(1.30)	(−1.13)		
Decile 4	−0.48	−0.41***	0.56	−0.20		1.46	−0.45		
	(−1.62)	(−2.70)		(1.06)	(−0.71)		(1.56)	(−0.98)	
Decile 5	−0.30	−0.21		0.82	0.34		1.75*	0.16	
	(−0.95)	(−1.19)		(1.39)	(1.12)		(1.77)	(0.29)	
Decile 6	−0.27	−0.29*		0.94	0.04		2.16**	−0.24	
	(−0.83)	(−1.89)		(1.62)	(0.13)		(2.29)	(−0.46)	
Decile 7	−0.23	−0.22		0.86	0.20		1.74*	−0.04	
	(−0.82)	(−1.44)		(1.57)	(0.75)		(1.82)	(−0.07)	
Decile 8	−0.39	−0.22		0.38	0.09		1.29	−0.14	
	(−1.34)	(−1.50)		(0.72)	(0.31)		(1.42)	(−0.27)	
Decile 9	−0.67**	−0.36**		−0.28	−0.06		−0.05	−0.20	
	(−2.27)	(−2.10)		(−0.48)	(−0.17)		(−0.05)	(−0.30)	
Decile 10	−1.86***	−0.79***		−2.10***	−0.76**		−3.75***	−1.58**	
	(−4.05)	(−4.28)		(−3.29)	(−2.36)		(−4.75)	(−2.55)	
ETFArb			−0.03			−0.04			−0.05
			(−1.16)			(−1.08)			(−1.28)
Date fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lagged controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.197	0.033	0.032	0.220	0.042	0.041	0.236	0.049	0.047
Observations	37,759	37,759	37,759	36,711	36,711	36,711	35,139	35,139	35,139

Table VI. Trading strategy performance

Notes: Each month, ETFs are sorted based on the prior month's *ETFFlow*. ETFs with the highest redemption activity are sorted into Decile 1, while ETFs with the highest creation activity are sorted into Decile 10. Portfolio returns are calculated by forming long-short value-weighted portfolios, which are long high redemption ETFs and short high creation ETFs. For 1-month returns, the portfolios are only based on the prior month's *ETFFlow*, so the portfolio turns over every month. For 6-month returns, the portfolio equally weights each of the last 6 months' 1-month portfolios, so one-sixth of the portfolio turns over every month. Returns, standard deviations, and Sharpe ratios are all annualized. Unleveraged high activity ETFs are unleveraged ETFs in the top tercile of creation/redemption activity each period. Panel A includes the full sample. Panel B analyzes portfolio returns that occurred from 2007 to 2011 and Panel C analyzes portfolio returns that occurred from 2012 to 2016. The 2012–16 sample, therefore, uses data from the end of 2011 in forming the portfolios in early 2012.

Performance measures	One month			Six months		
	Mature ETFs	Leveraged ETFs	Unleveraged high Activity ETFs	Mature ETFs	Leveraged ETFs	Unleveraged high Activity ETFs
Panel A: Full sample						
Mean annual return (%)	11.96	27.21	3.07	8.03	8.08	6.05
Standard deviation (%)	19.84	74.42	13.97	8.08	39.76	7.42
Sharpe ratio	0.60	0.37	0.22	0.99	0.20	0.82
Maximum monthly loss (%)	−14.52	−54.98	−8.21	−7.80	−47.91	−6.34
Panel B: 2007 – 11						
Mean annual return (%)	19.86	21.42	9.12	12.27	−12.38	9.66
Standard deviation (%)	26.63	94.74	17.78	10.72	52.39	9.15
Sharpe ratio	0.75	0.23	0.51	1.14	−0.24	1.06
Maximum monthly loss (%)	−14.52	−50.41	−8.21	−7.80	−47.91	−6.34
Panel C: 2012–16						
Mean annual return (%)	4.66	31.98	−2.59	3.99	27.48	2.60
Standard deviation (%)	8.39	54.20	8.28	3.60	26.08	4.96
Sharpe ratio	0.55	0.59	−0.31	1.11	1.05	0.52
Maximum monthly loss (%)	−7.20	−54.98	−7.82	−2.61	−24.35	−4.06

demand shocks that push asset prices away from fundamentals. This fundamental mispricing reverses over the span of 1–6 months.

4.2 Trading Strategy Performance

To investigate the practical implications and the economic significance of our results, we evaluate the performance of several trading strategies that condition on ETF flows using the same portfolio formation process. Table VI details the performances of long-short trading strategies using three ETF samples: (i) mature ETFs, (ii) leveraged ETFs, and (iii) high activity unleveraged ETFs (i.e., those in the top tercile of primary market activity), over 1- and 6-month horizons. The value-weighted 1-month horizon strategy turns over monthly based on the prior month's flows. The 6-month horizon strategy is constructed by equally-weighting the six most recent 1-month horizon portfolios. Therefore, each month, the oldest 1-month portfolio rolls out of the set of six and the newest 1-month portfolio rolls in. For each ETF sample and at each horizon, we report the mean annual return, the annualized standard deviation of monthly returns, the annualized Sharpe ratio, and the maximum monthly loss.

Panel A of Table VI presents the results for our entire 10-year sample. For the mature ETF sample, the 1-month strategy's mean annual return is 11.96% and the 6-month strategy's mean annual return is 8.03%. For the leveraged ETF sample, the mean annual returns are 27.21% and 8.08%, and for the high activity unleveraged ETF sample, the mean annual returns are 3.07% and 6.05%. The samples' mean annual returns are consistent with the earlier findings showing that the leveraged ETF strategy predicts large return reversals over a short horizon, while the high activity unleveraged ETF strategy predicts more moderate return reversals over a longer horizon. The strong performance in the leveraged ETF sample, relative to the mature sample and the high activity unleveraged sample, is not without risk. The standard deviation of returns is 74.42% and 39.76% in the leveraged ETF sample, which corresponds to Sharpe ratios of 0.37 and 0.20. In comparison, the standard deviation of returns in the mature ETF sample is smaller, leading to stronger Sharpe ratios; the 1-month horizon and 6-month horizon annual Sharpe ratios are 0.60 and 0.99. For the high activity unleveraged ETF sample, the Sharpe ratios are the lowest for the 1-month strategy (0.22) but are higher for the 6-month strategy (0.82).

Panel B of Table VI presents the results over the sample period of 2007–11. During this window, in which the 2008 Financial Crisis falls, the mature ETF sample dominates the leveraged and high activity unleveraged samples with respect to Sharpe ratios. For the 1-month strategy, the mature sample achieves a Sharpe ratio of 0.75, while the leveraged sample and high activity unleveraged sample earn Sharpe ratios of 0.23 and 0.51. For the 6-month strategy, the mature sample achieves a Sharpe ratio of 1.14, and the leveraged and high activity unleveraged samples achieve Sharpe ratios of –0.24 and 1.06. Panel C presents the results over the second half of our sample, 2012–16. Again, the mature sample outperforms the leveraged and high activity unleveraged samples in terms of Sharpe ratios, both for the 1-month strategy and the 6-month strategy. The mature sample also outperforms in terms of maximum monthly loss.

Table VI provides two insights. First, while return predictability is strongest in the mature leveraged sample, it comes with two undesirable features: (i) the volatility of returns is three to seven times larger than the volatility in the mature sample and the high activity unleveraged sample and (ii) the increase in volatility does not correspond to proportionally greater returns, leading to lower overall Sharpe ratios. Second, both mean annual returns and Sharpe ratios are consistently higher in the mature ETF sample than in the subsamples of leveraged and high activity unleveraged ETFs. As such, there appears to be a diversification benefit in building the ETF portfolio using both leveraged and unleveraged ETFs.

5. ETF Investor Performance

The analyses in Section 4 suggest that non-fundamental demand shocks, measured by ETF flows, create relative mispricing among ETFs; those ETFs with the greatest inflows underperform those with the greatest outflows. In this section, we examine how the relation between non-fundamental demand shocks and returns affects investor performance. Specifically, for each ETF, we quantify the costs (if any) for investors due to a negative relation between ETF flows and subsequent returns. To do so, we consider two measures of ETF investor performance and examine ETFs on a fund-by-fund basis.

First, we examine each ETF's IRR, which considers both inflows and outflows to the ETF (Dichev, 2007; Dichev and Yu, 2011). Unlike the return on a single ETF share, the IRR measures the timing impact of ETF flows and it provides the return that a representative investor would have earned if she held all shares and was responsible for all ETF flows.

Second, we examine a measure that we call SGAR. SGAR measures the dynamic performance of an investor that allocates capital between ETF shares and a risk-free asset based on the prior month's flows. To calculate the SGAR for each ETF, we take its return series $\mathbf{r} = \{r_1, \ldots, r_T\}$and its one-period-lagged share growth series $\mathbf{g} = \{g_0, \ldots, g_{T-1}\}$and perform the following calculation,

$$\mathrm{SGAR} = \left(\prod_{\tau=1}^{T}((1+r_\tau)(1+g_{\tau-1}) - (1+r_{f,\tau})g_{\tau-1})\right)^{12/T} - 1, \tag{19}$$

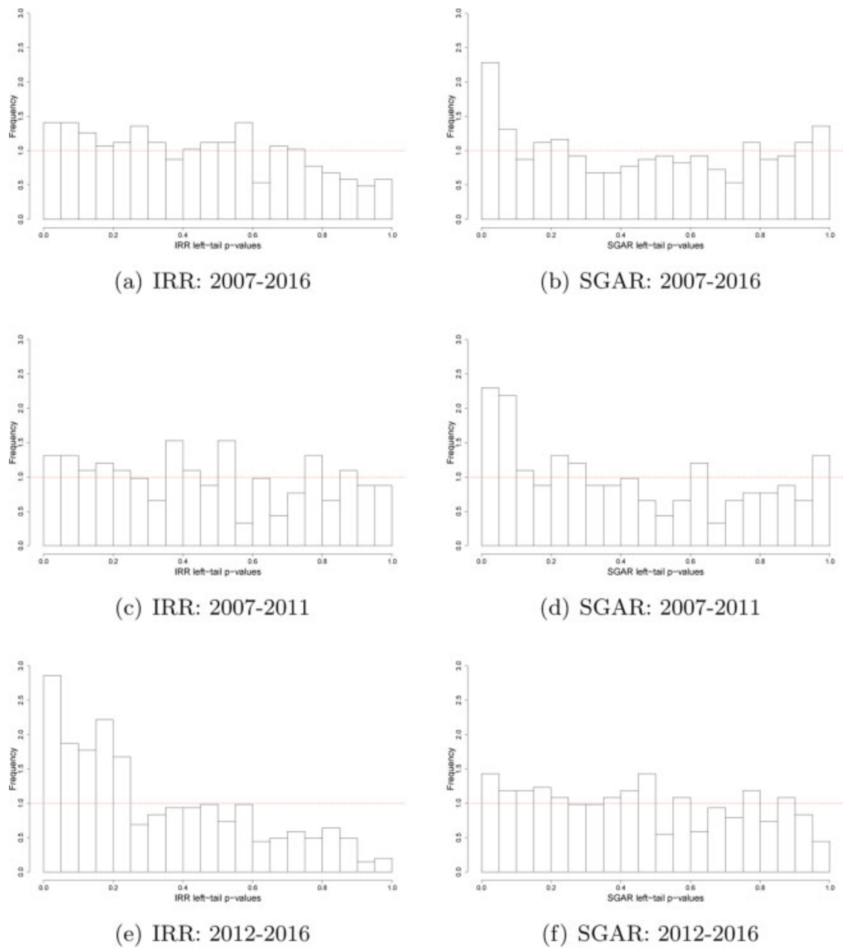
in which $r_{f,\tau}$ is the risk-free rate at time τ.[35] SGAR is a pseudo portfolio return that captures the notion that share creations and redemptions have a leverage-like effect on total return; the SGAR puts more weight on returns after inflows and less weight on returns after outflows. SGAR assumes that capital may be invested in the risk-free asset (negative values of $g_{\tau-1}$) and may also be borrowed at the risk-free rate (positive values of $g_{\tau-1}$).

The IRR and SGAR measures differ in important ways and each measure has its own strengths and weaknesses. IRR is sensitive to the correlation of flows and returns, regardless of the number of months separating a particular flow from a particular return. As such, IRR captures correlations between flows and returns without restrictions on the time series properties, allowing identification of fundamental mispricings that take several months to reverse. However, this flexibility comes with a cost as one cannot disentangle whether flows lead returns or if returns lead flows. For example, Hayley (2014) shows that IRRs may be pushed upwards by increasing inflows following bad performance and increasing outflows following good performance. SGAR only measures the relation between flows and subsequent returns. However, SGAR is rigid by construction and is only effective at measuring a relation between flows and subsequent returns if mispricing reverses in a month's time. Despite their strengths and weaknesses, the IRR and SGAR measures complement each other by providing two distinct means of evaluating investor performance.

Starting with our sample of mature ETFs, we restrict the analysis to 412 ETFs with at least 36 consecutive months of data. For each ETF, we calculate its realized IRR and SGAR and test the statistical significance of each measure via Monte Carlo simulation using 100,000 sample paths. To simulate each ETF's IRR distribution, we follow the methodology of Dichev (2007) and shuffle the vector of capital contributions (as a percentage of beginning-of-month assets) with replacement using the stationary bootstrap technique of Politis and Romano (1994).[36] To simulate each ETF's SGAR distribution, we shuffle the vector of share growth $\mathbf{g}$ with replacement also using a stationary bootstrap in each Monte Carlo path.

35 We get the monthly risk-free rate from Kenneth French's website.

36 The stationary bootstrapping is characterized by $p = 1/5$.



(a) IRR: 2007-2016 (b) SGAR: 2007-2016

(c) IRR: 2007-2011 (d) SGAR: 2007-2011

(e) IRR: 2012-2016 (f) SGAR: 2012-2016

Figure 3. Histogram of p-values for realized ETF IRRs and SGARs. Each histogram reports the p-values corresponding to realized IRRs and SGARs based on their simulated distributions. A uniform distribution's density function is depicted in each histogram for reference (dotted red line). The left column of histograms corresponds to IRRs and the right column corresponds to SGARs. (a) IRR: 2007–16. (b) SGAR: 2007–16. (c) IRR: 2007–11. (d) SGAR: 2007–11. (e) IRR: 2012–16. (f) SGAR: 2012–16.

Figure 3 and Table VII report the percentages of ETFs that have realized IRRs and SGARs in the left tails of their simulated distributions.[37] Figure 3 depicts histograms of realized p-values (i.e., the percentage of simulated returns below the realized return); in each plot, p-values appear on the horizontal axis and frequencies appear on the vertical axis. For reference, the density function for a uniform distribution is included as a horizontal line in each plot. If IRRs and SGARs were not skewed distributions, the realized p-values would resemble a uniform distribution. Instead, both IRRs and SGARs exhibit negative skew; the left-hand portion of realized frequencies is consistently above the uniform

37 In Table IA20 of the Online Appendix, we show the percentages of ETFs which have realized IRRs and SGARs in the right tails of their simulated distributions.

Table VII. Distributions of IRR and SGARs

Notes: The table presents realized IRRs and SGARs compared to their simulated distributions. Three-time period samples are analyzed: 2007–16 (the full sample), 2007–11, and 2012–16. For each sample, the table reports the percentage of ETFs falling in the left tail of their simulated distributions at thresholds of 1%, 5%, and 10%, respectively. The first three columns correspond to IRRs and the second three columns correspond to SGARs.

	IRR (%)			SGAR (%)		
	1%	5%	10%	1%	5%	10%
Entire Sample						
Equal-weighted	1.96	7.09	14.18	3.16	11.41	17.96
$N = 412$						
January 2007–December 2011						
Equal-weighted	2.19	6.56	13.11	3.83	11.48	22.40
$N = 183$						
January 2012–December 2016						
Equal-weighted	4.43	14.29	23.65	0.74	7.14	13.05
$N = 406$						

distribution, and the right-hand portion of realized frequencies is consistently below the uniform distribution.

Table VII analyses the entire sample of ETFs and then considers two subsamples: 2007–11 and 2012–16. For each sample of data, we report the percentages of ETFs falling below 1%, 5%, and 10% of the observations in their simulated distribution. In the 2007–16 sample, both the IRR and the SGAR exhibit negative skew. 1.96% of ETFs (8 of 412) have an IRR that is smaller than 1% of the simulated IRRs in their respective distributions and 3.16% of ETFs (13 of 412) have an SGAR that is smaller than 1% of the simulated SGARs in their respective distributions. To put these numbers in perspective, the probabilities of 8 and 13 observations (or more) falling below a 1% threshold with 412 draws from a uniform distribution are 5.77% and 0.03%.[38] The percentage of ETFs with IRRs in the 5% and 10% left tails are 7.09% (by chance, this would occur with probability 4.23%) and 14.18% (by chance, this would occur with probability 0.52%). For SGARs, the percentages of ETFs in the 5% and 10% left tails are 11.41% (by chance, <0.01%) and 17.96% (by chance, <0.01%).

In the 2007–11 sample, the IRR skew is less pronounced as compared to the full sample. The percentage of ETF IRR p-values falling in the left tail are 2.19% (by chance, 11.26%), 6.56% (by chance, 20.70%), and 13.11% (by chance, 10.30%) for the 1%, 5%, and 10% thresholds, respectively. The SGAR skew, however, is more pronounced in the 2007–11 sample as compared to the full sample. The percentage of ETF SGAR p-values falling in the left tail are 3.83% (by chance, 0.26%), 11.48% (by chance, 0.04%), and 22.40% (by chance, <0.01%) for the three threshold values, respectively. The 2012–16 sample exhibits greater IRR distribution skew and less SGAR distribution skew as compared to the full

38 The probabilities are computed under the assumption that each ETF's p-value is i.i.d. and drawn from a uniform distribution.

sample. The percentage of ETF IRR p-values falling in the left tail are 4.43% (by chance, <0.01%), 14.29% (by chance, <0.01%), and 23.65% (by chance, <0.01%), and the percentage of ETF SGAR p-values falling in the left tail are 0.74% (by chance, 77.20%), 7.14% (by chance, 3.61%), and 13.05% (by chance, 2.78%).

The results in Figure 3 and Table VII show that realized ETF IRRs and SGARs tend to have a negative tilt relative to the expectation under the null hypothesis, implying that ETF investors underperform over time. The IRR analysis suggests that ETF investors collectively underperform the performance on a single share. As an anecdotal example, consider State Street Global Advisor's fund SPY which is the largest ETF in the world and represented $226 billion in assets at the end of 2016. ETF investors' IRR in SPY underperformed the return on a single share by 37 bps per year. The SGAR analysis also suggests that ETF investors mistime their asset allocation decisions between risky and safe assets. SPY's SGAR of 5.43% differs from its simulated expected SGAR of 6.91% by 148 bps per year. Thus, while SPY's management fee is 9 bps annually, our IRR and SGAR analyses suggest that non-fundamental demand imposes indirect costs on investors that are an order of magnitude larger than the ETF's management fee.[39]

6. Conclusion

We show theoretically and empirically that ETF flows contain a strong signal of non-fundamental demand. ETFs with large inflows predictably earn lower future returns than ETFs with large outflows. Moreover, we find that non-fundamental demand causes ETF investors to systematically mistime their investments, leading to a reduction in their profits. Our findings suggest several areas of future research. By showing that non-fundamental demand shocks have different effects at the 1-, 3-, and 6-month horizons, our results suggest that there may be additional variation in non-fundamental demand that impacts asset prices at different horizons. Put differently, future research should investigate the term structure of non-fundamental demand shocks and the speed at which they reverse. More generally, our analysis shows that ETFs are a novel laboratory for studying the impact of non-fundamental demand shocks.

Supplementary Material

Supplementary data are available at *Review of Finance* online.

Appendix

Proof of Lemma 1. Each authorized participant i's optimization is given explicitly by,

$$\max_{\delta \in \Re} \delta_i \left(\left(\beta + \Omega_0 + \nu_2^{etf} - \eta(q_0 + \delta_i + \delta_{-i}) \right) - \left(\Omega_0 + \nu_2^{nav} + \lambda(\delta_i + \delta_{-i}) \right) \right), \tag{A.1}$$

in which δ_{-i} is the total redemption/creation activity for all other $N - 1$ authorized participants and,

39 Our results do not say that the ETF mechanism leads to underperformance. Rather, the ETF mechanism provides the means to measure the costs of non-fundamental demand.

$$\nu_2^{etf} = \omega_1 + \epsilon^{etf}, \tag{A.2}$$

$$\nu_2^{nav} = \omega_1 + \epsilon^{nav}, \tag{A.3}$$

since the authorized participants cannot disentangle the fundamental shock from the non-fundamental shocks. The optimization in Equation (A.1) simplifies to,

$$\max_{\delta \in \Re} \delta_i \Big(\Big(\nu_2^{etf} - \eta(\delta_i + \delta_{-i}) \Big) - \Big(\nu_2^{nav} + \lambda(\delta_i + \delta_{-i}) \Big) \Big), \tag{A.4}$$

and the first-order condition with respect to δ_i yields,

$$0 = \Big(\Big(\nu_2^{etf} - \eta(\delta_i + \delta_{-i}) \Big) - \Big(\nu_2^{nav} + \lambda(\delta_i + \delta_{-i}) \Big) \Big) + \delta_i(-\eta - \lambda). \tag{A.5}$$

Solving for δ_i gives,

$$\delta_i = \frac{\nu_2^{etf} - \nu_2^{nav} - \delta_{-i}(\lambda + \eta)}{2(\lambda + \eta)}. \tag{A.6}$$

All authorized participants face the same optimization problem and their equilibrium choice δ^* is implicitly given by,

$$\delta^* = \frac{\nu_2^{etf} - \nu_2^{nav} - (N-1)\delta^*(\lambda + \eta)}{2(\lambda + \eta)}, \tag{A.7}$$

and explicitly given by,

$$\delta^* = \frac{\nu_2^{etf} - \nu_2^{nav}}{(N+1)(\lambda + \eta)} \tag{A.8}$$

$$= \frac{\epsilon^{etf} - \epsilon^{nav}}{(N+1)(\lambda + \eta)}. \tag{A.9}$$

Furthermore, the aggregate ETF flow is given by,

$$\Delta^* = N\delta^* \tag{A.10}$$

$$= \frac{N(\epsilon^{etf} - \epsilon^{nav})}{(N+1)(\lambda + \eta)}. \tag{A.11}$$

The limit of Δ^*as $N \to \infty$is given by,

$$\lim_{N \to \infty} \Delta^* = \frac{\epsilon^{etf} - \epsilon^{nav}}{\lambda + \eta}. \tag{A.12}$$

Proof of Lemma 2. The proof follows immediately from taking the variance of $\lim_{N \to \infty} \Delta^*$.

Proof of Proposition 1. The equilibrium ETF share price is given by,

$$p^* = \beta + \Omega_2 - \eta(q_0 + \Delta^*) + \epsilon^{etf}, \tag{A13}$$

which simplifies to,

$$p^* = \Omega_0 + \omega_1 + \epsilon^{etf} - \eta\left(\frac{N(\epsilon^{etf} - \epsilon^{nav})}{(N+1)(\lambda+\eta)}\right). \tag{A14}$$

The equilibrium NAV price is given by,

$$\pi^* = \Omega_2 + \epsilon^{nav} + \lambda\Delta_1, \tag{A15}$$

which simplifies to,

$$\pi^* = \Omega_0 + \omega_1 + \epsilon^{nav} + \lambda\left(\frac{N(\epsilon^{etf} - \epsilon^{nav})}{(N+1)(\lambda+\eta)}\right). \tag{A16}$$

Using the equilibrium ETF share price and NAV price, the $t=2$ equilibrium ETF premium is given by,

$$\psi^* = \left(1 - \frac{N}{N+1}\right)(\epsilon^{etf} - \epsilon^{nav}), \tag{A17}$$

the $t=2$ equilibrium ETF fundamental mispricing is given by,

$$\varphi^* = \epsilon^{etf} - \eta\left(\frac{N(\epsilon^{etf} - \epsilon^{nav})}{(N+1)(\lambda+\eta)}\right), \tag{A18}$$

and the $t=2$ equilibrium NAV fundamental mispricing is given by,

$$\alpha^* = \epsilon^{nav} + \lambda\left(\frac{N(\epsilon^{etf} - \epsilon^{nav})}{(N+1)(\lambda+\eta)}\right). \tag{A19}$$

Taking the limits of ψ^*, φ^*, and α^*as $N \to \infty$ yields,

$$\lim_{N\to\infty} \psi^* = 0, \tag{A20}$$

$$\lim_{N\to\infty} \varphi^* = \epsilon^{etf}\frac{\lambda}{\lambda+\eta} + \epsilon^{nav}\frac{\eta}{\lambda+\eta}, \tag{A21}$$

$$\lim_{N\to\infty} \alpha^* = \epsilon^{etf}\frac{\lambda}{\lambda+\eta} + \epsilon^{nav}\frac{\eta}{\lambda+\eta}. \tag{A22}$$

References

Agarwal, V., Hanouna, P., Moussawi, R., and Stahel, C. (2019): Do ETFs increase the commonality in liquidity of underlying stocks?. Working Paper, Georgia State University.

Baker, M. and Wurgler, J. (2000): The equity share in new issues and aggregate stock returns, *Journal of Finance* 55, 2219–2257.

Baker, M. and Wurgler, J. (2006): Investor sentiment and the cross-section of stock returns, *Journal of Finance* 61, 1645–1680.

Baker, M. and Wurgler, J. (2007): Investor sentiment in the stock market, *Journal of Economic Perspectives* 21, 129–152.

Baltussen, G., van Bekkum, S., and Da, Z. (2019): Indexing and stock market serial dependence around the world, *Journal of Financial Economics* 132, 26–48.

Barberis, N. and Thaler, R. (2003): A survey of behavioral finance, in: Constantinides. G. M, Harris. M, and Stulz R. M. (eds.), *Handbook of the Economics of Finance*, Vol. 1, Amsterdam, Elsevier, pp. 1053–1128.

Ben-David, I., Franzoni, F., and Moussawi, R. (2012): Hedge fund stock trading in the financial crisis of 2007–2009, *Review of Financial Studies* 25, 1–54.

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

Ben-Rephael, A., Kandel, S., and Wohl, A. (2012): Measuring investor sentiment with mutual fund flows, *Journal of Financial Economics* 104, 363–382.

Berk, J. B. and Green, R. C. (2004): Mutual fund flows and performance in rational markets, *Journal of Political Economy* 112, 1269–1295.

Berk, J. B. and Stanton, R. (2007): Managerial ability, compensation, and the closed-end fund discount, *Journal of Finance* 62, 529–556.

Bessembinder, H. (2015): Predictable ETF order flows and market quality, *The Journal of Trading* 10, 17–23.

Black, F. (1986): Noise, *Journal of Finance* 41, 528–543.

Brown, D. C. and Davies, S. W. (2017): Moral hazard in active asset management, *Journal of Financial Economics* 125, 311–325.

Brown, D. P. and Wu, Y. (2016): Mutual fund flows and cross-fund learning within families, *Journal of Finance* 71, 383–424.

Carhart, M. M. (1997): On persistence in mutual fund performance, *Journal of Finance* 52, 57–82.

Cherkes, M., Sagi, J., and Stanton, R. (2008): A liquidity-based theory of closed-end funds, *The Review of Financial Studies* 22, 257–297.

Chevalier, J. and Ellison, G. (1997): Risk taking by mutual funds as a response to incentives, *Journal of Political Economy* 105, 1167–1200.

Cooper, M. J., Gulen, H., and Rau, P. R. (2005): Changing names with style: mutual fund name changes and their effects on fund flows, *Journal of Finance* 60, 2825–2859.

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

Daniel, K. and Titman, S. (2006): Market reactions to tangible and intangible information, *Journal of Finance* 61, 1605–1643.

Daniel, K. D., Hirshleifer, D., and Subrahmanyam, A. (2001): Overconfidence, arbitrage, and equilibrium asset pricing, *Journal of Finance* 56, 921–965.

Dannhauser, C. (2017): The impact of innovation: evidence from corporate bond exchange-traded funds (ETFs), *Journal of Financial Economics* 3, 537–560.

Dannhauser, C. and Hoseinzade, S. (2017): The transformation of corporate bond investors and fragility: Evidence on mutual funds and ETFs. Working Paper, Villanova University.

Dannhauser, C. and Pontiff, J. (2019): Flow. Working Paper, Villanova University.

Davies, S. W. (2020): Speculation sentiment. Working Paper, University of Colorado Boulder.

De Long, J. B., Shleifer, A., Summers, L. H., and Waldmann, R. J. (1990): Noise trader risk in financial markets, *Journal of Political Economy* 98, 703–738.

Dichev, I. D. (2007): What are stock investors' actual historical returns? Evidence from dollar-weighted returns, *American Economic Review* 97, 386–401.

Dichev, I. D. and Yu, G. (2011): Higher risk, lower returns: what hedge fund investors really earn, *Journal of Financial Economics* 100, 248–263.

Driscoll, J. C. and Kraay, A. C. (1998): Consistent covariance matrix estimation with spatially dependent panel data, *Review of Economics and Statistics* 80, 549–560.

Duffie, D. (2010): Presidential address: asset price dynamics with slow-moving capital, *Journal of Finance* 65, 1237–1267.

Evans, R. B., Moussawi, R., Pagano, M. S., and Sedunov, J. (2019): ETF failures-to-deliver: naked short-selling or operational shorting?. Working Paper, University of Virginia.

Fama, E. F. and French, K. R. (1993): Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Fama, E. F. and French, K. R. (2008): Dissecting anomalies, *Journal of Finance* 63, 1653–1678.

Fama, E. F. and French, K. R. (2015): A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.

Frazzini, A. and Lamont, O. (2008): Dumb money: mutual fund flows and the cross-section of stock returns, *Journal of Financial Economics* 88, 299–322.

Friesen, G. C. and Sapp, T. R. A. (2007): Mutual fund flows and investor returns: an empirical investigation of fund investor timing ability, *Journal of Banking and Finance* 31, 2796–2816.

Fulkerson, J. and Jordan, B. D. (2013): Reading tomorrow's newspaper: predictability in ETF returns, *The Journal of Index Investing* 4, 23–31.

Glosten, L. R., Nallareddy, S., and Zou, Y. (2020): ETF trading and informational efficiency of underlying securities, *Management Science*, doi:10.1287/mnsc.2019.3427.

Hayley, S. (2014): Hindsight effects in dollar-weighted returns, *Journal of Financial and Quantitative Analysis* 49, 249–269.

He, Z., Kelly, B., and Manela, A. (2017): Intermediary asset pricing: new evidence from many asset classes, *Journal of Financial Economics* 126, 1–35.

Hirshleifer, D. (2001): Investor Psychology and asset pricing, *Journal of Finance* 56, 1533–1597.

Hou, K., Xue, C., and Zhang, L. (2015): Digesting anomalies: an investment approach, *Review of Financial Studies* 28, 650–705.

Hsu, J., Myers, B. W., and Whitby, R. (2016): Timing poorly: a guide to generating poor returns while investing in successful strategies, *The Journal of Portfolio Management* 42, 90–98.

Ikenberry, D., Lakonishok, J., and Vermaelen, T. (1995): Market underreaction to open market share repurchases, *Journal of Financial Economics* 39, 1765–1790.

Ippolito, R. A. (1992): Consumer reaction to measures of poor quality: evidence from the mutual fund industry, *The Journal of Law and Economics* 35, 45–70.

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

Jiang, W. and Yan, H. (2020): Financial innovation, investor behavior, and arbitrage: implications from the ETF market. Working Paper, The Chinese University of Hong Kong.

Kamstra, M. J., Kramer, L. A., Levi, M. D., and Wermers, R. (2017): Seasonal asset allocation: evidence from mutual fund flows, *Journal of Financial and Quantitative Analysis* 52, 71–109.

Krause, T., Ehsani, S., and Lien, D. (2014): Exchange traded funds, liquidity, and volatility, *Applied Financial Economics* 24, 1617–1630.

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

Lee, C., Shleifer, A., and Thaler, R. H. (1991): Investor sentiment and the closed-end fund puzzle, *Journal of Finance* 46, 75–109.

Lorenzoni, G. (2008): Inefficient credit booms, *Review of Economic Studies* 75, 809–833.

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

Loughran, T. and Ritter, J. R. (1995): The new issues puzzle, *Journal of Finance* 50, 23–51.

Loughran, T. and Vijh, A. M. (1997): Do long-term shareholders benefit from corporate acquisitions?, *Journal of Finance* 52, 1765–1790.

Madhavan, A. and Sobczyk, A. (2016): Price dynamics and liquidity of exchange-traded funds, *Journal of Investment Management* 14, 86–102.

Marshall, B. R., Nguyen, N. H., and Visaltanachoti, N. (2013): ETF arbitrage: intraday evidence, *Journal of Banking and Finance* 37, 3486–3498.

Pan, K. and Zeng, Y. (2019): ETF Arbitrage under liquidity mismatch. Working Paper, Harvard University.

Pástor, L. and Stambaugh, R. F. (2012): On the size of the active management industry, *Journal of Political Economy* 120, 740–781.

Politis, D. N. and Romano, J. P. (1994): The stationary bootstrap, *Journal of the American Statistical Association* 89, 1303–1313.

Pontiff, J. (1996): Costly arbitrage: evidence from closed-end funds, *The Quarterly Journal of Economics* 111, 1135–1151.

Pontiff, J. and Woodgate, A. (2008): Shares issuance and cross-sectional returns, *Journal of Finance* 43, 921–945.

Sapp, T. and Tiwari, A. (2004): Does stock return momentum explain the "smart money" effect?, *Journal of Finance* 59, 2605–2622.

Shleifer, A. and Vishny, R. (2011): Fire sales in finance and macroeconomics, *Journal of Economic Perspectives* 25, 29–48.

Sirri, E. and Tufano, P. (1998): Costly search and mutual fund flows, *Journal of Finance* 53, 1589–1622.

Staer, A. (2016): Fund flows and underlying returns: the case of ETFs, *International Journal of Business* 22, 275–304.

Staer, A. and Sottile, P. (2018): Equivalent volume and comovement, *The Quarterly Review of Economics and Finance* 68, 143–157.

Stambaugh, R. F. (1999): Predictive regressions, *Journal of Financial Economics* 54, 375–421.

Warther, V. A. (1995): Aggregate mutual fund flows and security returns, *Journal of Financial Economics* 39, 209–235.