ETF Short Interest and Failures-to-Deliver:
Naked Short-selling or Operational Shorting?

Richard B. Evans
University of Virginia

Rabih Moussawi
Michael S. Pagano
John Sedunov
Villanova University

First Draft: March, 2016
This Draft: November, 2017

Abstract

While ETFs constitute just under 10% of U.S. equity market capitalization, they account for over 20% of short interest and over 78% of failures-to-deliver in U.S. equities. While the disproportionate share of short activity in ETFs has raised concerns about excessive shorting and naked short-selling, we identify an alternative cause for this activity related to the market making activities associated with the ETF creation/redemption process, which we label "operational shorting." We propose a simple methodology to estimate operational shorting and show that our measure is consistent with the economics behind the mechanism. In examining the market implications of operational shorting, we find that it is associated with improved liquidity but that it is also predictive of market-wide indicators of systemic and counterparty risk. In exploring possible mechanisms for this predictive relationship, we find there is commonality in operational shorting across ETFs that have the same lead market maker/authorized participant and that market makers' financial leverage might be a channel that amplifies this commonality, both of which are suggestive of an increase in counterparty risk.

Keywords: Exchange-Traded Funds, Failure to Deliver, Financial Markets, Short Selling, Short Interest, Counterparty Risk, Equities, Investments

JEL Codes: G1, G12, G14, G23

We are thankful to Aleksandar Andonov, Francois Cocquemas, Austin Gerig, Wes Gray, Bryan Johanson, James Simpson, Jack Vogel, as well as seminar participants at the FINRA-Columbia Market Structure Conference, 5th Luxembourg Asset Management Summit, and the Southern Finance Association and seminar participants at the Federal Reserve Board, Loyola University of Maryland, Penn State – Harrisburg, University of Georgia, and University of Virginia, and in addition to participants at the meeting for helpful comments and feedback. We also greatly appreciate the capable research assistance of Alejandro Cuevas, Shreya Rajbhandary, and Austin Ryback.

1. Introduction

With over $2.5 trillion invested worldwide[1], exchange traded funds (ETFs) are a financial innovation that has been embraced by investors. In addition to providing a low-cost way to obtain long exposure to different asset classes, ETFs also offer investors a simple way to gain short exposure. Due to their hybrid nature, ETF shares can be borrowed and sold short. Figure 1 shows that as ETFs have grown, so has short-selling activity in ETFs. At the end of 2016, the aggregate dollar value of ETF short interest was upwards of $150 billion, accounting for 20% of the overall dollar value of short interest in U.S. equity markets, even though ETFs constituted just under 10% of total U.S. equity market capitalization.

While the disproportionate share of ETF short interest relative to ETF market capitalization may simply indicate that investors find short-selling ETFs more compelling than short-selling individual equities, there is concern among regulators and market participants that this significant short-selling activity may also be an indication of naked or abusive short-selling practices. Recent enforcement actions against authorized participants[2] by FINRA and Nasdaq underscore this concern about the improper short-selling of ETFs.[3] While these actions are annecdotal, one signal of naked short-selling is the incidence of failures-to-deliver (hereafter, FTDs).[4] Using equity FTDs as a point of comparison, Figure 2 shows the aggregate

[1] U.S.-registered investment companies include all mutual funds, ETFs, closed-end funds and unit investment trusts registered under the Investment Company Act of 1940. The market share of ETFs was calculated using the "2016 Facts at a Glance" from page ii of the 2017 Investment Company Fact Book published by the Investment Company Institute.

[2] In March of 2016, FINRA and Nasdaq fined Wedbush Securities, an ETF authorized participant, for submitting "naked" ETF redemption orders on behalf of a broker/dealer client, Scout Trading, in a number of levered ETFs. If Scout Trading wanted to profit from the well-documented price decline/decay of these leveraged ETFs (i.e. Zhang and Judge, 2016), but was unable or unwilling to borrow shares due to short selling constraints, one way to achieve short exposure would be to redeem or sell shares they did not own ("naked" redemption/short-selling), and subsequently fail-to-deliver (FTD) those shares to Wedbush.

[3] Thomas Gira, the FINRA Head of Market Regulation, explains, the regulatory concern of interest is "naked" short-selling of ETFs: "Timely delivery of securities is a critical component of sales activity in the markets, particularly in ETFs that rely on the creation and redemption process. Naked trading strategies that result in a pattern of systemic and recurring fails flout such principles and do not comply with Regulation SHO. Authorized Participants and their broker-dealer clients need to have adequate supervisory procedures and controls in place to ensure that they are properly redeeming and creating shares of ETFs." FINRA News Release, "FINRA and Nasdaq Fine Wedbush Securities Inc. $675,000 For Supervisory Violations Relating to Chronic Fails to Deliver by a Client in Multiple Exchange-Traded Funds", 3/21/2016, accessed 6/2/2017 at http://www.finra.org/newsroom/2016/finra-and-nasdaq-fine-wedbush-securities-inc-675000-supervisory-violations-relating.

[4] Stratmann and Welborn (2013) describe FTDs as "electronic IOUs" where a market participant who has engaged in a short sale does not deliver the underlying security at the time of settlement (typically 3 days after the sale in the U.S., and referred to as "T+3" in the parlance of securities trading and settlement).

daily dollar volume of FTDs over time. During the 2008 financial crisis, SEC Rules 203 and 204 were introduced, in part, to address naked short-selling and the associated FTDs.[5] While Figure 2 shows that these rule changes led to a dramatic decline in both stock and ETF FTDs in early 2009 and a relatively low level and negligible growth of stock-related FTDs during 2010-2016, there is an upward trend in the dollar volume of ETF FTDs over the past seven years. At the end of 2016, ETF FTDs accounted for over 78% of all equity FTDs.

Whether the issue is excessive short-selling, naked short-selling, or both, the high levels of ETF short interest and FTDs are concerning. These concerns are based on the idea that all of the observed short interest and FTDs arise from "directional shorting" or investors attempting to benefit from a negative directional move in ETF prices or returns. At the same time, there is an alternative mechanism, unique to ETFs, whereby market making activities associated with the creation/redemption process could generate short-selling and FTDs. This mechanism, which we call "operational shorting", is described as follows:

> "Market makers, often commercial banks or hedge funds, create ETFs for their issuers by buying the securities that the funds are supposed to represent. But they've discovered that they can make a predictable return by delaying the purchases and selling you nonexistent exchange-traded fund shares that they will create later. These transactions—a form of shorting—eventually may involve 50,000 shares—the amount typically in a "creation unit" authorized by the issuer..."[6]

Under prevailing market making rules, an authorized participant (AP) / lead market maker (MM) (hereafter AP)[7] can sell new ETF shares to satisfy bullish order imbalance, but can opt to delay the physical share

[5] Jain and Jain (2015)

[6] Jim McTague, "Market Maker's Edge: T+6", *Barron's*, 12/24/2011, accessed online 10/4/16 at http://www.barrons.com/articles/SB50001424052748703679304577108520307148702.

[7] Section 2 provides more information about the roles and responsibilities of ETF Market Makers and Authorized Participants. Antoniewicz and Heinrichs (2015) finds an average ETF has 34 authorized participants, out of which only 5 APs are active (have at least one create/redeem order over a period of 6 months), and 5 APs, on average, are registered market makers, with obligations to provide continuous buy and sell quotes for ETF shares on secondary markets. We follow the findings in Antoniewicz and Heinrichs (2015) and assume that the active ETF authorized participants have also market making capacity, and we refer to them interchangeably in our paper as AP or MM.

creation (purchasing the basket of underlying securities and swapping that basket for the corresponding number of ETF shares) until a future date. There are a number of operational reasons why an AP might want to delay creation. First, ETF creation is done in discrete blocks of ETF shares called creation units (typically 50,000 ETF shares). If the order imbalance is smaller than the creation unit size, APs may wait until the the imbalance builds to size equal to or greater than the creation unit. Second, if the underlying basket of securities is less liquid than the ETF itself and purchasing the stocks to form the creation basket incurs price impact and liquidity costs, during the time that creation is delayed, order flow might reverse. This reversal would enable the AP to earn the ETF bid-ask spread, without paying the trading costs associated with buying the basket of underlying securities. Both of these motivations become even more compelling if a cheap hedge is available through the futures and options markets.[8] The motivation for these 'operational' short positions stands in stark contrast to informed, 'directional' shorting, that has been the primary focus of the short-selling literature (i.e. Senchack and Starks, 1993, Asquith, Pathak and Ritter, 2005, and Boehmer, Jones, and Zhang, 2008).

We propose a simple and novel methodology to estimate the operational shorting of ETFs and show that the estimate is consistent with the economics behind the proposed mechanism. The description of the above activity suggests that operational shorting occurs when new ETF shares are purchased by investors but there is a delay in the creation of those shares by the AP. To measure operational shorting, we use: a) the buy-sell imbalance (measured using signed intra-daily trade data) of a given ETF to proxy for the purchase of new ETF shares by investors and b) changes in the daily shares outstanding of the ETF to proxy for the delayed, or non-contemporaneous, net share creation activity. If the buy-sell imbalance is positive at a given point in time but there is no contemporaneous creation of the ETF shares, the AP is operationally short those shares because they have yet to create and deliver them to investors. Figure 3 presents a daily timeline that depicts the evolution of an operational short position for an AP. This timeline demonstrates how the rules related to "bona fide market making" can extend the actual delivery of the ETF shares for

[8] In the appendix, we work through a numerical example of the value of this "option to delay" the creation and delivery of ETF shares that have already been sold.

several days past traditional T+3 settlement.[9]

With our measure of operational shorting, we first examine the relationship between operational shorting and FTDs. Figure 4 plots the aggregate dollar value of operational shorting and FTDs across all ETFs. Comparing the two time series, we see that there is a strong positive correlation between the two, consistent with operational shorting playing an important role in ETF FTDs. Repeating the analysis at the ETF level and controlling for the other potential determinants, we confirm this strong statistically and economically significant relationship. The result is especially striking given that our operational shorting measure only identifies cases where there is excess demand for ETF shares (i.e., there is a buy imbalance that is greater than the number of shares created). Our evidence on AP incentives to delay creation that give rise to operational shorting support our conjectures that arbitrage profits represent a major incentive that draw APs into ETF market making activities. Our results suggest that satisfying order flow imbalances with new create orders is not instantaneous and create orders typically lag actual order imbalances by days creating higher operational shorting. More opeartional shorting is found to be driven by higher liquidity mismatch with underlying basket and in the presence of efficient hedges. Our results provide important support and justifications that incentivize APs to wait and delay the assembly of the basket and creation of new ETF shares until a future date.

As a separate test of the underlying economics behind operational shorting, we also examine its predictive power for future risk-adjusted returns for the subset of U.S. equity ETFs. There is a long literature documenting that short-selling activity is predictive of future underperformance, consistent with a "directional" motive for informed investors to short sell. Unlike other measures of short-selling demand for an ETF (i.e. short interest or lending fees), which are negatively related to future returns, operational shorting is unrelated to the future return on the ETF, consistent with the underlying economics of liquidity provision by APs. In this case, we expect no relationship between operational shorting and future ETF returns because these short sales are driven by the liquidity provision activities of APs and are not "taking

[9] On September 5, 2017, a shortened t+2 settlement cycle was implemented for most securities. Bona-fide market making transactions have 3 additional days (until t+5). For more information, see: https://www.sec.gov/tm/t2-sbrefa.

a view" on the direction of the ETF's market value. This finding has important implications for the extant short selling literature because it underscores the need to account for the different motivations behind ETF short selling: directional / informational vs. operational shorting/liquidity provision. While previous research has shown that common stock short interest is an important predictor of aggregate stock returns consistent with a primarily directional motivation for short selling (i.e. Rapach, Ringgenberg, and Zhou, 2016), we document that operational shorting is one of the most significant drivers of an ETF's short interest.

After establishing how operational shorting results from the liquidity provision in the ETF share market, we turn our attention to the underlying constituents basket, and we examine the impact of such "operational shorting" on common stocks that are held by those ETFs. The exception to Rule 204 for market makers is granted only when the operational short is "attributable to bona fide market making activities." Given that important caveat, examining the impact of operational shorting on ETF liquidity is an important verification that indeed these short-sales represent legitimate market making activities. Fotak, Raman, and Yadav (2014)[10], along with Merrick, Naik, and Yadav (2005), argue that FTDs can serve as an "important release valve" that removes any binding constraints on market participants' ability to supply liquidity and perform valuable arbitrage activities.[11] We run a similar analysis of the impact of operational shorting on the liquidity of underlying stocks by examining the relation between ETF operational shorting activities, stock volatility and best bid and offer spreads on an intraday basis. Consistent with Ben-David, Franzoni and Moussawi (2015) and recent literature,[12] we find that ETF ownership is positively associated

[10] Fotak, Raman, and Yadav (2014) examined stock FTDs during 2005-2008 and found that increased levels of stock-related FTDs led to improved market quality in terms of reduced pricing errors, as well as lower levels of intraday volatility, bid-ask spreads, and order imbalances. In addition, they find that FTDs during the 2008 financial crisis did not distort prices.

[11] This notion of a "release valve" is also supported in terms of short selling activity's impact on loosening institutional constraints and sharpening price discovery. For example, Chu, Hirshleifer, and Ma (2016) show that the introduction of Regulation SHO (which reduced short selling constraints) has led to a reduction in returns to asset pricing anomalies. The authors suggest that this increase in short selling ability has made arbitrage of asset pricing anomalies easier and thus has decreased the returns to these strategies. In effect, like FTDs, Regulation SHO acted as another form of release valve which can lead to increased market efficiency.

[12] For example, Da and Shive (2014), Hamm (2014), Sullivan and Xiong (2012), Chinco and Fos (2016), Bhattacharya and O'Hara (2016), Dannhauser (2016), and Israeli, Lee, and Sridharan (2017). See Ben-David, Franzoni, and Moussawi (2017) for a survey of ETF literature.

with higher volatility and intraday spreads. However, we also show that operational shorting is negatively related to intraday spreads and volatility, thus acting as a "release valve". As operational shorting increases due to a sudden surge in buying demand, the APs seem to provide liquidity in the ETF market without (or before) entering the market for the underlying stocks. Therefore, our evidence suggests that operational shorting serves as a buffer that reduces the transmission of the ETF liquidity shocks to underlying stocks, especially when higher frequency investors are increasingly attracted to ETFs due to their incremental liquidity (Ben-David, Franzoni and Moussawi, 2015).

While operational shorting does improve underlying stock liquidity, there remains a concern that FTDs impact financial stability through a commonality in the liquidity provision activities of an inter-connected network of ETF market makers. This mechanism may be important, because ETFs, as hybrid investment vehicles, form an essential nexus between several areas of the financial system. Thus, a shock to an ETF can be reflected in its FTD activity which, in turn, leads to potentially destabilizing effects on other parts of the financial market. In a 2011 report, the Financial Stability Board (FSB) raised concerns about ETFs and their potential impact on financial markets because the size and complexity of the ETF market could increase both counterparty risk and systemic risk.[13] The FSB report noted "the expectation of on-demand liquidity may create the conditions for acute redemption pressures on certain types of ETFs in situations of market stress." The unique redemption / creation process of ETFs, as well as the risks of trading, clearing, and settling these securities, are different than those present in the equity markets.[14]

To test for evidence of this broad concern about financial stability, we aggregate operational shorting over time and examine its relationship with the St. Louis Federal Reserve Financial Stress Index

[13] As financial crises involving U.S. and European financial institutions in recent years have shown, problems in one market can quickly create negative "spillover" or "contagion" effects to financial institutions that were not thought to be closely related. These spillover effects can lead to sudden, sharp spikes in a financial system's overall risk, commonly referred to as systemic risk. To the extent that ETFs can also employ financial leverage and derivatives, one can see that ETFs are at the nexus of the markets for cash equities, options, futures, credit, and securities lending. Thus, shocks to any of these markets can affect many other areas of the financial system via their linkages to ETFs and the institutions that serve as ETF market makers.

[14] For research on the FTDs of U.S. equities, see Boni (2006); Stratmann and Welborn (2013); Fotak, Raman, and Yadav (2014); Autore, Boulton, and Braga-Alves (2015); and Jain and Jain (2015). For FTDs in option markets and linkages to common stocks, see Evans, Geczy, Musto, and Reed (2009); Battalio and Schultz (2011); and Stratmann and Welborn (2013).

(FSI). As ETF FTDs have increased in aggregate over time, we find that they have become more closely related to financial system stability. Before the 2008 SEC rule change, an increase in stock FTDs was predictive of a rise in the FSI, but ETF FTDs were not. After the change, however, we find that stock FTDs no longer relate to FSI, but that ETF FTDs strongly relate. We also aggregate our measure of operational shorting across ETFs and show that it too is related to the FSI.

While there are a number of potential channels through which ETF FTDs could relate to financial stress, Malamud (2015) models ETF liquidity provision and proposes one such channel. He shows that the creation and redemption mechanism in the ETF markets can serve as a "shock propagation channel through which temporary demand shocks may have long-lasting impacts on future prices." Additionally, Malamud (2015) notes that ETF liquidity providers are fundamentally different because they typically play a dual role not only as ETF market makers but also as arbitrageurs between the market for ETF shares and the market for the ETF's basket of underlying securities.

To assess whether or not this channel plays a role, we first identify all of the different ETFs served by a given lead market maker[15]. We then explore whether or not operational shorting or FTDs in a given ETF could lead to decreased liquidity and increased FTDs in other ETFs that the same market maker participates in. In addition, most ETFs have more than one market maker and thus a sudden spike in FTDs (coinciding with a drop in liquidity) by, say, a "lead" market maker in one ETF can spill over to other liquidity providers if these firms make markets in a common set of ETFs. This, in turn, can quickly create a ripple effect throughout the entire ETF market and consequently increase counterparty risk and system-wide stress not only with ETFs but also with ETF-related common stocks and derivatives. Indeed, we find that increases in FTDs and operational shorting for one ETF are affected by the operational shorting of other ETFs that are traded by the same lead market maker. Thus, there appears to be commonality in operational shorting and FTDs across ETF market makers which suggests that such a contagion

[15] Because the authorized participants for a given ETF are not reported in public sources, we use the lead market maker as our proxy for the authorized participant. Antoniewicz and Heinrichs (2015) report similar numbers of active APs and APs registered as market makers, suggestive that lead market maker would be a viable proxy for an active AP.

phenomenon may play a role in explaining the observed relationship between operational shorting and systemic risk. Further, we examine a channel by which FTDs and operational shorting are associated with financial stability: leverage. We find that the financial leverage of lead market makers is positively related to both of our key dependent variables. This finding is consistent with our earlier observations and an AP's business strategy of increasing its return on equity by economizing not only on creation fees and trading costs but also capital requirements. Although this strategy might be profitable at the level of an individual AP, it can also lead to a more highly levered and inter-connected ETF market that is vulnerable to financial stress on an aggregate basis.

The remainder of the paper is organized as follows. Section two motivates and defines the empirical models used in our analysis. Section three describes the data, while section four presents our results, and section five concludes.

2. ETF Market Making and Fails-to-Deliver

2.1 The Mechanics of ETF Trading and Market Making

ETFs can be viewed as a hybrid investment vehicle with some of the attributes of both a mutual fund and an exchange-traded common stock.[16] Madhavan (2014) describes ETFs as more than "exchange-traded versions of index mutual funds," as they have a mixture of elements related to both open-end and closed-end mutual funds, as well as the ability to be traded intraday and engage in "in-kind" securities transfers that have tax advantages for investors.[17] Similar to stocks and closed end funds, ETF shares trade on exchanges, and such secondary market trading is attributed to constitute the majority of ETF trading activity. ETF market makers ensure the liquidity of ETF trading in secondary markets by assuming

[16] This section draws largely from Madhavan (2014), Antoniewicz and Heinrichs (2014), and Evans et al. (2009).

[17] Antoniewicz and Heinrichs (2014) note that ETFs can use in-kind redemptions by redeeming "low basis" securities for purchases of new securities to reduce unrealized capital gains. In effect, ETF investors can defer most of their capital gains until they sell their shares. These authors also indicate that price transparency is another benefit of ETFs due to their ability to be traded within the day at, or near, their net asset value (NAV), as arbitrage between the cash equities and ETF markets typically keep ETF prices quite close to the NAV. ETFs also provide an easy way for investors to gain exposure to a particular sector of the financial market with a well-diversified portfolio that is monitored by professional investment managers. For these reasons, ETFs have continued to grow in popularity with both retail and institutional investors.

obligations to provide continuous bid and ask quotes on ETFs. In instances of buy sell imbalances in the ETF secondary markets or when trading cannot be met with existing shares, ETF market makers work with affiliated APs to create (or redeem) ETF creation units in order to enhance the liquidity of the ETF shares in secondary markets. For this reason, it is not suprising to find that many ETF market makers are registered as authorized participants, in order to facilitate the create/redeem process of of ETF shares. In this paper, we make the assumption that APs are registered market makers, which is consistent with Antoniewicz and Heinrichs' (2015) findings that, out of the 35 APs on an average US equity ETF, there are about 7 active APs and the same number of APs that are registered as market makers.

APs are institutions that have contractual agreements with the ETF sponsor allowing them to trade directly with the ETF sponsor to create and redeem ETF shares in the primary market.[18] For US equity ETFs, such transactions are typically in kind, and a creation basket of securities is exchanged for a creation unit of ETF shares.[19] AP do not receive compensation from the ETF sponsor and have no legal obligation to participate in ETF primary markets. However, they have strong incentives to participate in order to take advantage of the create/redeem process and eliminate price discrepancies by purchasing the cheaper asset on the market and selling the more expensive one. Through such incentives, APs help keep the ETF prices in the secondary market aligned with their intrinsic values. Important features that allow the market price of an ETF to track closely the fund's NAV are the creation / redemption process and the ability to engage in arbitrage. ETF shares are redeemed (in effect, taken out of circulation and thus lowering the supply of shares outstanding) when this process is reversed: the AP delivers a block of ETF shares equivalent to one or more creation units to the ETF investment manager in exchange for the specific basket of cash securities.

[18] An AP is typically a market maker or large institutional investor that has a legal agreement with the ETF to create and redeem shares of the fund. APs do not receive any compensation from the ETF and have no obligation to create or redeem shares of the ETF. Instead, APs earn commissions and fees from customer orders as well as potential profits from ETF-common stock arbitrage. APs must also pay a flat fee for any creation or redemption orders.

[19] As Antoniewicz and Heinrichs (2014) describe, ETF shares are created when an AP places an order for one or multiple creation units, which typically range in size from 25,000 to 200,000 shares. According to Ben-David, Franzoni and Moussawi (2017), 70% of ETFs traded in the US have creation units with blocks of 50,000 ETF shares. The AP can obtain ETF creation units when a pre-specified creation basket of cash securities is delivered to the ETF investment manager (e.g., a market cap-weighted portfolio of stocks that comprise the S&P 500 in exchange for an ETF creation unit).

Note that this redemption process is the result of selling pressures on ETF shares in the open market that can cause a discount in ETF prices relative to NAV, which creates an arbitrage opportunity for APs to redeem ETF shares with the fund for the constituent basket that are worth more in such instances.

2.2 Fails to deliver

An ETF market maker, also an AP, would resort to ETF primary market to settle a order imbalance due to excess demand of ETF shares, that resulted in this AP being net short in ETF shares in the course of providing liquidity and making the market on ETF shares. Their incentive is to harvest the difference between ETF price, pushed higher by excess demand pressures, and the basket NAV at which they construct the creation basket. However, the process of eliminating such short positions due to excess buy order imbalance through the create/redeem process is not instantaneous. In the matter of fact, those ETF APs have incentives to delay the actual creation of ETF, and therefore incur a bona fide “operational short” position for three reasons.

Under prevailing market making rules, the AP sells the new ETF shares to satisfy bullish order imbalance but can opt to delay the physical share creation until a future date, especially if the AP expects at least a partial reversal in the order flow. For underlying basket securities that are less liquid than ETF securities, and might incur steeper price impact and liquidity costs for amassing creation baskets, such an option to delay creation becomes even more valuable, especially in the presence of transient ETF order flows. Additionally, deferring the share creation until future order flows are observed might end up being a more valuable option available to market makers especially when a cheap hedge is available through the futures and options markets.

If an AP delays the creation of ETF shares to satisfy and offset their operational short positions beyond day T+3, a fail to deliver (FTD) position is established. More precisely, an FTD occurs when an AP or other market participant sells ETF shares that it does not already own and then does not deliver those shares to the NSCC within T+3 days. This can happen due to operational shorting, as part of bona fide market making activity, as well as directional shorting, or naked short selling with the purpose of obtaining a negative exposure in the ETF shares in anticipation of a future decline in ETF price. Our tests aim to directly test and distinguish behind those two distinct motivations for ETF FTDs. The NSCC can then force a "buy-in" of an outstanding FTD by typically contacting the market participant with the oldest FTD and requiring them to purchase or borrow the shares in the open market. As Evans et al. (2009) reports, buy-ins are a relatively rare occurrence and the expected cost of failures is relatively low. Thus, there are economic incentives to failing, especially in the ETF market because of the difficulty in distinguishing between FTDs that are due to illegal naked shorting and those FTDs that are due to legal transactions such as ETF redemptions. Thus, there is potential for a market-wide increase in FTDs, particularly in ETFs.

Existing research in the U.S. equities and options markets suggests that FTDs can have both positive and negative effects related to "limits to arbitrage" and "search and bargaining frictions" models. This literature includes Merrick et al. (2005) and Fotak et al. (2014) who argue that a more permissive policy towards FTDs can improve market quality. Additionally, Battalio and Schultz (2011) and Stratmann and Welborn (2013) are aligned with Fotak et al.'s (2014) "release valve" view that FTDs can have positive benefits for the overall market by encouraging traders to supply more liquidity and engage in useful arbitrage activities.

In contrast to the above findings, there are several studies that provide evidence that FTDs may have negative effects. Given the increasing amount of FTDs in ETFs, there is greater potential for these hybrid investment vehicles to perturb financial markets. For example, as noted earlier, Madhavan (2012) and Ben-David et al. (2015) demonstrate that ETFs may have consequences for the volatility of financial markets. Furthermore, in contrast to earlier findings, Stratmann and Wellborn (2016) find that ETF-related FTDs Granger-cause higher stock market volatility and lower future returns which can ultimately lead to increased market instability.

Additionally, Boni (2006) shows that FTDs were pervasive and persistent in U.S. equities during three settlement dates: September 2003, November 2003, and January 2004. This finding is consistent with market makers' incentive to "strategically fail" when borrowing costs are high. Boni's result suggests that one market participant's FTDs can spill over to other parts of the market and cause increased stress on the broader market. Using detailed data from a large options market maker, Evans et al. (2009) finds similar strategic failure behavior in U.S. equity options markets during 1998-1999. The authors observe that the use of FTDs is due to the relatively low cost of failing. They compute an FTD's cost as "the cost of a zero-rebate equity loan plus the expected incidence of buy-in costs" and find that it amounts to only 0.1 basis points in their sample.[20] Accordingly, Evans et al. (2009) conclude that failing to deliver securities can be profitable for market makers and that this activity can affect options prices.

Recent research on equity-related FTDs has focused on the effects on stock valuation and regulatory changes. Autore, Boulton, and Braga-Alves (2015) show that stocks with high levels of failures are more likely to be over-valued but this apparent trading opportunity is difficult to arbitrage due to the high costs of short selling in these relatively illiquid securities. Thus, less-liquid stocks can remain over-valued even in the presence of high levels of FTDs. In contrast, Jain and Jain (2015) report not only a significant decline in the level of equity FTDs but also a weakening in the relationship between short selling activity and FTDs after the implementation of SEC Rules 203 and 204 in 2008-2009. However, our results show that the implementation of these rules has not had a similar effect on FTDs for ETFs. Overall, it remains an empirical question as to: 1) how FTDs affects the risks, returns, and costs of trading ETFs, 2) whether or not the underlying rationale behind FTDs is the same for stocks and ETFs, as well as 3) identify what are the effects of FTDs on the underlying securities held by these funds.

2.3 Incentives for ETF Shorting and Failing

While equity FTDs are concerning to investors and regulators largely because they are symptomatic of naked short selling, we identify an alternative cause for ETF FTDs, namely, market making activities associated with the ETF creation/redemption process. We propose a simple measure to estimate short-selling that arises from ETF liquidity provision, which we call *operational shorting*. The motivation and empirical predictions behind operational shorting are distinct from those of *directional shorting*, or naked short selling, that can also result in FTDs.

[20] "Buy-in costs" refer to the expenses incurred by a market participant who is forced to close out its FTD via the clearinghouse, the National Securities Clearing Corp. (NSCC). For an excellent description of the process of short selling, rebates, FTDs, and buy-ins, see Appendix A of Evans et al. (2009).

Through the creation/redemption process and the market making of ETF shares, authorized participants and market makers[21] have two sources of revenue that are typically considered by the literature: mispricing arbitrage and the bid-ask spread. As the market price of the ETF deviates from the NAV of the underlying securities, the AP can either create or redeem shares along with purchasing or selling the underlying securities to earn arbitrage profits from the discrepancy between the two. A market maker in the ETF shares could also earn the bid-ask spread by trading the firm's existing inventory of ETF shares.

While the discussion of the arbitrage strategy focuses on the two legs of the trade (i.e. buying the underlying basket of shares at NAV, swapping them for the ETF shares, and selling the ETF shares at market price), it ignores the timing of both legs of the trade. It might seem natural that the AP/market maker would have to purchase the underlying basket before selling the ETF shares.[22] By delaying the creation and subsequent delivery beyond T+3 of the ETF shares which it has already sold, the market maker is failing-to-deliver the ETF. As discussed above, the existence, frequency, and magnitude of cases with large fail-to-deliver shares in ETFs suggest strong incentives for market makers to delay settlement in order to generate large, predictable profits (e.g., by avoiding creation fees and delaying the outlay of capital to accumulate the full creation basket of underlying securities). As we discuss below, we develop a numerical example in the Appendix that attempts to model the AP's incentives to delay settlement of its short position, especially if the ETF's price can be easily hedged in the futures and/or options markets. Additionally, ETFs with large expense ratios or embedded costs (such as the cost of maintaining swaps for

---

[21] The ICI reports for a sample of 1,896 ETFs, the average (median) ETF had 34 (36) authorized participants, of which 5 (4) were active per their definition. Of those APs, the ICI also reports that 5 (4) were both APs and registered market makers in the ETF shares (see Antoniewicz and Heinrichs, 2014).

[22] Index Universe explains below using "Bob", a hypothetical market maker, they can actually sell the ETF shares before they enact the ETF creation, effectively generating an uncovered short position:
"Market makers are given more time to settle their accounts than everyone else: While most investors' trades must settle in T+3, market makers have up to T+6. Market makers often have reason to delay settlement for as long as they can, particularly for ETFs. If Bob is a market maker trading ETFs, it might deliberately sell more and more shares of SPY short until it's sold enough to warrant creating a basket with the ETF issuer, thus making good on its sales. The longer Bob delays basket creation, the longer it can avoid paying the creation fee (often $500 or $1,000) and related execution costs. Moreover, it can delay the time it takes before taking on responsibility for a full creation basket of ETF shares (often 50,000 shares)."
"ETF.com Briefing Book", *Index Universe*, 10/18/2011, pg. 14.

leveraged/inverse ETFs) provide a second incentive to delay settlement as long as possible because these recurring fund-related expenses can create predictable and profitable short-selling opportunities for APs and other short-term traders.

2.4 Estimating Operational Shorting Activity

To estimate operational shorting, we compare the buy-sell imbalance for trading in the ETF (our proxy for excess demand or supply of the ETF) to changes in the share creation. The formula for our measure of operational shorting is:

$$Operational\ \ Shorting = \frac{\max[0, \left(\text{Cumulative Buy/Sell Imbalance}(t-3, t-1) - \Delta\text{Shares Outstanding}(t-1, t)\right)]}{\text{Shares Outstanding(t - 3)}} \tag{1}$$

To calculate the buy-sell imbalance, we classify intradaily trades in the ETF as buys or sells by comparing the execution price of the trade with the national best bid and offer (NBBO).[23] We then aggregate the buy-sell imbalance from time t-3 to t-1 because 3 days is the typical time between a short sale and its delivery for tradeds other than bona fide market making by an AP We subtract from this the daily net create/redeem activity, which is computed as the changes in ETF shares outstanding from t-1 to t because it is at time-t when prior short sales are expected to be covered. Finally, we use the maximum function to focus on those cases where excess buys as measured by a large, positive buy-sell imbalance and exceed the actual creation of shares as measured by the changes in shares outstanding. We normalize this result by dividing by the number of ETF shares outstanding to scale the numerator.[24] To ensure that our measure of operational shorting is solely capturing excess buys beyond contemporaneous creation activity, and not driven by excess redemptions relative to a sell-imbalance (i.e. ΔShares Outstanding(t-1,t) < Cumulative Buy/Sell

[23] NBBO stands for the national best bid and offer, which is obtained from the NYSE TAQ database.

[24] Note that our measure is biased against us finding a significant effect because it is only positive when there is more net buying demand than shares actually created. This occurs when there is strong, "bullish" investor demand for the ETF shares. However, one would expect the influence of the operational shorting metric to be weaker when investors are actually bullish on a specific ETF. Thus, our specification works against us finding a significant result because one would expect that operational shorting's effect would be stronger when investors are "bearish" on the ETF but, in that case, our measure is set to zero. As we will show later, the operational shorting metric has a significant effect on FTDs and other variables even though it is non-zero only for periods when investor demand is bullish.

Imbalance(t-3,t-1) < 0), we set operational shorting to 0 whenever there is a sell imbalance.

To understand the timing of these measures, consider the AP's decision of whether or not to submit a create order on date t. Observing excess demand for the ETF shares on date t (e.g., Cumulative Buy/Sell Imbalance (t-3,t-1) > 0), APs "acting as market makers or agents to market makers" might submit a create order on that date and have 3 trading days, until t+3, to deliver the basket of underlying to complete the creation.[25] If they deliver the underlying basket by the cutoff time on t+3, the ETF shares are created and the shares outstanding at t+4 would reflect the increased number of shares outstanding. However, if they fail-to-deliver, the ETF shares outstanding will not change.

Figure 5 contains an illustrative example of how the cumulative buy-sell imbalance, change in shares outstanding and fails-to-deliver might relate, further motivating our measure. The figure shows these cumulative quantities for the iShares Core S&P Total U.S. Stock Market ETF (ticker: ITOT) over the year 2012. Early on, there are sharp increases in the cumulative *buy/sell imbalance* (black line) indicative of excess demand for the ETF. The cumulative *change in shares outstanding* (dark grey line) responds to this imbalance consistent with APs submitting orders to create new ETF units. However, the response of the cumulative *change in shares outstanding* lags behind the excess demand, possibly due to the reasons described above. Precisely when demand for the ETF increases sharply and the increase in the supply of ETF shares lags is when a spike in the percentage of *fails-to-deliver* (light grey line) occurs in ITOT shares. It would appear that APs and market makers are accommodating the demand, but the delay in creating them generates the FTDs observed. The operational shorting measure we propose above compares the cumulative *buy-sell imbalance* to the cumulative *change in shares outstanding* as an estimate of the potential short positions and failures-to-deliver that result due to the lagged response of APs/market makers to the excess demand.

---

[25] Antoniewicz and Heinrichs (2014) explain how failing-to-deliver in the primary market can generate fails in the secondary market: "Market makers, which can include APs acting as market makers or agents to market makers, have up to three additional days to settle trades (a total of T+6) if their failure to deliver is the result of bona fide market making. This mismatch in timing can create delays in the settlement of both primary market ETF redemptions and secondary market ETF trades, as market makers often use ETFs to hedge their inventories."

3. Data

We gather data from a number of sources. Primarily, FTD data[26] are disseminated publicly at the SEC's website and are made available to the SEC by National Securities Clearing Corporation's (NSCC).[27] The FTD database contains CUSIP numbers, issuer names, prices, and the total number of fails-to-deliver shares recorded in the NSCC's Continuous Net Settlement (CNS) system on a daily basis. The total number of fails-to-deliver represents the total outstanding balance of shares failed, that are aggregated over all NSCC members, regardless of when the original fail position was initiated.[28] We collect these data from March 22, 2004, which is the beginning of the dataset, through December 31, 2016. It is worth noting here that prior to September 16, 2008, only securities with aggregated fails of 10,000 shares or more were reported in the data After that date, however, all fails regardless of the outstanding fail amounts are included in the fail to deliver data that the SEC disseminates.

In order to capture daily ETF creation and redemption activity (*Net Create/Redeem Activity* or *ETF flow* variable), we rely on the daily changes in the ETF total shares outstanding. We follow Ben-David, Franzoni, and Moussawi (2015) and extract the data from Bloomberg as they are not not reported accurately in CRSP and Compustat. Bloomberg sources the ETF shares outstanding data directly from ETF administrators and custodians, which report the new shares oustanding that reflects accepted new create and redeem orders after market hours on the transaction date. While Bloomberg reports this info on the day the create/redeem orders are and submitted and accepted, it might take several days for other data vendors and exchanges to reflect this info. *Overall, we believe that our ETF create/redeem flow data computed from Bloomberg shares outstanding figures are a timely and accurate representation of when ETF shares are created/redeemed, regardless of when they (or the baskets) are settled later on.*

[26] The FTD data can be downloaded from the following SEC page: http://www.sec.gov/foia/docs/failsdata.htm.
[27] The National Securities Clearing Corporation (NSCC) is regulated by the SEC, and is a subsidiary of the Depository Trust and Clearing Corporation (DTCC). See http://www.dtcc.com/about/businesses-and-subsidiaries/nscc and http://www.dtcc.com/~/media/Files/Downloads/legal/rules/nscc_rules.pdf for more info.
[28] The total number of fails reported on day (t) reflect the fails originating at day (t) as well as the remaining outstanding fails that were not closed out from previous days. FINRA and the SEC do not distribute the actual timing of the share settlement fails, and instead disseminate the outstanding balance of fails at a given day.

We supplement the SEC data with additional variables from other sources. We merge the data with Compustat, CRSP, and FISD Mergent to determine the asset class of each of those securities, as well as the total shares outstanding or issue size. Stock price and volume data come from the CRSP database, and are used to calculate variables such as *market capitalization, stock turnover, illiquidity*, and *idiosyncratic volatlity*. ETF characteristics are extracted from CRSP Mutual Fund database, and we use ETFG database for additional ETF-specific information, such as the ETF lead market maker and the creation unit size and fee amounts. The ETF holdings of underlying stocks is drawn from Thomson-Reuters and the intraday spread and return volatility are calculated from the NYSE TAQ database. Short interest information is extracted from Compustat on a biweekly basis, and represents the level of consolidated short interest in shares as reported by exchanges and compiled by FINRA. We supplement these short interest data with daily information on securities lending supply, utilization, and lending fees using Markit Securities Finance database (formerly Data Xplorers).

In order to compute the daily buys and sell imbalances in ETF shares, we need first to appropriately sign the ETF trades into buys and sells. To do that, we use the TAQ millisecond database to classify every trade between 2004 and 2016 into a buy or sell trade using a modified algorithm that combines the methods of Lee and Ready (1991) and Ellis, Michaely and O'Hara (2000). First, for each trade, we compute the national best bid and offer (NBBO) quote at the end of the previous millisecond. Then, we compare the trade price to the best bid and best offer. The midpoint reference inherent to the Lee and Ready (1991) algorithm does not take into consideration the "outside trades" which are not permitted under the Reg. NMS rules, and therefore are less likely to occur in recent periods. For this reason, we use a modified quote test based on Ellis, Michaely and O'Hara (2000) who proposed a clever methodology that acknowledges the clustering of buys on the offer price, and sales on the bid prices.[29]

[29] According to Ellis, Michaely and O'Hara (2000), the quote test is less accurate when the trades are not executed at the ask or the bid. Most importantly, the Ellis, Michaely and O'Hara (2000) argument is especially valid when the Lee and Ready algorithm fails to take into consideration trades executed outside the quotation. Additionally, once an executed trade price crosses the prevailing NBBO within a millisecond, we stop using the quote test for the rest of the millisecond. Instead, and for the rest of the trades during this millisecond, we rely on the tick test as it is likely that the quote test is not accurate, especially when there is intense high frequency algorithmic trading that is faster than

Table 1 presents summary statistics for the key variables in our analysis. These data are computed on a daily basis for the entire ETF sample in the top portion of the table while the bottom portion reports statistics based on a sub-sample comprised solely of ETFs that invest in U.S. equities. Strikingly, the short interest ratio for the full sample, measured as a percentage of shares outstanding has a standard deviation of over 11.84%, and the 99^{th} percentile of its distribution is equal to 83.76%. This may be a product of the operational shorting mechanism that we described above. Moreover, we find that 0.42% of the average ETF's shares are considered failures (FTDs) at any given time. Lastly, the average value of our operational shorting measure is 1.01%, with a standard deviation of 2.89%. Thus, there is considerable cross-sectional and time variation in the key variables of interest in our analysis.

4. Empirical Methodology and Results

4.1 FTD Summary Statistics and Trends

Table 2 presents summary statistics for our sample of FTDs. Panel A provides the average daily FTD dollar volume by asset class. Over the course of our sample, the total volume of FTDs across all asset classes is primarily concentrated in stocks and ETFs. The total dollar volume of FTDs increased until 2007, where it reached over $7 billion. Following this, due to the regulatory changes described earlier, the total volume of FTDs fell dramatically. In 2016, the total volume of FTDs was $3.3 billion. Despite this trend, the dollar volume of ETF-related FTDs has been generally increasing over our sample period, starting at $936 million in 2004 and reaching nearly $2.6 billion in 2016.

Moreover, stock FTD volume has decreased over time, and specifically declined following the regulatory changes in 2008. At their peak, stock FTDs reached $3.9 billion. By 2016, this figure had dropped to $522 million. Figure 2 provides a graphical representation of these trends. We plot the total volume of FTDs for stocks and ETFs on a monthly basis from 2004-2016. ETF FTDs comprise a large

---

the refresh rate of the quotes within a millisecond period. So, our modified Ellis, Michaely and O'Hara method takes into consideration that buys are more likely to be executed at the ask, and sales on the bid price, and whenever an outside trade is observed during that millisecond, then the algorithm relies instead on the tick test until the end of the millisecond. After signing all trades during market hours, we sum all the buys and sales at 4:00 pm to construct our buy and sale volume for the day.

portion of total FTDs following the 2008 regulatory change and crisis period. Together, these trends show that ETFs are becoming a larger part of the overall financial market, which may lead to a greater number of FTDs in this segment of the financial markets.

Panel B of Table 2 reports FTDs as a proportion of total shares outstanding. Following the regulatory changes in 2008, these values also dramatically decrease. For ETFs, FTDs as a percentage of shares outstanding reached a high of 5.24% in 2007, and dropped to 0.83% in 2016. Similarly, for stocks, FTDs as a percentage of shares outstanding declined from 0.63% in 2004 to 0.02% in 2016. Finally, Panel C provides the maximum daily FTD dollar volume by asset class. The most important insight in this portion of table is that the dollar level of ETF-related FTDs has been growing and is now much greater than stock-related FTDs. Thus, the SEC rule changes have been effective in curbing stock-related FTDs but the maximum levels of ETF-related FTDs appear to still be problematic in 2016 (e.g., $0.9 vs. $5.8 billion, respectively). In fact, the average dollar value of ETF-related FTDs now represents 78.5% of all FTDs (up from 29.5% in 2008).

4.2 Operational Shorting and its Impact on Failures-to-Deliver (FTDs)

To better understand the nature of ETF FTDs, we run an initial series of tests that examine the impact that *Operational Shorting*, described by Equation (1), has on the levels of short interest and FTDs for ETFs. We use daily FTDs scaled by the number of ETF shares outstanding as our dependent variable. For the short interest regressions, our dependent variable is the biweekly short interest level as a proportion of shares outstanding. Proportionally to their relative market capitalization, ETFs have exhibited elevated short interest ratios in recent years that are much higher than those of common stocks. As seen in Figure 1, ETF shorting volume corresponds to more than 20% of the total dollar outstanding short interest in U.S. equity markets in recent years, which is much higher than the relative market capitalization of ETFs (about 9% in recent years).

We also develop a two main hypotheses to describe phenomena related to operational shorting: First, we posit that short interest will be associated with operational shorting. In other words, we postulate that an ETF's short interest is comprised of an additional element that has been overlooked in prior literature, and that this element is different than the directional shorting and hedging elements. Second, we include both the lagged short interest and the operational shorting variables in the FTD regression and gauge which of the shorting variables is a stronger determinant of the subsequent FTD activity. We control for the cost of borrowing the ETF shares, which also coincides with increased FTDs since higher securities lending fees make it more expensive for APs and other market participants to locate, borrrow, and then short the shares. At the same time, ETFs that are popular to short should have higher lending fees.[30] We measure this borrowing cost with a variable defined as the *Daily Cost of Borrow Score*, which measures the cost of borrowing the ETF shares based on a decile rank score of the securities lending fees provided by Markit Securities Finance Database. This variable equals 10 for the ETFs with the highest borrowing costs.

Beyond the *Short Interest Ratio, Operational Shorting*, and *Daily Cost of Borrow Score*, our regressions also include control variables based on the findings in Fotak et al. (2014) and Stratmann and Welborn (2016) related to the effects of ETF liquidity and options listing. We control for the ETF's liquidity by including size (*log of Market Cap*) and trading volume (*Share Turnover*). We also include an options listing dummy equal to 1 if options are traded on the ETF. We expect that the ETF's asset size should be negatively related to FTDs because larger funds are typically more liquid and thus it is easier to locate shares prior to the T+6 deadline. We expect ETF trading volume to be positively related to FTDs because greater share turnover increases the likelihood that some shares might not be delivered in a timely fashion. Last, *Option Listing* suggests that the ETF is larger and positions in this fund are easier to hedge, thus leading to a lower amount of FTDs.

Table 3 examines the link between short interest and FTDs or O*perational Shorting*.[31] All regressions used in our analysis include ETF and date fixed effects, and standard errors are clustered by ETF and date. Table 3 shows that all coefficients have the expected sign and are statistically significant at

---

[30] We do not have clear data on exactly who all the Authorized Participants are for each ETF. This inhibits our ability to fully run this test on the group of Authorized Participants.

[31] In our analysis, we focus on daily measures rather than longer time intervals so that the short-term dynamics and inter-relationships between FTDs and operational shorting can be examined more fully.

the 1% level. Regressions (2), (3), (5) and (6) confirm that greater trading volume, short selling activity, and securities borrowing costs are related to increased ETF FTDs and short interest levels. Of particular interest is the positive and statistically significant coefficient on *Operational Shorting* in regressions (3) and (6). Operational shorting appears to be a very strong determinant of short interest, suggesting that it represents a significant component of the short interest ratio. Additionally, the operational shorting coefficient is a stronger determinant of FTDs and appears to have a higher economic and stastistical significance than short interest ratio. Thus, when operational shorting is high, short interest and FTDs both increase, even after controlling for prior short selling activity, securities lending costs, and an ETF's liquidity-related variables such as the fund's asset size and trading volume. This finding underscores the need to decompose the effects of short selling that might be directional or informational in nature from shorting activity that is due to liquidity provision (as measured by our *Operational Shorting* variable).

4.3 The Persistence of ETF net creation activity and order imbalances

To better explain FTDs and operational shorting behavior, it is important to understand the daily patterns of ETF creations (net of any redemptions) and ETF order flows, their persistence and potential reversal patterns, if they exist. If, for example, past net creation activity and order imbalances affect future creation units and order flows in an autoregressive manner, then APs and other ETF market makers might delay the creation/redemption process to take advantage of these patterns. This would then lead to increased operational shorting and potentially higher FTDs. However, if there are no clear patterns associated with net creations and order flow, then APs/market makers would have less incentive to engage in operational shorting of ETF shares. Thus, when faced with a large buying imbalance, APs have two alternative trading strategies: 1) locate or create a sufficient number of shares to satisfy this buyer-initiated demand, or 2) sell the ETF shares now without locating or creating them and then wait up to T+6 days to obtain and deliver the shares. As described below, the AP typically has a strong incentive to follow the second strategy, especially if order flows are persistent and alternate between positive and negative imbalances over time.

APs must create ETF shares in typical large blocks of 50,000 shares. The market maker is unable to create, say, 1.5 blocks even though he/she would ideally want to create 75,000 shares to possibly cover an open short position of this size. In this example, the AP would be forced to either create 1 block of 50,000 shares or 2 blocks of 100,000 shares, both of which deviate from the AP's desired quantity of 75,000 shares. Due to the indivisibility of creation units, the AP might defer the creation of the second unit if he/she thinks the ETF's order flow is persistent and mean-reverting over time. By creating one unit of 50,000 shares today and then waiting for the next day's order flow to (hopefully) mean-revert to a negative 25,000 share order imbalance, the AP can cover the full 75,000 share short position because the -25,000 share imbalance can be offset by the AP buying 25,000 shares. Thus, by "partially cleaning up" the position with 1 creation unit and then waiting a day (or longer) with an open short position of 25,000 shares, the AP might be able to create a zero net position without having to incur the extra transaction costs and capital outlay for a second block of 50,000 shares. This example illustrates some of the incentives that can explain why APs might want to delay the creation of new ETF shares and engage in operational shorting.

In addition to the above discussion, we present a more explicit numerical example in the Appendix which shows the trade-off between the costs and benefits of an AP covering a short position either immediately or by waiting up to 6 days. As the Appendix and Figure A1 demonstrate, the variability and predictability of ETF order imbalances are important factors affecting the AP's "value of waiting" to cover its short position rather than immediately creating new ETF shares. Given the assumptions in the Appendix's example, this value of waiting by the AP can be quite profitable and thus the AP has a strong incentive to delay (or avoid altogether) the creation of new ETF shares.

Accordingly, we examine the dynamics and inter-relationships between *Net Creation Activity* and *ETF Order Imbalance* using lagged values of the dependent variables along with the liquidity-related *Controls* (fund size and trading volume), as follows:

$$Net\ \ Creation\ \ Activity_t \quad or\ \ ETF\ \ Order\ \ Imbalance_t = \alpha_0 + \ \alpha_1 Controls + \sum_{n=0}^{8} \beta_n ETF\ \ Order\ \ Imbalance_{t-n} + \sum_{n=0}^{8} \gamma_n Net\ \ Creation\ \ Activity_{t-n} + \ \epsilon_t \tag{2}$$

Equation (2) provides a parsimonious way to identify any autoregressive patterns in the dependent variables as well as possible inter-relationships between order imbalances and past creation activity, and vice versa.

To test the behavior of net creations and order imbalances described in Eq. (2), models (1)-(3) of Table 4 use contemporaneous and lagged values of order imbalances (days t-8 to t), as well as lagged values of net creation activity (days t-8 to t-1) to estimate their effects on the current level of *Net Create/Redeem Activity*.[32] Net Create/Redeem Activity is constructed on a daily basis as the percentage change in the overall ETF shares outstanding. Since this variable is a percentage change, which similar to return is lognormally distributed, and for the purposes of the autoregressive regressions in this table, we construct our flows variable, *Net Create/Redeem Activity,* as the *log*(1+% change in shares outstanding) which is likely to be more symmetrical for AP creation as well as redemption activities. After controlling for the two ETF liquidity variables, regressions (1) – (3) show that net creation activity is highly persistent with all of the net creation and order imbalance variables yielding positive and significant parameters at the 1% level. Thus, the prior sequence of net creation activity and order imbalances support the idea that past behavior plays an important role in the subsequent creation and redemption of ETF shares.

Models (4)-(6) repeat this analysis using *ETF Order Imbalance* as the dependent variable . The persistent, autoregressive pattern is also apparent in these regressions although there are some important differences when compared to *Net Create/Redeem Activity*. For example, a comparison of the parameter estimates for the first autoregressive variable shows that the lagged 1-day *ETF Order Imbalance* parameter is much higher in model (6) (0.105) than its corresponding lagged *Net Create/Redeem Activity* parameter in model (3) (0.0358). This indicates that order imbalances are much more persistent than net creations. The finding is consistent with the discrete nature of net creation activity.

In contrast to the discrete nature of net creation activity, order imbalances are continuous in nature and can respond quickly to changes in the buying and selling demand of ETF investors. Thus, it is not surprising that we find in models (5) and (6) of Table 4 that today's ETF order imbalances are more positively autocorrelated with yesterday's order imbalances than the *Net Create/Redeem Activity*

[32] We use lags up to 8 days to control for possible effects from prior short selling and FTD activity. To compute the operational shorting and order imbalance measures, we focus on buyer- and seller-initiated trades during U.S. market hours (9:30 am – 4:00 pm Eastern time) and do not include after-hours trading activity.

regressions reported in models (1)-(3). In addition, when lagged values of both net creations and order imbalances are included in model (6), there is evidence of an inverse relationship between today's *ETF Order Imbalance* and lagged *Net Create/Redeem Activity* variables, as can be seen by the negative parameters for lagged values of net creations/redemptions from day t-6 to t-2. For example, the *Net Redeem/Create Activity* parameter at t-3 is the most significant and most negative (-0.00797) while the contemporaneous time-t parameter for this variable is 0.0404, thus suggesting that order imbalances are highest when APs' net creations are currently positive while prior net creations were negative over the past 2-7 trading days (i.e., the APs were experiencing net redemptions in the past, especially at time t-3). Taken together, the results reported in Table 4 for order imbalances and net creation activity show that order imbalances are more persistent than net creations and this can be due, in part, to the discrete and discretionary behavior of APs when creating blocks of ETF shares.

4.4 The effects of ETF net creation activity and order imbalances on FTDs

Given the potential autoregressive and dynamic patterns outlined in the above discussion, it is also useful to examine the effect of order imbalances and net creations on ETF-related FTDs. We then regress FTDs and/or short interest level as a percentage of shares outstanding variable on lagged values of *ETF Order Imbalance* and *Net Create/Redeem Activity*, as well as the controls for ETF liquidity. The model also includes 1-day lagged values of the dependent variable to control for autoregressive tendencies in FTDs. Such setting can help identify the timing in terms of which lagged values of order imbalances and net creations are most closely related to FTDs and short interest ratio. In this way, we can determine whether our *Operational Shorting* variable properly captures the dynamics related to order imbalances, net creation activity, short interest and FTDs.

Panel A of Table 6 presents the regression results for FTDs and Panel B contains the results of the Short Interest Ratio regressions. By focusing on the full specification of model (6) in Table 5, Panel A, we can see that the lagged value of *ETF Order Imbalance* at t-3 has the largest and most signifcant positive coefficient when compared to all other variables in the FTD regression (i.e., 0.121 with a t-statistic of 13.82) as well as the short interest regression (i.e. 0.0126 with a t-statistic of 6.18). Given that FTDs occur after time t+3, it is not that suprising that order imbalances from 3 days prior can have such a large impact on today's FTD metric. This result shows that large positive order imbalances (which are usually symptomatic of unexpectedly strong excess buying demand by ETF investors) can lead to higher operational shorting, which consequently shows up in higher short interest, and eventually higher FTDs. The finding is consistent with the idea that APs can provide liquidity in an excess buying situation by engaging in operational shorting activity. However, some of these operational short positions might not be covered within 3 days and thus can result in a surge in FTDs. This pattern is confirmed by the relatively large positive coefficient on the t-3 *ETF Order Imbalance* variable.

Model (6) of Table 5 also shows an alternating pattern between lagged values of *Net Create/Redeem Activity* at days t-4 to t-1 and the current level of short interest and FTDs (at day t). For the shortest lag, net creations are positively related to short interest and FTDs (0.0976) and could be driven by the "partial clean-up" of past operational short positions. In contrast, net creations at t-3 are negatively related to FTDs (-0.0715) and short interest (-0.0103). It is also noteworthy that the higher and more positive the Net Create activity before t-3, the lower the the ETF short interest level, as the operational short positions are closed out. Keep in mind that the short interest data are disseminated on a biweekly basis and are refreshed once every two weeks in our sample. The large variation in coefficients in the FTD regression for net creations over a few days is similar to the relationship observed between net creations and order imbalances reported earlier in Table 4. Thus, the discretion that APs exhibit when making creation/redemption decisions in the recent past appears to not only correspond to current order imbalances but also the current level of FTDs. Further, Table 5 shows that the time period between t-3 and t-1 is the most important in terms of economic and statistical signficance. Consequently, we have formulated our definition of *Operational Shorting* in Equation (1) over this critical t-3 to t-1 period and then use this variable in the following section to analyze the key factors that explain variations in this type of shorting activity across ETFs.

4.5 Incentives for AP's Operational Shorting Activity

We next examine the determinants of the operational shorting activity as a way of understanding the incentives available for APs to participte in market making the ETF market and provide liquidity for various market participants. After establishing that operational shorting is more likely due to buy pressure and when flows are volatile with some likelihood for partial reversals, we test directly our hypothesis whether operational shorting activity is driven by AP's incentives to capture arbitrage profits resulting from buy pressure on ETF shares, by selling ETF shares at a price driven higher than the unerlying basket fundamental value. [33] Arbitrage profits represents the principal incentive for APs to engage in providing liquidity in the ETF market. The *ETF Mispricing* variable is included as operational shorting activity could also be motivated by APs' and other investors' incentives to arbitrage differences between the ETF's market price and its NAV. This variable is not calculated on an absolute basis because the sign matters. In other words, we decompose ETF mispricing into two separate variables (i.e., a "premium" vs. a "discount") which are equal to the *absolute* value of the mispricing variable only when the mispricing is positive vs. negative, and zero otherwise. This enables us to relate operational shorting to potential arbitrage activities only when the price of the ETF is *above* the NAV (i.e., the existence of a premium). By constructing the mispricing variable in this way, we can then examine the true incentive behind operational shorting. Positive values mean that the ETF's market price is too high relative to the NAV and thus one would expect *more* operational shorting to bring these two values in line (and vice versa when this variable is negative).

Additonally, we expect that opeartional shorting to be easier and less risky with the avail of reasonable and cost efficient hedges for their operational short positions in order to shield those market makers from unanticipated market swings. Market makers to be more inclined to delay creation when the underlying basket stocks are less liquid until they are have a better gauge of the permanent component of the ETF order flow before committing to a basket create order and therefore incurring related trading costs. Our model include multiple proxies for the availability of efficient hedges. The *Maximum Futures R-*

[33] We focus our analysis on *Operational Shorting* rather than *Short Interest* because we want to identify that portion of short-selling that is not offset by ETF net creation activity. Since ETF short interest is the sum of both operational shorting and directional shorting, ETFs with high operational shorting due to a high buy/sell imbalance in secondary markets are likely to have higher levels of overall short interest. Thus, we want to decompose the effect of operational and directional short selling so that the effects of AP liquidity provision can be seen more clearly. We also find that the correlation of daily movements in Operational Shorting and Short Interest is +0.14. This observation confirms that our operational shorting measure is consistent with short selling activity but does not behave exactly like short interest. Also, our calculation has the advantage of being calculated on a daily basis rather than on a bi-weekly basis.

squared variable measures the maximum explanatory power of ETF returns across a set of three different equity futures contracts. For each date, the previous 252-days of ETF NAV returns are regressed on the futures return from S&P 500-mini, the S&P MidCap 400-mini, and the Russell 2000-mini contracts.[34] The maximum R^2 across these three regressions is the value assigned to the *Maximum Futures R-squared* variable. If an AP or other investor wanted to hedge their exposure to an ETF, this R-squred variable serves as a proxy for the suitability of using futures on one of the three equity indexes as a reliable hedging vehicle.[35] Options listed on the ETF would also facilitate the hedging of ETF-specific risk. One would expect a positive relation between these hedging-related variables and *Operational Shorting* because the presence of such hedging instruments can allow an AP to provide more liquidity when they can use the futures and/or option markets to hedge this short position (e.g., via a long futures position or long call option).[36]

We also include a proxy for *Liquidity Mismatch* between the ETF and its underlying basket of securities, to capture another incentive to delay creation. We follow Pang and Zeng (2016) and measure liquidity mismatch as the difference between the trade-weighted average intraday bid-ask spread of the ETF's underlying securities and the trade-weighted average bid-ask spread for the ETF. We expect that the option to delay creation by APs is more valuable when there is a greater mismatch between the liquidity of the basket of securities relative to the ETF, as APs would prefer to observe the ETF order flows in subsequent days before committing to gathering the less underlying basket stocks and incurring related

[34] The futures data is taken from Quandl and the roll assumption used in constructing the daily futures returns is the 'last-trading-day' or 'end-to-end roll' method. This assumption "…allows you to use the front contract for as long as possible; however, the danger is that activity may have switched to the back contract prior to your roll. A trading strategy based upon this rule runs the risk of unwanted delivery and/or close-out of your positions, if you do not roll in time (the margin for error is very limited)."

[35] As the example in the Appendix demonstrates, the use of a long position in a futures contract to hedge an AP's short position can be an effective way to lock in an arbitrage profit while providing time for any order imbalance to reverse so that the AP's costs to deliver the ETF shares are reduced. Thus, a strategy of operationally shorting first, then hedging in the futures market, and ultimately covering the short position at a later date, can be more profitable than immediately covering any short position with the creation of new ETF shares. This approach can also be accomplished using options on the ETF but would entail greater upfront costs to purchase a long call position (but also provide potentially greater profit potential).

[36] An AP can also hedge by using a "proxy ETF" that is highly correlated with the AP's ETF. For example, an AP's short position in an ETF could also be hedged by entering a long position in a liquid, low-cost ETF such as the largest S&P 500 Index ETF, SPY.

transaction costs. This is especially valuable if APs expect partial reversal in the flows, and in the presence of a reasonable hedge.

Our model also includes controls for ETF-specific transaction costs / frictions: *the natural log of Creation Unit Dollar Size* and *Creation Unit Fee* (per share). Higher creation unit sizes and fees are expected to encourage APs to engage in more operational shorting in order to economize on these costs / frictions. Regression (2) begins with four independent variables: the ETF liquidity-related control variables as well as two proxies for ETF-specific transaction costs or frictions: *the natural log of Creation Unit Dollar Size*, and *Creation Unit Fee* (per share). We find in model (2) that the coefficients on *ln(Creation Unit Dollar Size)* and *Creation Unit Fee* are positive and statistically different from zero at the 1% level. In both cases, we find that the more costly it is to create or maintain ETF shares, the more likely it is that APs will turn to operational shorting, perhaps to wait for excess buying demand to subside and order flows to reverse. Alternatively, the APs could simply be buying time until they need to pay a relatively higher creation fee and save on the capital outlay required to accumulate the requisite shares in the underlying securities.

Regression (6) adds proxies for the ability to use futures and options markets to hedge a long or short exposure to an ETF (*Maximum Rolling R-Squared with Available Futures Contract* and *Available Options Dummy*). This model confirms the positive relation between operational short positions and the hedging proxies. Interestingly, when these additional variables are included in the regression, we also see that the ETF transaction cost variables are no longer significant, suggesting that frictions such as ETF creation size and fee are of secondary importance in explaining the true determinants and incentives behind AP operational shorting activities.

As expected arbitrage profits seem to be the main driver behind AP's participation in ETF market making and related operational shorting activity. Positive values for *ETF Mispricing* mean that the ETF's market price is too high relative to the NAV and thus one would expect *more* operational shorting to bring these two values in line (and vice versa when this variable is negative). Model (6) confirms our expectation that greater permium coincides with increased operational shorting due to the highly significant positive parameter estimate on *ETF Mispricing* (0.370 with a t-statistic of 10.23). In addition, consistent with our prior that greater liquidity mismatches should be positivitey related to operational shorting, model (6) additionally shows that *Proxy for Liquidity Mismatch* is positively and significantly related to the *Operational Shorting* dependent variable (although it is not as strong as the *ETF Mispricing* variable).

Overall, the signs and significance of the variables in model (6) suggest that smaller, actively traded ETFs that have greater potential profits from capturing ETF premium, with decent available hedging alternatives, and with larger liquidity mismatches engage in more operational shorting activity. Since arbitrage activity by APs is an important role in the proper functioning of the ETF market, it is important to examine how operational shorting influences the mispricing between ETF prices and NAVs. In addition, the liquidity mismatch between the underlying basket and the ETF could also lead APs to wait longer before covering their operational short positions. This could alleviate demands for liquidity in the underlying securities market while also increasing FTDs. Table 6's results are consistent with the numerical example in the Appendix, which formulates the trade-offs an AP faces when it decides to hedge its short position in order to wait for excess buying imbalances to reverse. Table 6's findings also lead us to explore the effects of operational shorting activity on ETF mispricing and the liquidity of the underlying securities which are held by ETFs.

4.6 The effects of Operational Shorting activity on future risk-adjusted returns

The academic literature has documented a strong negative relationship between short-selling constraints and future stock returns. Whether the measure of short-selling constraints is: a) short interest (e.g. Figlewski (1981); Asquith and Meulbroek (1996); Desai, Ramesh, Thiagarajan, and Balachandran (2002)), b) short interest relative to institutional ownership (e.g. Asquith, Pathak, and Ritter (2005); Nagel (2005)), c) IPO lockups (e.g. Ofek and Richardson (2003)), d) rebate rates (e.g. Jones and Lamont (2002)), e) rebate rates combined with the lendable supply of shares (e.g. Cohen, Diether, and Malloy (2007))) trade-level indications of a short-sale (e.g. Boehmer, Jones and Zhang (2008); Diether, Lee and Werner (2009))), or f) FTDs (Autore, Boulton, and Braga-Alves (2015)), the result is the same: constrained short-selling is associated with over-valuation. Ourinterpretation of this strong predictive relationship is that short-sellers are informed, but constraints prevent them from fully incorporating their information.

While operational shorting is an important component of ETF short interest and FTDs, as we have shown, our earlier discussion indicates that the operational shorting activity of APs is a nuanced one that is centered on liquidity provision and not informed, speculative short-selling. A large positive order imbalance represents unexpectedly high excess buying demand for an ETF and the AP can offset this buying pressure and provide liquidity by selling the ETF shares. The AP then faces a choice to either locate the shares immediately (at day T) or wait and deliver these shares at a later date (up to day T+6). If the AP chooses the latter approach, operational shorting and FTDs rise. Since APs are simply providing liquidity and not engaging in directional selling, we would expect operational shorting to be unrelated to future ETF returns.

To test this hypothesis, we regress future ETF returns of the sub-sample of domestic equity[37] ETFs on operational shorting, ETF and date fixed effects, and other control variables. To construct our dependent variable for this test, we calculate the returns based on the ETF's market and NAV prices over the next month (t+1 to t+22). The key independent variable is *Operational Shorting.* In order to properly risk-adjust the return performance, we include controls for well-known asset pricing factors related to the average book-to-market ratio and market cap of the stocks held by the ETF, as well as other factors such as proxies for momentum and reversals, idiosyncratic volatility, and institutional ownership. In addition, we include the *Daily Cost of Borrow Score* to control for the effects of short-selling demand. We expect the *Operational Shorting* variable to be insignificant if APs are conducting pure market making activities rather than speculative, directional short selling.

Table 7 presents results of regressions of future ETF returns on operational shorting for a sub-sample of domestic equity ETFs. These returns are based on the ETF market prices in models (1)-(2) while ETF NAVs are used in models (3) and (4). The key independent variable is *Operational Shorting.*

[37] To identify the sample of domestic equity ETFs, we use the CRSP 4-character investment objective code. From the sample of domestic equity ETFs (crsp_obj_cd = "EDxx"), we remove gold, real estate, and commodity sector funds (crsp_obj_cd = "EDSO", "EDSR" and "EDSC" respectively). We also remove hedged and dedicated short funds (crsp_obj_cd = "EDYH" and "EDYS" respectively).

Regression (1) presents baseline results and Regression (2) adds *Operational Shorting*. We find that the coefficient on *Operational Shorting* is insignificant, indicating that operational shorting is unrelated to ETF returns. Regressions (3) and (4) repeat this exercise using NAV-based returns and find similar results showing that *Operational Shorting* activity is not a significant predictor of future ETF returns. This finding is consistent with our conceptualization of *Operational Shorting* as a measure of an AP's liquidity provision services and not an indicator of informed, directional short selling activity. Since APs are providing liquidity rather than speculating on the future direction of ETF prices, it is not surprising that operational shorting behavior does not affect future ETF returns.

4.7 The effects of Operational Shorting on an ETF's Underlying Securities

4.7.1 The effects of Operational Shorting on ETF Mispricing

After establishing our operational shorting measure and its determinants in prior sub-sections, we further explore the primary motivation of APs and MMs: providing liquidity through operational shorting while capturing not only the bid-ask spread but also potentially sizable arbitrage profits. *ETF Mispricing* variable is included because operational shorting activity could also be motivated by APs' and other investors' incentives to engage in arbitrage. Since operational shorting exists only when APs are statisfying buying demand pressure, we expect an asymmetry in how deviations of ETF prices from the NAV are related to operational shorting incentives. In other words, we we use the decomposed ETF mispricing created earlier (i.e., a "premium" vs. a "discount") which are equal to the *absolute* value of the mispricing variable only when the mispricing is positive vs. negative, and zero otherwise.[38] This enables us to relate operational shorting to potential arbitrage activities only when the price of the ETF is *above* the NAV (i.e., the existence of a premium). By constructing the mispricing variable in this way, we can then examine operational shorting's effect on our main variables of interest after controlling for other important factors such as the presence of hedging alternatives.

To further elaborate on the profit incentives for operational short positions, we quantify the amount

[38] Recall that we compute mispricing by subtracting the NAV price from the market price. So, a positive (negative) value for the mispricing variable indicates a premium (discount).

of profits APs garner by engaging in operational shorting to exploit potential arbitrage mispricing opportunities. If operational shorting is aimed to profit and arbitrage away mispricing (particularly premium) opportunities, then we should therefore examine the change in mispricing following operational shorting activities as such arbitrage should be associated with a subsequent decrease in mispricing. We use two variables that measure ETF mispricing: the daily change in mispricing (*Mispricing Change*) and the absolute value of this daily change (*Absolute Mispricing Change*). These variables measure changes in the difference between an ETF's market price and NAV, measured as a percentage of ETF price. We include the absolute value of the change in mispricing to account for the possibility that mispricing might be persistent but alternating from positive to negative throughout the trading day. We also include the contemporaneous or 1-day lagged values of our *Operational Shorting* variable, and lagged versions of the dependent variables to control for possible autoregressive patterns related to ETF mispricing, as follows:

The significance of operational shorting suggests that it not only represents strong profit incentives for market makers to provide liquidity, but most importantly it can be used to help eliminate arbitrage opportunities *without* the need to physically create underlying stocks. The ability to hedge these operational shorting exposures with futures and options allow market makers to capture the mispricing change, in absolute value, as well as the spread revenue in excess of the costs of their hedge.[39] Therefore, *Operational Shorting* activity can be effective in arbitraging away mispricing opportunities without the need to immediately create ETF shares or cause price pressures to the underlying stock market.

After controlling for the significant negative coefficients on ETF fund size, model (2) of Table 8 shows that the contemporaneous level of *Operational Shorting* has a strong negative relationship with the (signed) ETF mispricing variable while model (3) also confirms that the lagged *Operational Shorting* variable is also an important determinant of ETF mispricing. Models (5) and (7) repeat these tests with

[39] For example, an AP can enter an operational short position and hedge this exposure until the short is covered via a long futures position or a long call position. Typically, this derivatives position is based on an index which is highly correlated with the ETF's portfolio. Thus, the cost of these long positions in futures or calls will need to be deducted from the amount of arbitrage mispricing in order to compute the AP's net profit on an operational short position. Unfortunately, we do not have data on the AP's hedging activities and instead rely on proxies.

the *Absolute Mispricing Change* and confirm the significant negative relationship between operational shorting and ETF mispricing. We also find that the futures hedging proxy (*Maximum Rolling R-Squared*) is negatively related to the absolute change in ETF mispricing. The results in Table 8 provide consistent evidence that larger mispricing in ETF trading coincides with greater levels of operational shorting and the presence of reliable futures hedging vehicles. Thus, increased operational shorting and the opportunity to hedge in the futures market can make it easier for APs and other market participants to engage in arbitrage actions that reduce ETF mispricing. This outcome can therefore make the pricing of ETF shares more efficient and potentially improve the liqudity not only for the ETF market but also its underlying basket of securities.

4.7.2 The effects of Operational Shorting on an ETF's Underlying Securities

As noted in Fotak et al. (2014), Stratmann and Welborn (2013), and others, FTDs for common stocks can have an impact on the liquidity of these equity securities. Also, we hypothesize that operational shorting can lead to FTDs in some instances. Thus, we can examine the impact of operational shorting behavior on the liquidity of the underlying securities held by ETFs. To do this, we use two measures of liquidity for these underlying securities: the daily *Average Intraday NBBO Spread of Underlying Stocks* and the daily *Average Intraday Second-by-Second Return Volatility of Underlying Stocks*. Following Fotak et al. (2014), we use average spreads and intraday volatility as measures of the liquidity and market quality of trading in individual securities. These spreads are computed by weighting each intraday NBBO spread by the size of the trade that immediately follows this NBBO quote. The second-by-second return volatility is calculated from the last traded price recorded during each second of the trading day.

The basic structure of the empirical model is similar to the one used in earlier sections and includes contemporaneous and 1-day lagged forms of *Operational Shorting*. For each ETF, i, and each day d, we compute the intraday second-by-second returns, using the last trade price every second, s, between 9:30 am and 4:00 pm. We then use these returns to compute the intraday volatility measures for the ETF as well as the average intraday volatility for the underlying basket of stocks held by that ETF on that day[40].

For the basket's average volatility, we first compute the intraday volatility of each common stock in the ETF portfolio on a daily basis. We then construct the average intraday volatility of the underlying stocks using the weight of each stock *j* in the ETF basket, computed as the weight in the ETF portfolio at the end of the previous month (based on Ben-David, Franzoni, and Moussawi, 2015), and multiplied by the ETF's market capitalization at the end of the previous day.[41] We follow the same approach in computing the ETF's intraday quoted spread, using the prevailing NBBO quote matched to each trade in the trade - signing procedure described earlier.

Additionally, we control for the "liquidity level" effect of ETF ownership that is documented in earlier literature. In particular, Ben-David, Franzoni, and Moussawi (2015) document causal evidence that links ETF ownership with increased volatility of underlying securities. To control for this effect, we construct and include a measure of average ownership by ETF *i* in the underlying basket stocks, using all the stocks *j* in the ETF *i* basket at the end of the previous month,[42] computed using mutual fund onwership data, as described in Ben-David, Franzoni, and Moussawi (2015).

Since our measure of operational shorting activity is computed at the ETF level, we run all of our analysis of liquidity and volatility effects using these ETF-level liquidity measures, as well as the underlying stock measures aggregated at the stock level. On the left-hand side, we use the average liquidity measures of the underlying stocks in the ETF basket, which are proxied by the average intraday stock volatility and the average intraday NBBO spread of the underlying stocks held by the ETF. We explicitly

---

[40] For the ETF: $Intraday\ \ ETF\ \ Volatility_{i,d} = \frac{\sum_{s=1}^{23,400}(ret_{s,i} - \overline{ret_i})^2}{23,400-1}$, where $ret_s = log\left(\frac{P_s}{P_{s-1}}\right)$, and 23,400 is the number of seconds over which we compute returns, during the regular market hours between 9:30am and 4:00pm.

[41] $Average\ \ Intraday\ \ Basket\ \ Volatility_{i,d} = \sum_{j=1}^{J} w_{j,d-1} \times Intraday\ \ Stock\ \ Volatility_{j,d}$

[42] $Average\ \ ETF\ \ ownership\ \ in\ \ basket\ \ stocks_{i,d} = \frac{\sum_{j=1}^{J} Dollar\ \ ETF\ \ Ownership\ \ in\ \ Stock_{j,d}}{\sum_{j=1}^{J} Market\ \ Capitalization\ \ of\ \ Stock_{j,d}}$. If we assume that the ETF has no cash or leveraged positions, which is likely to be a reasonable approximation for all physical replication plain vanilla U.S. ETFs, then the $\left[\sum_{j=1}^{J} Dollar\ \ ETF\ \ Ownership\ \ in\ \ Stock_j\ \right]$ term reduces to the market capitaliztion of the ETF *j* on day *d*. This is a reasonable assumption as most of the ETFs in our sample are physical ETFs that hold underlying securities (stocks, bonds, futures, commodities etc.) while "synthetic" ETFs are more common in Europe.

control for the lagged dependent variable and for the ETF-based liquidity measures by including up to three lags of these ETF liquidity measures to control for the persistence in volatility and spread measures, as well as address potential reverse causality concerns.

Following Ben-David, Franzoni, and Moussawi (2015), we expect that average ETF ownership is positively associated with the average underlying stock volatilities due to increased exposure to high frequency traders and other liquidity demanders that transmit their liquidity shocks to the ETF and ultimately to the ETF's underlying basket. Similarly, we expect that increased ETF ownership in basket stocks will be associated with higher spreads in these underlying stocks, as prior evidence suggests there can be a migration by liquidity-demanding investors from the underlying securities to the more-liquid ETF securities (e.g., see Dannhauser, 2016; Hamm, 2014; Israeli, Lee, and Sridharan, 2017).

On the other hand, if FTDs and, by extension, operational shorting, act as a "release valve" that improves liquidity (Fotak et al. (2014)), then we would expect the *Operational Shorting* variable to be negatively related to intraday spreads and volatility. As operational shorting increases due to a sudden surge in buying demand, the AP provides liquidity in the ETF market *without* entering the market for the underlying stocks. Through operational shorting, an AP in the ETF acts as a buffer that does not immediately transmit the liquidity shocks that hit ETFs to the underlying basket, thus cushioning the underlying stocks from higher volatility or widening spreads. If, on the other hand, the AP does not engage in operational shorting and decides to create new units of ETF shares immediately, then the AP will have to buy shares of the underlying stocks and transmit those liquidity shocks directly to the underlying securities. This can perturb the market for these underlying stocks, especially if this market is less liquid than the market for ETF shares. Thus, operational shorting at the ETF level can improve liquidity for the underlying stocks by enabling APs to forego potentially disruptive trades in the fund's basket of less-liquid securities. In turn, market makers in the underlying stocks can lower their spreads because large traders such as APs and other ETF market makers do not need to trade these individual stocks every time there is an order imbalance in the ETF shares.

Building on the results in Tables 6 and 8, the next set of regressions use two measures of liquidity for the underlying securities held by ETFs: the daily *Average Intraday NBBO Spread of Underlying Stocks* and the daily *Average Intraday Second-by-Second Return Volatility of Underlying Stocks*. Following Fotak et al. (2014), we use average spreads and intraday volatility as measures of the liquidity and market quality of trading in individual securities. In Table 9, our sample is restricted to U.S. equity-only ETFs because these are the only funds that we can reliably identify the holdings in each of the underlying stocks from Thomson Reuters. This sub-sample also facilitates our estimates of the national best bid and offer (NBBO) bid-ask spreads for both the ETFs as well as their underlying holdings. This reduces our sample sizes to around 800,000 ETF-day observations (compared to over 2.5 million in earlier tables).

Panel A of Table 9 reports the results of regressions (1)-(5) with underlying stocks' average bid-ask spread as the dependent variable. The structure of the model is similar to the one used in Table 8 and includes contemporaneous and 1-day lagged forms of *Operational Shorting*, which are our main variables of interest. We also include controls for ETF liquidity and hedging alternatives. In addition, we include contemporaneous and lagged forms of the *ETF's NBBO spread*, as well as the lagged NBBO spread of the underlying stocks held by the ETF (*Average Intraday NBBO Spread of Underlying Stocks in ETF Basket*). We find that *Operational Shorting* is negatively related to the underlying stocks' average spread, thus coinciding with an improvement in liquidity for these stocks. This evidence is consistent with our earlier results which suggested that ETF market makers can have an incentive to delay covering these short positions in order to avoid large capital outlays and ETF transaction costs.

In Panel B of Table 9, a similar set of regressions are performed on another proxy for the liqudity of a fund's underlying stocks, and we find that Operational Shorting is also negatively related to intraday (second-by-second) return volatility after controlling for a stock's market capitalization and trading volume. These results are consistent with those in Panel A and confirm that liquidity in the underlying stocks held by ETFs improves as *Operational Shorting* increases. An AP's operational shorting activity can thus have a overall beneficial effect on the market for the underlying stocks. In addition, Table 9 shows that the availability of a viable futures hedging vehicle can help improve underlying stock liquidity. This suggests that APs can use both operational shorting and hedging to provide liquidity in the ETF market and that this activity has a tangential benefit that enhances liquidity in the underlying cash market.

Consistent with prior evidence that relates ETF flows to subequent volatility (Ben-David, Franzoni and Moussawi, 2015) we find that when APs do not engage in operational shorting and decide to physically create new units of ETF shares immediately, then these APs will buy shares of the underlying stocks and transmit liquidity shocks to the underlying securities. These shocks, related to the creation activity, in turn, can worsen the liquidity in the underlying stocks. Overall, our evidence suggests that operational shorting can dampen the potentially adverse effects of ETFs on the volatility and liquidity of underlying stocks in their baskets. Operational shorting appears to act as a buffer that reduces the effects of liquidity shocks that ETFs are receiving from their clients' orders, which is consistent with the notion that operational shorting is a potentially beneficial by-product of liquidity provision by these market makers / APs. We also find that the contemporatous levels operational shorting at time-t lead to greater improvements in underlying stock liquidity at time-t.[43]

4.8 Is it All that Beneficial? Operational Shorting and Financial System Stress

We now investigate the link between operational shorting, FTDs, and financial system stress. Specifically, we focus on counterparty risk. Given that FTDs delay a form of compensation from one party to another, increases in the overall volume of FTDs might lead to the impairment of some market participants' ability to meet their other obligations in a timely way, thus leading to increased stress on the financial system. As noted in the prior section, the inter-connectedness between APs and market makers in terms of providing liquidity to a common set of ETFs can also cause FTDs and operational shorting in a subset of funds to ripple in a contagion-like manner through the entire market, thus increasing counterparty and systemic risk. We employ a widely used measure of financial system risk that relates to counterparty risk: the St. Louis Federal Reserve's weekly financial risk measure called the *Financial Stress Index* (*FSI*).

We regress the St. Louis Fed's FSI measure on lagged weekly values of FTDs and operational short

[43] We limit the regression results to the lagged operational shorting levels for cleaner and more rigorous identification, despite the fact that the improvement to underlying stock liquidity is the strongest on the day the operational shorting is initiated.

positions summed over all ETFs during each week. Aggregate FTDs for common stocks are also included because, as shown in Figure 2, stock-related FTDs were more prevalent before Rule 204's implementation in late 2008. The inclusion of stock FTDs also enables us to examine the potential differential effects of stock-related FTDs relative to ETF FTDs during the pre- and post-Rule 204 periods. The key independent variables are *Total ETF FTDs, Total ETF Operational Shorting,* and *Total Common Stock FTDs*, all of which are lagged one week. We also include the FSI measure lagged one week to control for any autoregressive effects. Since rapid growth in FTDs and operational shorting can lead to increased counterparty risk within the ETF market, we expect the FSI to be positively related to these aggregate measures. In addition, the Rule 204 regulatory change could lessen the effect of stock-related FTDs while increasing the impact of ETF-related activity after the 2008-2009 financial crisis.

Table 10 presents the results. Regressions (1)-(2) focus on the period before the implementation of SEC Rule 204 in 2009, while the remaining regressions use the 2010-2016 post-regulation sub-sample. We examine first the role of *ETF FTDs* and *Total Common tock FTDs.* We find in regressions (1) and (2) during the pre-Rule 204 period, that the Stock-related FTDs are statistically signficant and positively related to *FSI*, while the *ETF FTDs* are insignificant.

Alternatively, when we focus on the post-regulatory period in regressions (3) and (4), *ETF FTDs* become a statistically significant and positive determinant of the *FSI* while *Stock FTDs* are not statistically different from zero. Thus, regulatory changes during the 2008-2009 financial crisis which reduced the stock FTDs had a tangible effect on their ability to impact financial markets. However, the potentially harmful effects of FTDs on market-wide stress have not been completely removed because, as is evident in the post-crisis period, FTDs for ETFs are growing and now have a positive and significant impact on the FSI measure. Thus, the potential source of stress on the financial system appears to have shifted from common stocks during the pre-crisis period to ETFs during the post-crisis period. Finally, we find that the positive coefficient on *ETF Operational Shorting* is statistically significant in the post-regulatory period (see regression 4). This last finding confirms the additional impact of operational shorting behavior above and beyond the effects of FTDs.

4.9 Potential Spillover effects on FTDs and Operational Shorting activity

Our discussion has posited that that *Operational Shorting* can be an important factor that affects FTDs at the level of an indivdual ETF. Here, we explore the idea of contagion, or spillover, across many ETF FTDs. It may be the case that when an AP is having difficulty creating shares of one ETF, it is also difficult for the AP to create shares of other ETFs in which it is a lead market maker. Similarly, problems at one market maker may spill over to other APs that are trading these same ETFs, thus creating a market-wide contagion effect across several APs and the other ETFs they trade. Thus, when there is widespread difficulty in creating new ETF shares and operational shorting surges, the problem can spread across the market to all ETFs. This contagion problem is one reason that ETF FTDs and operational shorting behavior can increase counterparty risk and may contribute to overall financial system stress, as decribed in the previous section.

We use as dependent variables the FTDs at the individual ETF level as well as our *Operational Shorting* variable. Our key independent variables are the *FTD* and *Operational Shorting* variables aggregated at the lead market maker level. For each of these variables, we sum the lead market maker's activity (for both FTDs and operational shorting) on a specific day over all ETFs that it trades but we exclude the specific individual ETF's FTDs or operational shorts on that day. These varibles are denoted *Lead Market Maker FTDs* and *Lead Market Maker Operational Shorts*, respectively. In order to test for broader commonality effects associated with FTDs and operational shorting, we also compute these FTD and operational shorting variables for the entire market of ETFs. We do this by summing all FTDs and operational shorts on a given day for all ETFs while excluding the individual ETF's FTDs and operational shorts on that day. These are denoted *Overall Market FTDs* and *Overall Market Operational Shorts*, respectively. For completeness, we also include the liqudity-related control variables and the proxies for futures and options hedging vehicles. If there are contagion effects between APs and other ETF market makers, then we expect the dependent variables to be positively related to the Lead Market Maker's FTDs and Operational Shorting activity.[44]

In Table 11, we find that the coefficients on FTDs generated in other ETFs by the same lead market maker in regressions (1) and (2) are positive and statistically different from zero. Further, we find a positive and significant association with aggregate, market-wide FTDs in other ETFs from all other APs (net of the FTDs by the lead market maker affiliated with each ETF). This positive relation between individual ETF FTDs and overall market-wide FTDs suggests that the impact of other non-affiliated market maker's FTD activity is similar to the contagion effects of a lead market maker's FTDs. These regressions also show that the liquidity-related control variables are consistent with earlier tests and the hedging-related variables are negatively related to FTDs. This latter finding is consistent with the notion that investors are more likely to use futures and/or options to engage in directional short-selling, thus causing an ETF's FTDs to be lower because these FTDs have been shown in our analysis to be more highly correlated with liquidity provision activities (which are non-directional in nature).

Regression (3) and (4) repeat these tests using individual ETF-level *Operational Shorting* as the dependent variable. Similar to the results for FTDs, we find that a lead market maker's other operational shorts are positively related to an individual ETF's operational shorting activity and the *Overall Market-wide Operational Shorts* also exhibit a positive relationship with an individual ETF's operational shorting. These findings suggest there is a commonality in AP liquidity provision across a lead market maker's portfolio of ETFs and therefore FTDs and operational shorting in the ETF markets may spread in a contagion-like fashion. Additionally, the activity by non-affiliated APs other than the lead market maker also appear to play a reinforcing role in exacerbating some of these spillover effects. As in regressions (1) and (2), the liquidity control variables are consistent with our earlier findings. Regressions (3) and (4) also indicate that the hedging-related variables are positively related to operational shorting, as the presence of hedging vehicles can encourage APs to engage in more operational shorting. This latter result is also

[44] In these tests, we only include fixed effects for each ETF because the market-wide measures of FTDs and Operational Shorting are the aggregated values across all ETFs on a given day, after excluding the individual ETF values. Thus, we omit date fixed effects in order to exploit the time varying element of both the market-wide levels and the lead market maker levels in our regression.

consistent with our earlier findings.

4.10 Market Maker Leverage as a Potential Channel for Spillover effects

We examine financial leverage by APs as one of the channels by which FTDs and operational shorting can affect financial system stress. The leverage of financial institutions is a primary concern for regulators because excessively high leverage and jeopardize stability in the financial system. We gather capital positions for many of the lead market makers in our sample from the CFTC's Futures Commission Merchants Financial Reports.[45] We measure leverage by estimating the following ratio:

$$Capital\ \ Adequacy = \frac{Net\ \ Capital\ \ Required}{Adjusted\ \ Net\ \ Capital} \tag{3}$$

This ratio is bounded between 0 and 1, where the ratio for more highly leveraged market makers will approach 1. Thus, the variable is increasing in financial leverage. We then use this variable as an independent variable in a regression setting to estimate FTDs and opeational shorting with the following model:

Table 12 displays the results of this test. Regresions (1)-(3) examine FTDs and Regressions (4)-(6) examine operational shorting. Across all specifications, we find that leverage is positively related to both FTDs and operational shorting at the individual ETF level. This relation is statistically significant at the 1% level. We do not find a similar effect for market-wide leverage. The positive relation between market maker leverage and our key variables of interest, ETF FTDs and operational shorting, is evidence of one of the mechanisms by which FTDs and operational shorting at the ETF level are related to the stability of the financial system. Thus, individual APs that follow a business strategy of economizing on both trading costs (via operational shorting) and capital investment (via higher leverage) might collectively impose a significant negative externality in terms of increased system-wide financial stress through the inter-connected nature of AP-led liquidity provision.

5. Conclusion

This study is the first comprehensive analysis of failures-to-deliver (FTDs) of Exchange-Traded

[45] http://www.cftc.gov/MarketReports/financialfcmdata/HistoricalFCMReports/index.htm

Funds (ETFs) and how the liquidity provision activities of APs / market makers for ETFs can have important effects on market quality, security returns, as well as counterparty risk and market-wide stress on the financial system. We propose, test, and find evidence that these FTDs are related to "operational shorting," a manifestation of market making efforts when APs are faced with unexpectedly large excess buying demand from investors. This behavior can be explained by the dynamics of daily order imbalance and frictions associated with the creation of new ETF units.

Consistent with the economics underlying the proposed explanation (and unlike other proxies for short-selling activity), we also find that operational shorting is associated with increases in ETF short interest and FTDs but it is not predictive of future ETF returns. Our novel operational shorting measure can help disentangle the effects of liquidity-induced short-selling from informed / directional short-selling and thus has important implications for extant theories and research on short-selling. Additionally, we document that share creation is observed in the changes of ETF shares outstanding several days after the true flows appear in the ETF order imbalance. This finding is consistent with the existence of a valuable "option to wait," where the AP has an incentive to delay the delivery of ETF shares in order to maximize its profit from liquidity provision and arbitrage activities (while also increasing the level of FTDs in the financial system).

Although these liquidity provision activities can be profitable to APs and other market makers, they might create a negative externality in terms of greater levels of FTDs that spill over from one market maker's ETFs to other, unrelated ETFs which, in turn, can increase counterparty and systemic risk. We find that increased levels of FTDs and operational shorting at the lead market maker's other ETFs can have a contagion-like effect which is positively related to a system-wide measure of financial market stress. In addition, the FTDs and operational shorting activity aggregated across the overall market are positively related to an individual fund's FTDs and operational shorts. These results suggest ETF trading relies on an inter-connected network of liquidity providers which, at times, pursue positively correlated trading strategies that can be detrimental to the overall market. We also identify a potential channel by which market makers' financial leverage can reinforce the effects of operational shorting on system-wide financial stress. These effects represent a shift away from the importance of stock-related FTDs during the pre-crisis period to a more prominent role for ETF-related FTDs here in the post-crisis period. Thus, the unique structure of ETFs and the AP's ability to create / redeem shares on an intraday basis can lead to improvements to trading in individual ETFs but, at an aggregate level, it can create greater counterparty risk that has potentially destabilizing effects in the broader market not only for ETFs but also the underlying securities held by these ETFs.

Given that the above results have focused on operational shorting, one possible avenue for future research pertains to examining the asymmetry of AP behavior when there is excess selling pressure from ETF investors. In this alternative situation, the AP could provide liquidity by engaging in "operational buying" of the (relatively) cheap ETF shares and potentially minimizing the cost of this activity by redeeming ETF shares quickly in order to receive the underlying basket of (more-valuable) securities. However, it is unclear how quickly operational buy positions are covered relative operational short positions. Thus, additional research into this asymmetry between and operational buying vs. operational shorting is warranted.

References

Acharya, V., 2009. A theory of systemic risk and design of prudential bank regulation. *Journal of Financial Stability*, 5, 224-255.

Antoniewicz, R., Heinrichs, J., 2014. Understanding Exchange-Traded Funds: How ETFs work. *ICI Research Perspective* 20(5), 1-39.

Antoniewicz, R., Heinrichs, J., 2015. The Role and Activities of Authorized Participants of Exchange-Traded Funds. *ICI Research Report* March, 1-13.

Ashley, R.A., Parmeter, C.F., 2015. When is it justifiable to ignore explanatory variable endogeneity in a regression model? *Economics Letters* 137, 70-74.

Asquith, Paul, and Lisa Meulbroek, 1996, An empirical investigation of short interest. Working paper.

Asquith, Paul, Parag A. Pathak, and Jay R. Ritter, 2005, Short interest, institutional ownership, and stock returns, *Journal of Financial Economics* 78, 243–276.

Autore, D.M., Boulton, T.J., Braga-Alves, M.V., 2015. Failures to deliver, short sale constraints, and stock overvaluation, *Financial Review* 50, 143-172.

Bhattacharya, A., and M. O'Hara, 2016, Can ETFs Increase Market Fragility? Effect of Information Linkages in ETF Markets, Working paper.

Battalio, R., Schultz, P., 2011. Regulatory uncertainty and market liquidity: The 2008 short sale ban's impact on equity option markets. *Journal of Finance* 66, 2013-2053.

Ben-David, I., Franzoni, R., Moussawi, R., 2015, Do ETFs Increase Volatility. Working paper: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1967599

Ben-David, I., Franzoni, R., Moussawi, R., 2017, Exchange Traded Funds (ETFs), *The Annual Review of Financial Economics*, 9 (6), November 2017.

Boehme, Rodney D., Bartley R. Danielsen, and Sorin M. Sorescu, 2006, Short sale constraints, differences of opinion, and overvaluation, *Journal of Financial and Quantitative Studies* 41, 455–487.

Boehmer, Ekkehart, Charles M. Jones, and Xiaoyan Zhang, 2008, Which Shorts Are Informed?. *The Journal of Finance* 63, 491–527.

Boni, L., 2006. Strategic delivery failures in U.S. equity markets, *Journal of Financial Markets* 13, 397-421.

Chu, Y., Hirshleifer, D., Ma., L., 2016. The causal effect of limits to arbitrage on asset pricing anomalies. Working paper.

Chinco, A., and V. Fos, 2016, The Sound of Many Funds Rebalancing, Working Paper, University of Illinois at Urbana-Champagne.

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

Da, Z., and S. Shive, 2014, Exchange Traded Funds and Asset Return Correlations, Working Paper, Notre Dame University.

Dannhauser, C., 2016, The Impact of Innovation: Evidence from Corporate Bond ETFs, *Journal of Financial Economics*, Forthcoming.

Desai, Hemang, K. Ramesh, S. Ramu Thiagarajan, and Bala V. Balachandran, 2002, An Investigation of the Informational Role of Short Interest in the Nasdaq Market, *Journal of Finance* 57, 2263-2287.

Diether, K.B., Lee, K-h., Werner, I.M., 2009. Short-sale strategies and return predictability. *Review of Financial Studies* 22, 576-607.

Duffie, D., Garleanu, N., Pedersen, L.H., 2007. Valuation in the over-the-counter markets. *Review of Financial Studies* 20, 1865-1890.

Ellis, K., Michaely, R., O'Hara, M., 2001. The accuracy of trade classification rules: Evidence from Nasdaq. *Journal of Financial and Quantitative Analysis* 35, 529-551.

Evans, R., C. Geczy, D. Musto, Reed, A., 2009. Failure is an option: Impediments to short selling and option prices, *Review of Financial Studies* 22, 1955–1980.

Figlewski, S., 1981, The informational effects of restrictions of short sales: Some empirical evidence, *Journal of Financial and Quantitative Analysis* 16, 463-476.

Fotak, V., V. Raman, and P. Yadav, 2014. Fails-to-deliver, short selling, and market quality. *Journal of Financial Economics* 114, 493–516.

Hamm, S., 2014, The Effect of ETFs on Stock Liquidity. Working paper.

Israeli, D., C. Lee, and S. Sridharan, 2017, Is there a Dark Side to Exchange Traded Funds (ETFs)? An Information Perspective, *Review of Accounting Studies* 22, 1048–1083.

Jain, A., Jain, C., 2015. Fails-to-Deliver before and after the implementation of Rule 203 and Rule 204. *Financial Review* 50, 611-636.

Kyle, A., 1985. Continuous auctions and insider trading, *Econometrica* 53, 1315-1335.

Madhavan, A., 2012. Exchange-Traded Funds, market structure, and the flash crash. *Financial Analysts Journal* 68(4), 20-35.

Madhavan, A., 2014. Exchange-Traded Funds: An overview of institutions, trading, and impacts. *Annual Review of Financial Economics* 6, 311-341.

Malamud, S., 2015, A Dynamic Equilibrium Model of ETFs. Working Paper.

Nagel, Stefan, 2005, Short sales, institutional investors, and the cross-section of stock returns, *Journal of Financial Economics* 78, 277–309.

Nutz, M., and J. Scheinkman, 2017. Supply and shorting in speculative markets, Working Paper, Princeton University.

Ofek Eli, Matthew Richardson, 2003, DotCom mania: The rise and fall of Internet stock prices, *Journal of Finance* 58, 1113-1138.

Pan, K., and Y. Zeng, 2016, ETF arbitrage under liquidity mismatch. Working paper.

Rapach, D. E., M. C. Ringgenberg, and G. Zhou, 2016, Short interest and aggregate stock returns, *Journal of Financial Economics* 121, 46-65.

Senchack, A., Starks, L., 1993, Short-sale restrictions and market reaction to short-interest announcements, *Journal of Financial and Quantitative Analysis* 28, 177-194.

Shleifer, A., Vishny, R.W., 1997. The limits of arbitrage. *Journal of Finance* 52, 35-55.

Stratmann, T., Welborn, J.W., 2013. The option market maker exception to SEC Regulation SHO, *Journal of Financial Markets* 16, 195–226.

Stratmann, T., Welborn, J.W., 2016. Informed short selling, fails-to-deliver, and abnormal returns, *Journal of Empirical Finance* 38, 81–102.

Sullivan, R., and J. Xiong, 2012, How Index Trading Increases Market Vulnerability, *Financial Analysts Journal* 68(2), 70–84.

Zhang, Tony Q., and Thomas Judge, 2016, "Investment Analysis of Leveraged ETFs", working paper.

Table 1 – Summary Statistics: This table presents summary statistics for key variables used in our analysis. The sample period is 2004-2016.

	Variable	Obs	Mean	Std.Dev.	p1	p25	p50	p75	p99
Entire ETF Sample	Fail-to-Deliver Shares / Shares Outstanding	3,007,239	0.42%	1.53%	0.00%	0.00%	0.00%	0.11%	11.45%
	Operational Shorting, as % of Shares Outstanding	3,006,555	1.01%	2.89%	0.00%	0.00%	0.00%	0.65%	20.83%
	Net Create/Redeem Activity: log (1 + % change in Shares Outstanding)	3,006,045	0.11%	1.37%	-5.72%	0.00%	0.00%	0.00%	8.82%
	ETF Order Imbalance: (Buys - Sells) / Average Shares Outstanding	2,772,648	0.15%	1.81%	-7.15%	-0.15%	0.03%	0.29%	10.63%
	Market Capitalization, $ million	3,007,054	$867.19	$2,600.87	$1.38	$16.81	$86.20	$427.69	$18,523.09
	Daily Share Turnover, % of Shares Outstanding	2,950,760	4.0%	8.6%	0.1%	0.6%	1.2%	2.8%	55.5%
	Amihud Illiquidity Measure	2,756,643	0.11	0.37	0.00	0.00	0.00	0.04	2.59
	% Mispricing: % difference between ETF price and NAV	2,912,330	0.029%	0.572%	-2.332%	-0.118%	0.016%	0.184%	2.115%
	Maximum Rolling R-Squared with Available Futures Contracts	2,673,729	53%	29%	0%	30%	59%	77%	96%
	Available Options Dummy	3,007,239	0.31	0.46	0.00	0.00	0.00	1.00	1.00
	Creation Unit Size	931,999	69,602	35,005	25,000	50,000	50,000	100,000	250,000
	Creation Unit Fee	931,999	$1,577.56	$2,664.75	$100.00	$500.00	$500.00	$1,400.00	$15,000.00
	Bid-Ask Spread, at Close	2,956,434	0.330%	0.542%	0.011%	0.067%	0.147%	0.339%	3.544%
	Intraday NBBO Bid-Ask Spread, Trade Size Weighted	2,772,053	0.269%	0.395%	0.012%	0.064%	0.135%	0.288%	2.470%
	Intraday Volatility, using second-by-second intraday returns	2,703,755	0.0083%	0.0083%	0.0000%	0.0037%	0.0061%	0.0100%	0.0511%
	Daily Cost of Borrow Score	1,768,565	3.19	1.47	1.00	2.00	3.00	4.00	7.00
	Indicative Fee	1,588,220	4.37%	3.44%	0.38%	1.75%	3.50%	6.00%	18.00%
	Short Interest Ratio	2,946,535	4.66%	11.84%	0.00%	0.28%	0.90%	3.20%	83.76%
US Equity ETF Sample	Intraday NBBO Bid-Ask Spread, Trade Size Weighted	856,148	0.1785%	0.2895%	0.0117%	0.0485%	0.0963%	0.1843%	2.4695%
	Intraday Volatility, using second-by-second intraday returns	847,880	0.0086%	0.0076%	0.0000%	0.0045%	0.0065%	0.0099%	0.0511%
	Daily Cost of Borrow Score	571,069	2.63	1.26	1.00	2.00	3.00	3.00	7.00
	Indicative Fee	525,851	3.134%	2.596%	0.375%	1.125%	2.500%	4.000%	18.000%
	Short Interest Ratio	863,150	6.40%	15.24%	0.00%	0.30%	0.97%	3.98%	83.76%
	Average Intraday NBBO Bid-Ask Spread for Underlying Basket Stocks	866,921	0.1036%	0.1176%	0.0263%	0.0454%	0.0683%	0.1122%	0.8835%
	Average Intraday Volatility of Underlying Basket Stocks	866,897	0.0195%	0.0102%	0.0084%	0.0130%	0.0164%	0.0223%	0.0668%
	Underlying Basket Stocks, Average Daily Cost of Borrow Score	813,294	1.12	0.24	1.00	1.00	1.03	1.12	2.564050
	Underlying Basket Stocks, Average Indicative Fee	866,921	0.572%	0.475%	0.281%	0.386%	0.430%	0.544%	3.688%
	Underlying Basket Stocks, Average Short Interest Ratio	866,921	4.15%	2.40%	0.81%	2.34%	3.51%	5.41%	12.63%

Table 2 – Failures-to-Deliver (FTDs) Summary Statistics: This table presents summary statistics for the analysis below. Panel A reports the average daily dollar volume of failures-to-deliver (FTDs), Panel B reports the average daily FTDs in terms of the percentage of shares outstanding, and Panel C reports the maximum daily FTDs per year. All three panels report figures by asset class, but Panel B reports the results only for securities that we were able to identify in CRSP, Compustat, and Mergent FISD databases.

A. Average Daily Fail-To-Deliver Dollar Volume, by Asset Classes, $ million

Year	Total Dollar FTD	ETF	Common Stock	OTC Stocks	Corporate Bond	ADR	Structured Products	Units and Trusts	Other Securities	# of Securities with Positive FTD
2004	$3,439.9	$936.0	$2,103.8	$36.7	$35.9	$212.7	$21.2	$102.6	$2.8	2,739
2005	$3,011.3	$974.4	$1,691.4	$43.2	$25.5	$201.1	$14.6	$65.4	$0.3	2,488
2006	$3,443.6	$994.1	$2,040.2	$42.6	$88.7	$211.1	$19.7	$50.7	$1.2	2,639
2007	$7,129.6	$2,540.9	$3,520.4	$50.5	$451.3	$359.4	$40.9	$57.5	$117.1	2,937
2008	$6,401.6	$1,887.7	$3,931.2	$47.2	$45.8	$342.6	$66.1	$46.7	$44.2	4,545
2009	$1,430.0	$866.4	$402.0	$10.3	$15.9	$91.7	$25.4	$13.0	$10.6	6,465
2010	$1,953.3	$1,272.4	$495.0	$14.9	$13.9	$114.1	$20.2	$15.7	$12.4	6,265
2011	$2,479.4	$1,705.2	$543.1	$16.9	$15.5	$142.3	$30.8	$15.5	$19.2	6,109
2012	$1,877.0	$1,183.7	$509.0	$11.3	$20.5	$99.3	$23.8	$20.8	$18.3	5,731
2013	$2,065.3	$1,313.6	$552.4	$10.4	$20.1	$106.7	$29.2	$24.4	$17.6	5,588
2014	$2,704.9	$1,734.0	$746.4	$11.8	$20.0	$137.3	$36.3	$14.7	$12.0	6,074
2015	$3,460.1	$2,506.3	$734.2	$9.1	$15.1	$137.6	$37.4	$11.2	$15.9	6,190
2016	$3,304.1	$2,592.5	$522.1	$8.2	$10.3	$122.0	$32.1	$14.5	$7.0	5,951

B. Average Daily Fail-To-Deliver % of Shares Outstanding, As Percent of Security Shares Outstanding, by Asset Classes

Year	Total FTD, % of Shares Outstanding	ETF	Common Stock	OTC Stock	Corporate Bond	ADR	Structured Products	Units and Trusts	Other Securities	# of Securities with Positive FTD
2004	0.83%	3.94%	0.63%	1.12%	1.29%	1.01%	1.49%	0.47%	1.57%	1,943
2005	0.57%	2.40%	0.39%	1.02%	0.78%	0.63%	0.65%	0.27%	0.58%	1,756
2006	0.73%	3.35%	0.33%	1.72%	1.05%	0.49%	0.48%	0.20%	1.42%	1,834
2007	0.99%	5.24%	0.37%	2.01%	1.01%	0.46%	0.55%	0.22%	0.82%	2,124
2008	0.82%	4.05%	0.31%	1.66%	0.32%	0.23%	0.97%	0.14%	0.45%	3,507
2009	0.22%	0.85%	0.03%	1.20%	0.05%	0.03%	0.21%	0.02%	0.03%	5,400
2010	0.18%	1.02%	0.03%	0.61%	0.09%	0.02%	0.17%	0.02%	0.00%	5,373
2011	0.23%	1.15%	0.04%	0.53%	0.07%	0.04%	0.33%	0.02%	0.00%	5,216
2012	0.17%	0.87%	0.03%	0.28%	0.07%	0.03%	0.24%	0.02%	0.00%	5,185
2013	0.23%	1.10%	0.03%	0.14%	0.05%	0.11%	0.27%	0.02%	0.00%	5,061
2014	0.17%	0.80%	0.03%	0.18%	0.04%	0.06%	0.31%	0.01%	0.00%	5,553
2015	0.17%	0.68%	0.02%	0.34%	0.03%	0.08%	0.31%	0.01%	0.00%	5,664
2016	0.18%	0.83%	0.02%	0.31%	0.02%	0.02%	0.14%	0.01%	0.00%	5,504

C. Maximum Daily Fail-to-Deliver Dollar Volume, by Asset Classes (in $ millions)

Year	Total Dollar FTD	ETF	Common Stock	OTC Stocks	Corporate Bond	ADR	Structured Products	Units and Trusts	Other Securities	# of Securities with Positive FTD
2004	$6,746.4	$2,966.5	$3,980.2	$70.9	$100.1	$340.5	$77.9	$166.1	$4.6	3,233
2005	$8,399.0	$6,144.9	$4,775.8	$96.5	$90.9	$397.9	$80.5	$106.6	$1.9	3,108
2006	$12,211.2	$4,666.0	$9,627.2	$138.7	$616.6	$484.9	$54.6	$76.1	$4.5	3,146
2007	$19,143.9	$14,410.2	$6,451.3	$145.5	$5,405.9	$1,421.0	$96.5	$110.7	$2,630.2	3,958
2008	$20,315.2	$11,014.8	$9,055.2	$370.1	$236.2	$940.4	$140.0	$98.2	$194.0	11,280
2009	$6,734.8	$6,062.7	$979.9	$36.1	$90.8	$229.0	$66.1	$49.4	$36.9	7,387
2010	$7,509.6	$7,021.2	$2,249.6	$174.2	$30.0	$418.8	$77.8	$117.1	$101.0	7,089
2011	$9,118.6	$7,863.0	$1,170.6	$55.6	$112.2	$448.0	$151.7	$59.2	$70.3	7,076
2012	$6,230.1	$4,960.3	$1,237.2	$36.3	$113.5	$371.4	$83.5	$60.8	$95.7	6,802
2013	$8,582.5	$6,910.3	$2,695.9	$60.4	$59.5	$244.8	$135.8	$616.2	$132.3	6,749
2014	$6,839.8	$4,215.9	$4,390.9	$38.0	$79.8	$390.7	$295.0	$37.8	$136.2	6,935
2015	$11,598.2	$10,018.5	$2,431.4	$75.0	$47.5	$471.5	$132.9	$46.6	$831.9	7,095
2016	$6,554.8	$5,819.9	$854.3	$14.8	$14.6	$173.6	$58.0	$21.1	$13.8	6,092

Table 3 – The Deterimants of Failures-to-Deliver: This table displays the regression results of the short interest as a fraction of share outstanding, and the proportion of FTDs relative to total shares outstanding of an ETF as a function of several determinants of shorting and fail-to-deliver activities. The key independent variables are *Operational Shorting*, which measures the propensity for the operational shorting of ETF shares, the lagged ETF's Short Interest Ratio, Maximum Rolling R-Squared with Available Futures Contract and Available Options Dummy which are proxies for the ability to use futures and options markets to hedge a long or short exposure of an ETF, and the Daily Cost of Borrow Score which measures the daily cost of borrowing based on a decile rank score of lending fee provided by Markit Securities Finance Database (formerly Data Xplorers), where 100 equals the highest securities borrowing cost. The sample period is 2004-2016. All specifications include ETF and date fixed effects, and the standard errors are clustered at the ETF and date levels. The t-statistics are based on standard errors clustered at the ETF and date level are in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Short Interest / Shares Outstanding (t)			Fail-to-Deliver Shares / Shares Outstanding (t)		
	(1)	(2)	(3)	(4)	(5)	(6)
log (Market Cap) at (t-1)	-0.00203***	-0.00183***	-0.000878**	-0.00352***	-0.00326***	-0.00258***
	(-5.708)	(-4.766)	(-2.213)	(-13.78)	(-8.799)	(-6.923)
Share Turnover, as % of Shares Outstanding at (t-1)	0.0394***	0.0310***	0.0284***	0.0721***	0.0755***	0.0737***
	(5.427)	(4.577)	(4.228)	(7.650)	(6.381)	(6.185)
Short Interest Ratio, as % of Shares Outstanding at (t-1)	0.697***	0.767***	0.767***	0.0469***	0.0332***	0.0322***
	(37.34)	(43.58)	(43.52)	(8.947)	(6.807)	(6.655)
Daily Cost of Borrow Score at (t-1)		0.000558***	0.000536***		0.000421***	0.000408***
		(2.803)	(2.683)		(3.038)	(2.933)
Available Options Dummy at (t-1)		0.00258***	0.00230***		-0.00282***	-0.00300***
		(3.244)	(2.873)		(-4.645)	(-4.891)
Operational Shorting, as % of Shares Outstanding at (t-1)			0.105***			0.0753***
			(7.493)			(9.613)
Observations	260,352	163,454	163,454	2,925,879	1,755,400	1,755,400
R-squared	0.787	0.848	0.849	0.100	0.125	0.129

Table 4 – The Dynamics of Net Creation Units and Order Imbalances: This table displays Ordinary Least Squares (OLS) regression results. The dependent variables are *Net Creation Units (Flows) and ETF Order Imbalance,* (Signed Buy Trades – Signed Sell Trades) / Average Shares Outstanding. These measures estimates the inter-relationships between the net demands on creating new ETF units and buying ETF shares. Independent variables such as the ETF's *log(Market Cap)* and *Share Turnover* are included along with contemporaneous and lagged forms of Flows and BSI (for days t through t-8). The sample period is March, 22 2004 – December 31, 2016, and t-statistics based on standard errors clustered at the ETF and date level are in parentheses. All specifications include ETF and date fixed effects, and the standard errors are clustered at the ETF and date levels. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Net Create/Redeem Activity at day t: log (1 + % change in Shares Outstanding)			ETF Order Imbalance at day t: (Buys - Sells) / Average Shares Outstanding		
	(1)	(2)	(3)	(4)	(5)	(6)
log (Market Cap), at (t-15)	-0.000509***	-0.000268***	-0.000208***	-0.00146***	-0.00217***	-0.00145***
	(-16.78)	(-5.648)	(-5.104)	(-16.24)	(-13.27)	(-15.89)
Share Turnover, as % of Shares	0.00530***	0.00525***	0.00422***	0.00350***	0.00512***	0.00343***
Outstanding, at (t-15)	(12.79)	(10.73)	(9.840)	(7.124)	(6.666)	(7.050)
ETF Order Imbalance at (t)		0.0246***	0.0249***			
		(8.368)	(8.763)			
ETF Order Imbalance at (t-1)		0.0668***	0.0660***	0.108***		0.105***
		(13.07)	(13.20)	(23.26)		(22.22)
ETF Order Imbalance at (t-2)		0.0466***	0.0433***	0.0643***		0.0626***
		(14.15)	(13.94)	(17.95)		(16.65)
ETF Order Imbalance at (t-3)		0.0288***	0.0240***	0.0455***		0.0450***
		(11.65)	(10.43)	(15.04)		(14.09)
ETF Order Imbalance at (t-4)		0.0193***	0.0144***	0.0419***		0.0421***
		(9.251)	(7.560)	(14.06)		(13.33)
ETF Order Imbalance at (t-5)		0.0151***	0.0103***	0.0375***		0.0379***
		(7.628)	(6.021)	(12.75)		(12.11)
ETF Order Imbalance at (t-6)		0.0126***	0.00766***	0.0321***		0.0326***
		(7.166)	(5.121)	(12.37)		(11.90)
ETF Order Imbalance at (t-7)		0.00968***	0.00513***	0.0331***		0.0337***
		(6.286)	(3.722)	(11.92)		(11.50)
ETF Order Imbalance at (t-8)		0.00695***	0.00208	0.0359***		0.0362***
		(5.039)	(1.568)	(15.47)		(14.68)
Net Create/Redeem Activity at (t)					0.0699***	0.0404***
					(19.18)	(10.91)
Net Create/Redeem Activity at (t-1)	0.0507***		0.0358***		0.0249***	-0.00232
	(11.93)		(7.821)		(11.78)	(-1.045)
Net Create/Redeem Activity at (t-2)	0.0463***		0.0362***		0.0169***	-0.00526***
	(20.54)		(14.87)		(10.55)	(-2.928)
Net Create/Redeem Activity at (t-3)	0.0318***		0.0235***		0.0109***	-0.00797***
	(11.16)		(7.887)		(6.906)	(-4.210)
Net Create/Redeem Activity at (t-4)	0.0223***		0.0148***		0.0110***	-0.00519***
	(8.134)		(4.801)		(6.444)	(-2.719)
Net Create/Redeem Activity at (t-5)	0.0299***		0.0246***		0.00947***	-0.00437**
	(8.221)		(6.078)		(5.724)	(-2.356)
Net Create/Redeem Activity at (t-6)	0.0125***		0.00784***		0.00809***	-0.00397***
	(5.298)		(2.966)		(6.030)	(-2.863)
Net Create/Redeem Activity at (t-7)	0.0195***		0.0160***		0.00813***	-0.000557
	(10.13)		(7.323)		(5.514)	(-0.372)
Net Create/Redeem Activity at (t-8)	0.0172***		0.0150***		0.00806***	0.00334**
	(10.20)		(7.763)		(5.376)	(2.223)
Observations	2,950,589	2,136,427	2,136,427	2,136,427	2,364,099	2,136,427
R-squared	0.024	0.038	0.043	0.091	0.055	0.092

Table 5 – The Effects of Net Creation Activity and Order Imbalances on Failures-to-Deliver and Short Interest: This table displays Ordinary Least Squares (OLS) regression results. The dependent variable is *Failure-to-Deliver* (scaled by Total ETF Shares outstanding). Independent variables such as the ETF's *Size* and *Share Turnover* are included along with contemporaneous and lagged forms of Flows and BSI (for days t through t-8). The sample period is March, 22 2004 – December 31, 2016, and t-statistics based on standard errors clustered at the ETF and date level are in parentheses. All specifications include ETF and date fixed effects, and the standard errors are clustered at the ETF and date levels. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Failures to Deliver Regressions

	Fail-to-Deliver Shares / Shares Outstanding (t)					
	(1)	(2)	(3)	(4)	(5)	(6)
log (Market Cap), at (t-15)	-0.00384***	-0.00112***	-0.00377***	-0.00120***	-0.00375***	-0.00113***
	(-11.07)	(-10.92)	(-13.99)	(-14.65)	(-10.85)	(-10.80)
Share Turnover, as % of Shares	0.0820***	0.0237***	0.0847***	0.0254***	0.0803***	0.0239***
Outstanding, at (t-15)	(8.638)	(8.098)	(8.768)	(8.658)	(8.490)	(8.206)
ETF Order Imbalance at (t-1)	0.0219***	0.0103***			0.0156***	0.00760***
	(4.477)	(4.382)			(3.395)	(3.301)
ETF Order Imbalance at (t-2)	0.0368***	0.0215***			0.0174***	0.0167***
	(6.911)	(7.447)			(3.605)	(5.889)
ETF Order Imbalance at (t-3)	0.145***	0.119***			0.127***	0.121***
	(13.46)	(13.55)			(12.59)	(13.82)
ETF Order Imbalance at (t-4)	0.111***	0.00799**			0.100***	0.0134***
	(12.28)	(2.403)			(11.69)	(4.245)
ETF Order Imbalance at (t-5)	0.0867***	0.00726***			0.0804***	0.0113***
	(11.45)	(2.719)			(10.84)	(3.911)
ETF Order Imbalance at (t-6)	0.0675***	0.00513*			0.0634***	0.00736**
	(9.793)	(1.797)			(9.337)	(2.565)
ETF Order Imbalance at (t-7)	0.0523***	0.00246			0.0489***	0.00365
	(8.810)	(1.085)			(8.252)	(1.542)
ETF Order Imbalance at (t-8)	0.0469***	0.00641**			0.0436***	0.00707***
	(7.960)	(2.426)			(7.500)	(2.631)
Net Create/Redeem Activity at (t-1)			0.265***	0.101***	0.251***	0.0976***
			(27.01)	(17.25)	(23.27)	(15.46)
Net Create/Redeem Activity at (t-2)			0.137***	-0.0398***	0.102***	-0.0684***
			(13.76)	(-3.933)	(9.605)	(-5.872)
Net Create/Redeem Activity at (t-3)			0.0361***	-0.0531***	-0.00472	-0.0715***
			(4.986)	(-12.17)	(-0.584)	(-13.91)
Net Create/Redeem Activity at (t-4)			0.0177***	-0.00280	-0.0166**	-0.0103***
			(2.855)	(-0.916)	(-2.304)	(-2.928)
Net Create/Redeem Activity at (t-5)			0.0152***	0.00580**	-0.0102	0.00242
			(2.860)	(2.319)	(-1.584)	(0.839)
Net Create/Redeem Activity at (t-6)			0.0118**	0.00561**	-0.00790	0.00144
			(2.364)	(2.267)	(-1.312)	(0.503)
Net Create/Redeem Activity at (t-7)			0.0161***	0.00941***	0.00240	0.00667**
			(3.778)	(4.024)	(0.476)	(2.512)
Net Create/Redeem Activity at (t-8)			0.0130***	0.00471**	0.00491	0.00245
			(3.599)	(2.479)	(1.183)	(1.147)
Fail-to-Deliver Shares / Shares		0.697***		0.699***		0.698***
Outstanding (t-1)		(84.72)		(88.66)		(82.59)
Observations	2,151,271	2,151,271	2,950,592	2,950,592	2,151,271	2,151,271
R-squared	0.128	0.557	0.104	0.541	0.137	0.559

Panel B: Short Interest Regressions

	Short Interest / Shares Outstanding (t)					
	(1)	(2)	(3)	(4)	(5)	(6)
log (Market Cap), at (t-15)	-0.00912***	-2.23e-05	-0.00886***	-0.000166***	-0.00885***	-6.22e-05*
	(-5.291)	(-0.636)	(-5.933)	(-4.729)	(-5.146)	(-1.732)
Share Turnover, as % of Shares	0.300***	0.00262***	0.292***	0.00369***	0.295***	0.00323***
Outstanding, at (t-15)	(8.295)	(3.894)	(8.430)	(5.370)	(8.183)	(4.772)
ETF Order Imbalance at (t-1)	0.0478***	0.00404***			0.0442***	0.00355**
	(4.402)	(2.964)			(4.147)	(2.573)
ETF Order Imbalance at (t-2)	0.0514***	0.00534***			0.0349***	0.00436***
	(5.082)	(3.302)			(3.456)	(2.750)
ETF Order Imbalance at (t-3)	0.0623***	0.0127***			0.0370***	0.0126***
	(6.070)	(6.327)			(3.551)	(6.180)
ETF Order Imbalance at (t-4)	0.0732***	0.00977***			0.0429***	0.0108***
	(7.081)	(5.017)			(4.036)	(5.452)
ETF Order Imbalance at (t-5)	0.0807***	0.00622***			0.0488***	0.00857***
	(7.508)	(3.573)			(4.351)	(4.837)
ETF Order Imbalance at (t-6)	0.0840***	0.00139			0.0522***	0.00481***
	(7.467)	(0.752)			(4.413)	(2.614)
ETF Order Imbalance at (t-7)	0.0890***	0.000680			0.0584***	0.00513***
	(7.488)	(0.395)			(4.686)	(2.886)
ETF Order Imbalance at (t-8)	0.0948***	-0.00112			0.0653***	0.00447***
	(7.302)	(-0.675)			(4.834)	(2.684)
Net Create/Redeem Activity at (t-1)			0.213***	0.0214***	0.185***	0.0170***
			(17.27)	(8.786)	(12.83)	(7.024)
Net Create/Redeem Activity at (t-2)			0.200***	-5.51e-05	0.171***	-0.00475**
			(17.38)	(-0.0272)	(12.58)	(-2.304)
Net Create/Redeem Activity at (t-3)			0.181***	-0.00657***	0.153***	-0.0103***
			(16.41)	(-3.109)	(11.67)	(-4.544)
Net Create/Redeem Activity at (t-4)			0.157***	-0.0149***	0.129***	-0.0182***
			(14.92)	(-6.295)	(10.29)	(-7.511)
Net Create/Redeem Activity at (t-5)			0.135***	-0.0160***	0.109***	-0.0174***
			(13.10)	(-7.028)	(9.000)	(-7.496)
Net Create/Redeem Activity at (t-6)			0.108***	-0.0201***	0.0866***	-0.0214***
			(10.50)	(-8.533)	(7.332)	(-8.788)
Net Create/Redeem Activity at (t-7)			0.0862***	-0.0203***	0.0686***	-0.0214***
			(8.001)	(-8.813)	(5.768)	(-8.942)
Net Create/Redeem Activity at (t-8)			0.0540***	-0.0286***	0.0429***	-0.0290***
			(4.908)	(-11.34)	(3.628)	(-11.26)
Short Interest / Shares Outstanding (t-1)		0.980***		0.979***		0.981***
		(448.1)		(440.0)		(452.7)
Observations	2,476,342	2,475,921	2,926,486	2,925,790	2,476,342	2,475,921
R-squared	0.663	0.988	0.652	0.986	0.664	0.988

Table 6 – The Determinants of Operational Shorting: This table displays Ordinary Least Squares (OLS) regression results. The dependent variable is *Operational Shorting*. This measure estimates the propensity for operational shorting of ETF shares. Several independent variables, like the *Expense Ratio* and *Indicative Fee* are included. The sample period is March, 22 2004 – December 31, 2016, and t-statistics based on standard errors clustered at the stock and date level are in parentheses. All specifications include ETF and date fixed effects, and the standard errors are clustered at the ETF and date levels. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Operational Shorting, as % of Shares Oustanding at day (t)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
log (Market Cap), at (t-15)	-0.00730***	-0.00778***	-0.00806***	-0.00802***	-0.00899***	-0.00687***	-0.00962***
	(-14.19)	(-11.02)	(-14.16)	(-13.99)	(-6.777)	(-7.636)	(-4.590)
Average Share Turnover, as % of Shares	0.0320***	0.0246***	0.0319***	0.0316***	0.0376***	0.0330***	0.0409***
Outstanding, at (t-15)	(11.81)	(6.848)	(11.44)	(11.39)	(4.597)	(5.108)	(3.307)
Creation Unit Dollar Size, log, at (t-1)		0.00536***					0.00228
		(7.596)					(1.433)
Creation Unit Fee, per share, at (t-1)		0.0455***					0.00167
		(3.165)					(0.0784)
Maximum Rolling R-Squared with Available			0.0112***	0.0111***	0.0129***	0.00912***	0.0149**
Futures Contracts at (t-1)			(7.100)	(7.004)	(3.190)	(3.034)	(2.551)
Available Options Dummy at (t-1)			0.00243***	0.00253***	0.00387***	0.00261***	0.00276*
			(3.745)	(3.881)	(3.034)	(2.894)	(1.690)
Mispricing at (t-1): % difference between ETF				0.286***	0.243***		0.370***
price and NAV at the close of the previous day				(22.69)	(5.974)		(10.23)
Premium at (t-1), if mispricing>0, and zero						0.253***	
						(5.208)	
Discount at (t-1), in absolute value, if						-0.238***	
mispricing<0, and zero otherwise						(-3.230)	
Proxy for Liquidity Mismatch, at (t-1): Average					0.219***	0.157***	0.250**
Intraday Basket Spread - Intraday ETF Spread					(3.674)	(3.670)	(2.302)
Observations	2,950,667	1,988,950	2,633,071	2,624,669	787,099	820,652	499,849
R-squared	0.164	0.201	0.166	0.168	0.199	0.184	0.262

Table 7 – Operational Shorting, and Future Risk-Adjusted Returns: This table displays Ordinary Least Squares (OLS) regression results. The dependent variable is the 1-month forward looking ETF return (t+1, t+22) based on ETF price in specifications (1) and (2) and ETF NAV in specifications (3) and (4). The key independent variables are *Operational Shorting*, which measures the propensity for operational shorting of ETF shares and the *Daily Cost of Borrow Score* a measure of securities lending fees. The sample period is March, 22 2004 – December 31, 2016, and t-statistics based on standard errors clustered at the stock and date level are in parentheses. All specifications include ETF and date fixed effects, and the standard errors are clustered at the ETF and date levels. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	1-Month Forward Looking ETF Return (t+1, t+22)			
	Price Return		NAV Return	
	(1)	(2)	(3)	(4)
Average log (Book-to-Market) of Underlying Stocks, at (t)	-0.0505***	-0.0507***	-0.0556***	-0.0557***
	(-4.173)	(-4.176)	(-4.072)	(-4.074)
Average log (Market Cap) of Underlying Stocks, at (t)	0.00401	0.00400	0.00384	0.00382
	(1.329)	(1.324)	(1.264)	(1.259)
Reversal Proxy: Past 22-day Return, at (t)	-0.0406**	-0.0409**	-0.0405**	-0.0408**
	(-2.046)	(-2.057)	(-2.081)	(-2.092)
Momentum Proxy: Past 12-Month Return, with one month reversal, at (t)	0.0200***	0.0199***	0.0195***	0.0195***
	(3.291)	(3.280)	(3.245)	(3.234)
Institutional Ownership, as % of Shares Oustanding, at (t)	-0.00369	-0.00375	-0.00363	-0.00369
	(-1.124)	(-1.143)	(-1.116)	(-1.134)
Idiosyncratic Volatility, at (t)	-0.161	-0.162	-0.122	-0.122
	(-0.610)	(-0.614)	(-0.399)	(-0.402)
Daily Cost of Borrow Score, at (t)	-0.000712**	-0.000705**	-0.000808**	-0.000801**
	(-2.246)	(-2.226)	(-2.401)	(-2.383)
Operational Shorting, as % of Shares Outstanding at (t)		0.0138		0.0131
		(1.307)		(1.242)
Observations	304,708	304,708	304,708	304,708
R-squared	0.816	0.816	0.806	0.806

Table 8 – ETF Mis-Pricing and Arbitrage Activity: This table displays Ordinary Least Squares (OLS) regression results. The dependent variables are *Mispricing Change* and *Absolute Mispricing Change,* which are based on (ETF market price – ETF NAV) / ETF Price. This measure is regressed on our *Operational Shorting* variable and other variables that control for the ETF's asset size (*Size*), trading volume (*Turnover Average*), as well as hedging-related proxies (*Future R-squared* and *Options dummy*) and interactions between these hedging proxies and *Operational Shorting*. The sample period is March, 22 2004 – December 31, 2016, and t-statistics based on standard errors clustered at the stock and date level are in parentheses. All specifications include ETF and date fixed effects, and the standard errors are clustered at the ETF and date levels. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Mispricing Change at (t)				Absolute Mispricing Change (at t)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Operational Shorting, as % of Shares	-0.00233***	-0.00225***			-0.00128***	-0.000877***		
Outstanding at (t)	(-12.80)	(-12.60)			(-6.058)	(-5.927)		
Operational Shorting, as % of Shares			-0.000243**	-0.00133***			-0.00115***	-0.000716***
Outstanding at (t-1)			(-2.225)	(-9.912)			(-5.491)	(-4.810)
log (Market Cap), at (t-15)	-2.12e-05*	-2.37e-05**	-7.31e-06	-1.85e-05**	-0.000243***	-0.000164***	-0.000242***	-0.000162***
	(-1.772)	(-2.400)	(-0.743)	(-2.130)	(-9.731)	(-9.629)	(-9.692)	(-9.553)
Average Share Turnover, as % of Shares	4.90e-05	5.32e-05	-1.10e-05	1.63e-05	-0.000182	-0.000122	-0.000185	-0.000126
Outstanding, at (t-15)	(0.688)	(0.765)	(-0.159)	(0.257)	(-0.883)	(-0.853)	(-0.901)	(-0.887)
Maximum Rolling R-Squared with		6.18e-05	3.97e-05	6.84e-05	-0.00150***	-0.000969***	-0.00150***	-0.000971***
Available Futures Contracts at (t-1)		(0.280)	(0.180)	(0.367)	(-7.076)	(-6.404)	(-7.081)	(-6.414)
Available Options Dummy at (t-1)		5.99e-06	1.00e-06	6.26e-06	-9.36e-05*	-6.37e-05*	-9.39e-05*	-6.41e-05*
		(0.285)	(0.0477)	(0.337)	(-1.754)	(-1.746)	(-1.759)	(-1.756)
Mispricing Change at (t-1)				-0.485***				
				(-59.10)				
Absolute Mispricing Change at (t-1)						0.330***		0.330***
						(41.17)		(41.17)
Observations	2,864,290	2,624,038	2,624,039	2,623,622	2,624,038	2,623,621	2,624,039	2,623,622
R-squared	0.039	0.040	0.039	0.266	0.369	0.438	0.369	0.438

Table 9 – Effects of ETF Operational Shorting on the Liquidity of the Underlying Securities: This table displays Ordinary Least Squares (OLS) regression results. The dependent variable in Panel A is the intraday bid-ask spread of the underlying stocks held by U.S. equity-only ETFs. Panel B's dependent variable is the intraday 1-second return volatility The key independent variable is *Operational Shorting.* The sample period is March, 22 2004 – December 31, 2016, and t-statistics based on standard errors clustered at the stock and date level are in parentheses. All specifications include ETF and date fixed effects, and the standard errors are clustered at the ETF and date levels. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Effect of Operational Shorting Intraday NBBO Spread of Underlying Stocks

	Average Intraday NBBO Spread of Underlying Stocks in ETF Basket (t)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Average ETF Ownership in Underlying Stocks in ETF Basket (t-1)	0.00216210*** (3.14)	0.00066539*** (3.03)			0.00216031*** (3.14)	0.00066489*** (3.03)	0.00066526*** (3.03)
Operational Shorting, as % of Shares Outstanding at (t-1)			-0.00026142** (-2.16)	-0.00008075* (-1.93)	-0.00023448** (-2.40)	-0.00007327** (-2.03)	
Operational Shorting, as % of Shares Outstanding at (t)							-0.00009766*** (-2.67)
log (Market Cap), at (t-15)	-0.00002512** (-2.55)	-0.00000740** (-2.30)	-0.00001784* (-1.86)	-0.00000493 (-1.59)	-0.00002668*** (-2.65)	-0.00000789** (-2.40)	-0.00000805** (-2.44)
Average Share Turnover, as % of Shares Outstanding, at (t-15)	-0.00027368** (-2.07)	-0.00008531** (-2.03)	-0.00029175* (-1.77)	-0.00008913* (-1.74)	-0.00026616** (-2.01)	-0.00008297* (-1.96)	-0.00008221* (-1.94)
Intraday NBBO Spread of ETF, at (t)	0.00493170** (2.41)	0.00239582*** (2.60)	0.00935935** (2.24)	0.00342491** (2.39)	0.00490543** (2.40)	0.00238771*** (2.60)	0.00238477*** (2.59)
Intraday NBBO Spread of ETF, at (t-1)	0.00416669** (2.20)	0.00089497 (1.08)			0.00413961** (2.19)	0.00088665 (1.07)	0.00088090 (1.06)
Intraday NBBO Spread of ETF, at (t-2)	0.00404646** (2.19)	0.00105572 (1.62)			0.00401106** (2.18)	0.00104477 (1.61)	0.00104227 (1.60)
Intraday NBBO Spread of ETF, at (t-3)	0.00442017** (2.30)	0.00131459* (1.76)			0.00438403** (2.29)	0.00130343* (1.75)	0.00130040* (1.74)
Average Intraday NBBO Spread of Underlying Stocks in ETF Basket (t-1)		0.68773597*** (28.93)		0.69196466*** (29.12)		0.68770651*** (28.93)	0.68770000*** (28.93)
Observations	837,347	837,333	853,554	852,955	837,347	837,333	837,333
R-squared	0.755	0.869	0.753	0.870	0.755	0.869	0.869

Table 9 – Effects of ETF Operational Shorting on the Liquidity of the Underlying Securities: *(continued)*

Panel B: Effect of Operational Shorting Intraday Second-by-Second Return Volatility of of Underlying Stocks

	Average Intraday Second-by-Second Return Volatility of Underlying Stocks in ETF Basket (t)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Average ETF Ownership in Underlying Stocks in ETF Basket (t-1)	0.00018230***	0.00006649***			0.00018216***	0.00006645***	0.00006651***
	(2.79)	(2.78)			(2.78)	(2.77)	(2.78)
Operational Shorting, as % of Shares Outstanding at (t-1)			-0.00002750**	-0.00001036**	-0.00002604***	-0.00000998***	
			(-2.56)	(-2.51)	(-2.71)	(-2.73)	
Operational Shorting, as % of Shares Outstanding at (t)							-0.00000974***
							(-2.71)
log (Market Cap), at (t-15)	-0.00000288***	-0.00000103***	-0.00000207**	-0.00000073**	-0.00000306***	-0.00000110***	-0.00000110***
	(-3.08)	(-2.99)	(-2.08)	(-1.99)	(-3.26)	(-3.18)	(-3.18)
Average Share Turnover, as % of Shares Outstanding, at (t-15)	0.00004437**	0.00001635**	0.00004410*	0.00001601*	0.00004506**	0.00001662**	0.00001661**
	(2.09)	(2.09)	(1.85)	(1.85)	(2.11)	(2.12)	(2.12)
Average Intraday Second-by-Second Return Volatility of ETF, at (t)	0.11402739***	0.06758951***	0.11371293***	0.06704016***	0.11427900***	0.06769228***	0.06769561***
	(12.14)	(12.21)	(11.98)	(12.03)	(12.15)	(12.21)	(12.21)
Average Intraday Second-by-Second Return Volatility of ETF, at (t-1)	0.06845335***	-0.00042104	0.06819799***	-0.00115933	0.06872139***	-0.00030888	-0.00026267
	(10.17)	(-0.15)	(9.99)	(-0.39)	(10.18)	(-0.11)	(-0.09)
Average Intraday Second-by-Second Return Volatility of ETF, at (t-2)	0.06534263***	0.02290503***	0.06509124***	0.02246237***	0.06575231***	0.02306781***	0.02304690***
	(10.18)	(8.63)	(10.03)	(8.44)	(10.20)	(8.68)	(8.67)
Average Intraday Second-by-Second Return Volatility of ETF, at (t-3)	0.06200923***	0.01963887***	0.06205356***	0.01921328***	0.06239564***	0.01979273***	0.01975866***
	(9.47)	(7.55)	(9.34)	(7.18)	(9.49)	(7.58)	(7.57)
Average Intraday Volatility of Underlying Stocks in ETF Basket, at (t-1)		0.63641562***		0.64258288***		0.63632873***	0.63632892***
		(38.53)		(37.29)		(38.53)	(38.52)
Observations	822,739	822,712	823,270	822,712	822,739	822,712	822,712
R-squared	0.844	0.907	0.841	0.907	0.844	0.907	0.907

Table 10 – Financial System Stress and Failures to Deliver: This table displays Ordinary Least Squares (OLS) regression results. The dependent variable is a marcoeconomic variable which measures counterparty risk and stress in the financial system: the weekly St. Louis Federal Reserve's *Financial Stress Index* (FSI). The key independent variable, *Total ETF FTD*, is the FTDs aggregated over all ETFs, expressed as a percentage of shares outstanding. *Total Common Stock FTDs* represent aggregated FTDs of common stocks. Also included is *ETF Operational Shorting*, which measures the market-wide propensity for operational shorting of an ETF. The sample period is March, 22 2004 – December 31, 2016, and is split into two periods: from 2004 to the passage of SEC Rule 204 in 2009; and from 2010 to 2016. T-statistics based on standard errors clustered at the ETF and date level are in parentheses. All specifications include ETF and date fixed effects, and the standard errors are clustered at the ETF and date levels. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Weekly Series of St. LouisFinancial Stress Index (FSI)			
	(1)	(2)	(3)	(4)
FSI (t-7)	0.989***	0.991***	0.888***	0.880***
	(42.04)	(41.77)	(50.37)	(50.61)
Total ETF Fail-to-Deliver $ Shares, % of ETF Market	3.313	3.483	93.04***	90.83***
Capitalization (t-1 to t-7)	(0.516)	(0.537)	(5.947)	(6.147)
Total Common Stock Fail-to-Deliver $ Shares, % of Common	1,708***	1,728***	425.7	298.6
Stock Market Capitalization (t-1 to t-7)	(3.796)	(3.771)	(0.463)	(0.345)
ETF Operational Shorting (t-1 to t-7)†		-1.919		5.829**
		(-0.626)		(2.377)
Constant	-0.219***	-0.202***	-0.234***	-0.273***
	(-5.146)	(-4.718)	(-6.612)	(-7.121)
Observations	289	289	302	302
R-squared	0.958	0.958	0.936	0.937
Sample Period	Before Rule 204 in 2009	Before Rule 204 in 2009	After 2010	After 2010

† The operational shorting measure used previously in the paper focuses on the t-1 to t-3 time frame of the fail-to-deliver decision, the time frame of operational shorting in this table is t-1 to t-7 to match the weekly frequency of the St. Louis Financial Stress Index.

Table 11 – Market Makers' Spillover Effects on FTDs and Operational Shorting: This table displays Ordinary Least Squares (OLS) regression results. The dependent variable are ETF-related *Failures-to-Deliver (FTD)* and *Operational Shorting*. Several independent variables, like the FTDs and Operational Short positions of the *Affiliated Lead Market Maker* and the *Overall Market*, are included. The sample period is March, 22 2004 – December 31, 2016, and t-statistics based on standard errors clustered at the ETF and date level are in parentheses. All specifications include ETF fixed effects, and the standard errors are clustered at the stock and date levels. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Fail-to-Deliver Shares / Shares Outstanding, at day (t)		Operational Shorting / Shares Oustanding, at day (t)	
	(1)	(2)	(3)	(4)
log (Market Cap), at (t-15)	-0.00233***	-0.00209***	-0.00644***	-0.00625***
	(-19.79)	(-18.50)	(-19.76)	(-19.29)
Average Share Turnover, as % of Shares Outstanding, at (t-15)	0.0302***	0.0291***	0.0263***	0.0250***
	(13.85)	(13.06)	(11.39)	(10.84)
Affiliated Lead Market Maker Fail-to-Deliver, % of LMM Total Volume,	0.00950***			
excluding individual ETF FtDs and Volume	(10.34)			
Market-Wide Fail-to-Deliver, % of Overall Trading Volume, *excluding*	0.0225***			
Affiliated Lead Market Marker ETF FtDs and Total Volume	(10.29)			
Affiliated Lead Market Maker Fail-to-Deliver, % of All Affiliated ETF Market		0.354***		
Cap, *excluding individual ETF FtDs and Market Cap*		(10.59)		
Market-Wide Fail-to-Deliver, % of ETF Market Cap, *excluding Affiliated Lead*		0.737***		
Market Marker ETF FtDs and Market Cap		(13.52)		
Affiliated Lead Market Maker Operational Shorts, % of LMM Total Volume,			0.00119**	
excluding individual ETF Operational Shorts and Volume			(2.460)	
Market-Wide Operational Shorts, % of Overall Trading Volume, *excluding*			0.0110***	
Affiliated Lead Market Marker ETF Operational Shorts and Market Cap			(5.589)	
Affiliated Lead Market Maker Operational Shorts, % of All Affiliated ETF				0.108***
Market Cap, *excluding individual ETF Operational Shorts and Volume*				(6.176)
Market-Wide Operational Shorts, % of ETF Market Cap, *excluding Affiliated*				0.0372
Lead Market Marker ETF Operational Shorts and Volume				(1.171)
Maximum Rolling R-Squared with Available Futures Contracts at (t-1)	-0.00280***	-0.00298***	0.00601***	0.00501***
	(-4.056)	(-4.337)	(5.477)	(4.955)
Available Options Dummy at (t-1)	-0.000829***	-0.000868***	0.00214***	0.00218***
	(-3.397)	(-3.785)	(4.298)	(4.280)
Observations	2,307,010	2,307,615	2,307,010	2,307,615
R-squared	0.125	0.126	0.158	0.157
ETF Fixed Effects	Yes	Yes	Yes	Yes
Date Fixed Effects	No	No	No	No

Table 12 – The Effect of Market Makers' Leverage on FTDs and Operational Shorting: This table displays Ordinary Least Squares (OLS) regression results. The dependent variables are ETF-related *Failures-to-Deliver (FTD)* and *Operational Shorting*. Several independent variables, like the FTDs and Operational Short positions of the *Affiliated Lead Market Maker* and the *Overall Market*, are included. The sample period is March, 22 2004 – December 31, 2016, and t-statistics based on standard errors clustered at the ETF and date level are in parentheses. All specifications include ETF fixed effects, and the standard errors are clustered at the stock and date levels. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
	Fail-to-Deliver Shares / Shares Outstanding, at day (t)			Operational Shorting / Shares Oustanding, at day (t)		
Affiliated Lead Market Maker Capital Adequacy	9.30e-09***	1.00e-08***	1.00e-08***	2.28e-08***	2.44e-08***	2.44e-08***
	(4.578)	(4.182)	(4.182)	(4.208)	(3.471)	(3.471)
Market-Wide Capital Adequacy		-3.08e-09	-3.08e-09		2.27e-09	2.27e-09
		(-1.162)	(-1.162)		(0.170)	(0.170)
log (Market Cap), at (t-15)	1.89e-10***	1.87e-10***	1.87e-10***	5.46e-10***	3.35e-10	3.35e-10
	(3.827)	(3.663)	(3.663)	(2.721)	(1.381)	(1.381)
Average Share Turnover, as % of Shares	1.58e-09*	1.34e-09	1.34e-09	9.38e-10	4.09e-10	4.09e-10
Outstanding, at (t-15)	(1.707)	(1.361)	(1.361)	(1.003)	(0.418)	(0.418)
Maximum Rolling R-Squared with Available		1.72e-10	1.72e-10		3.68e-09***	3.68e-09***
Futures Contracts at (t-1)		(0.542)	(0.542)		(2.852)	(2.852)
Available Options Dummy at (t-1)		1.07e-10	1.07e-10		1.75e-09**	1.75e-09**
		(1.013)	(1.013)		(2.216)	(2.216)
ETF Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Date Fixed Effects	No	No	No	No	No	No
Observations	1,042,546	973,516	973,516	1,042,546	973,516	973,516
R-squared	0.188	0.188	0.188	0.166	0.167	0.167

Figure 1 – ETF Short Interest: This figure displays graphically the aggregate ETF short interest in billions of dollars and as a percentage of the overall short interest in U.S. Equity Markets on a daily basis from 2004-2016.

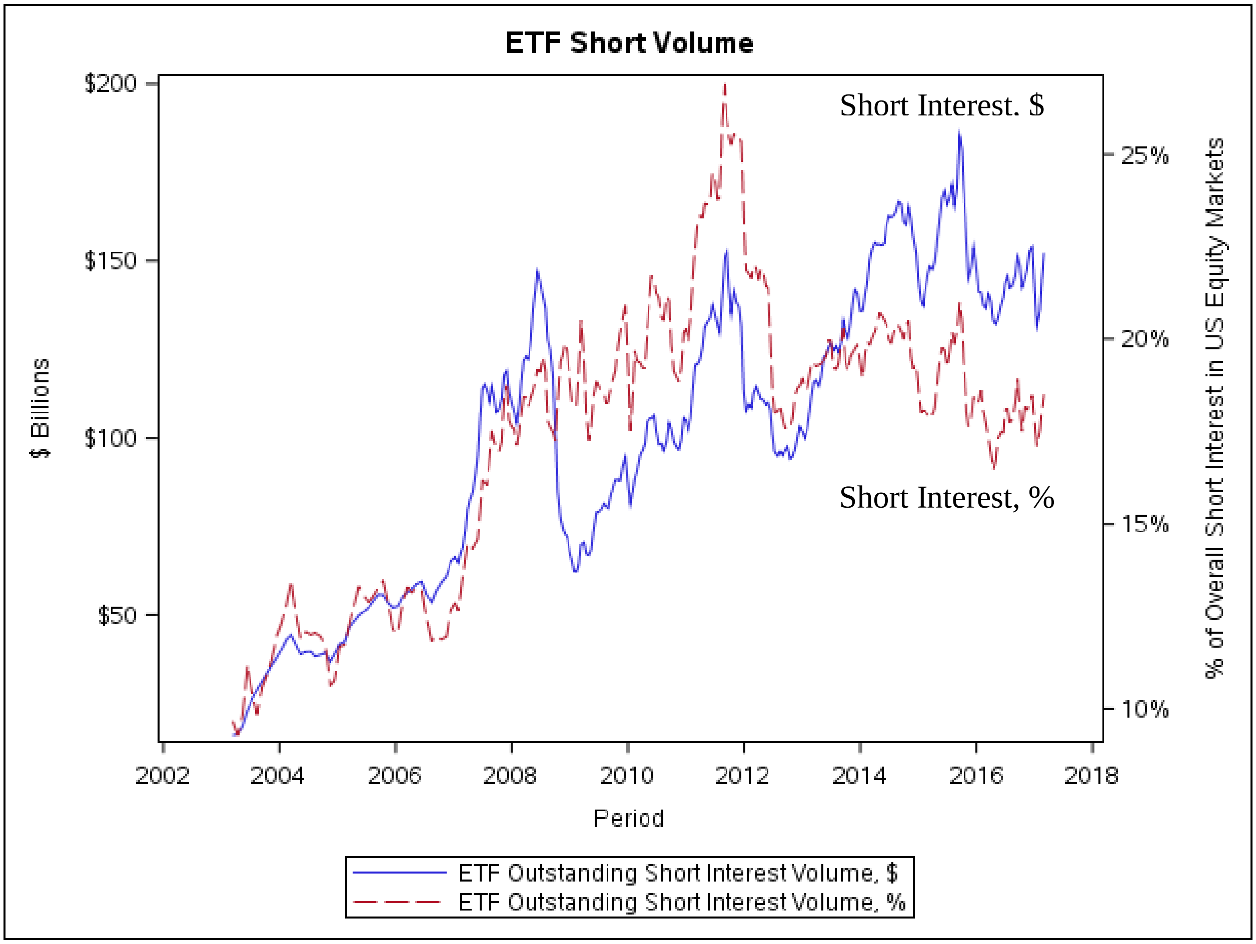


Figure 2 – Failure-to-Deliver (FTD) Activity of ETFs and Common Stocks: This figure displays graphically the average dollar volume of ETFs and common stocks on a daily basis from 2004-2016. We include only the rolling average daily FTD volume of stocks and ETFs in this graph, as they comprise the vast majority of total FTDs in the financial system.

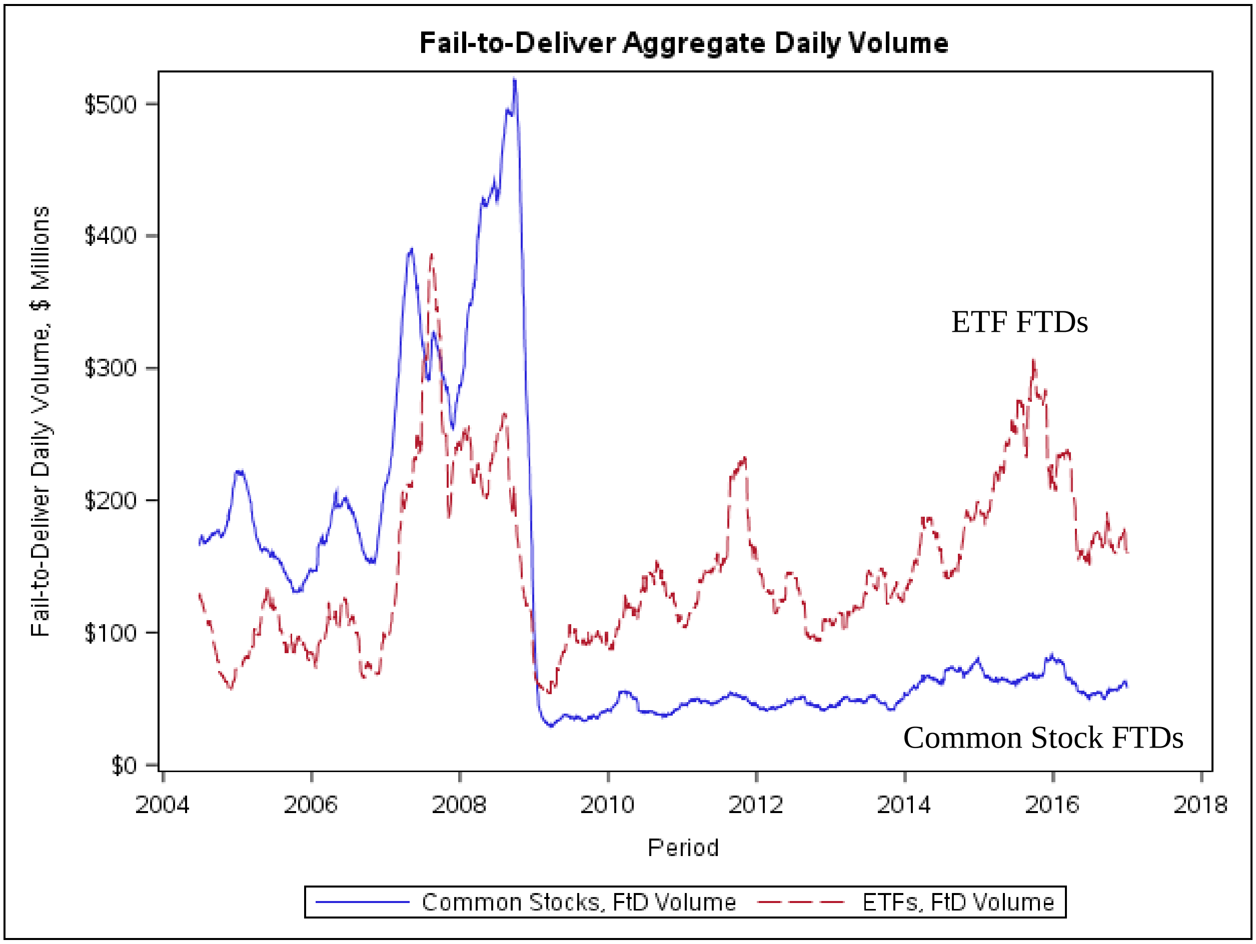


Figure 3 – ETF Settlement Failure Timeline: This figure displays the key events during a settlement failure for an ETF. Time *t* represents the time when an operational short is established. Dates *t+i*, where *i* is between 1 and 6, represent *i* days after the operational short position is established.

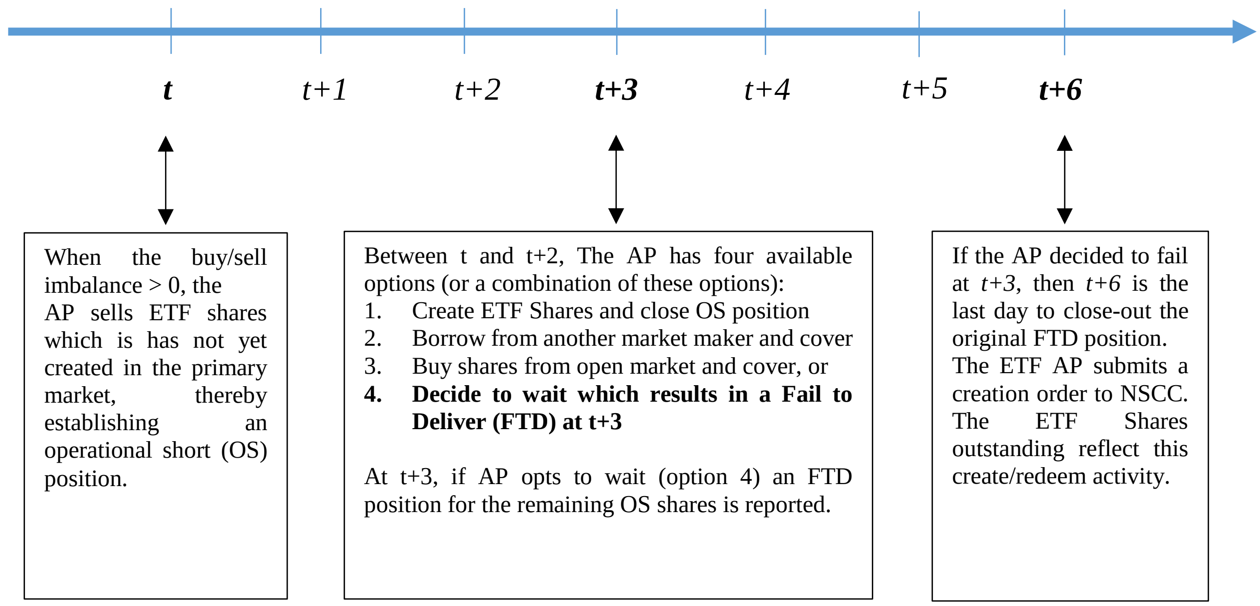


Figure 4 – Operational Shorting and Failure-to-Deliver (FTD) Activity of ETFs: This figure displays graphically the rolling-average daily dollar value of Operational Shorting activity and FTDs for ETFs from 2004-2016.

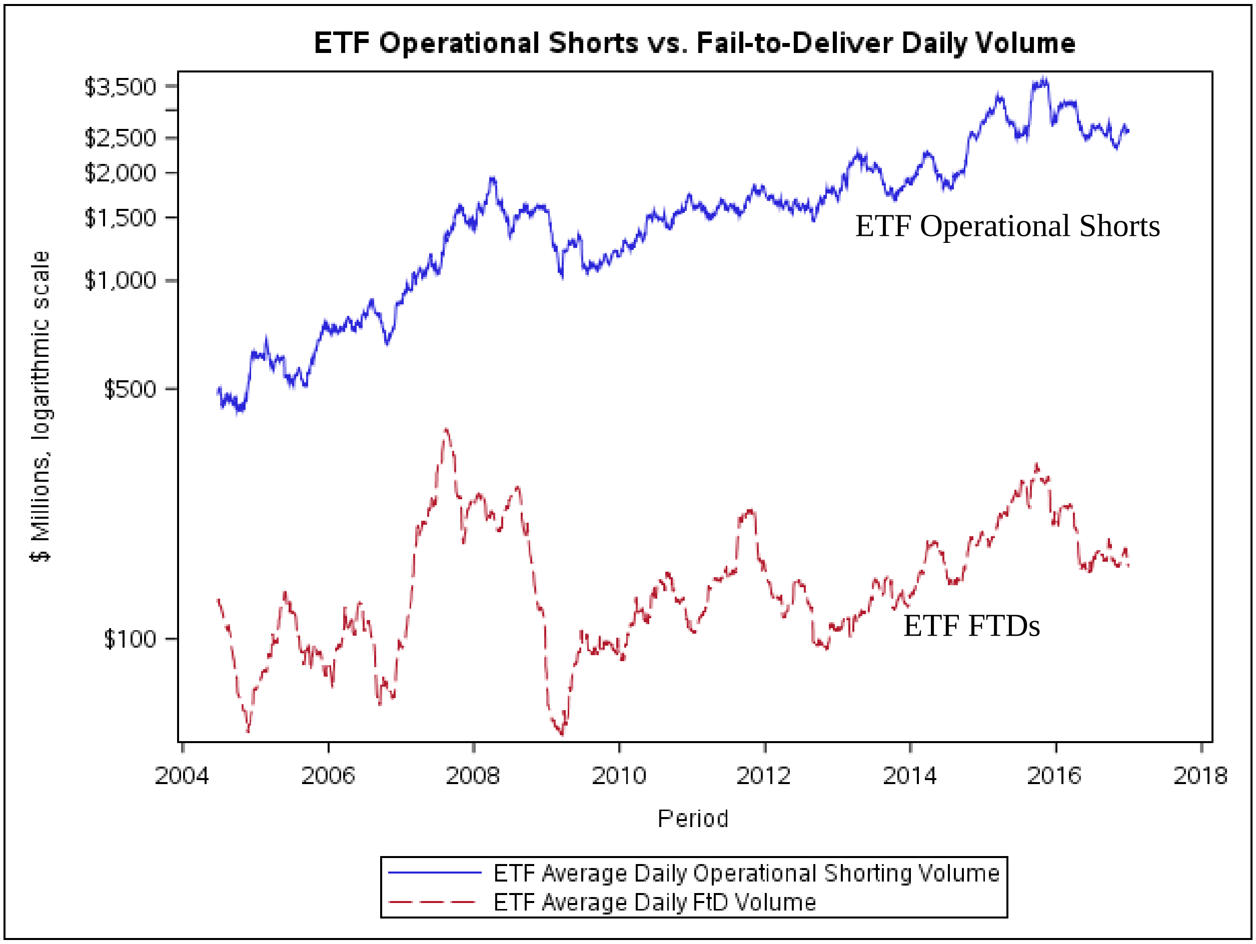
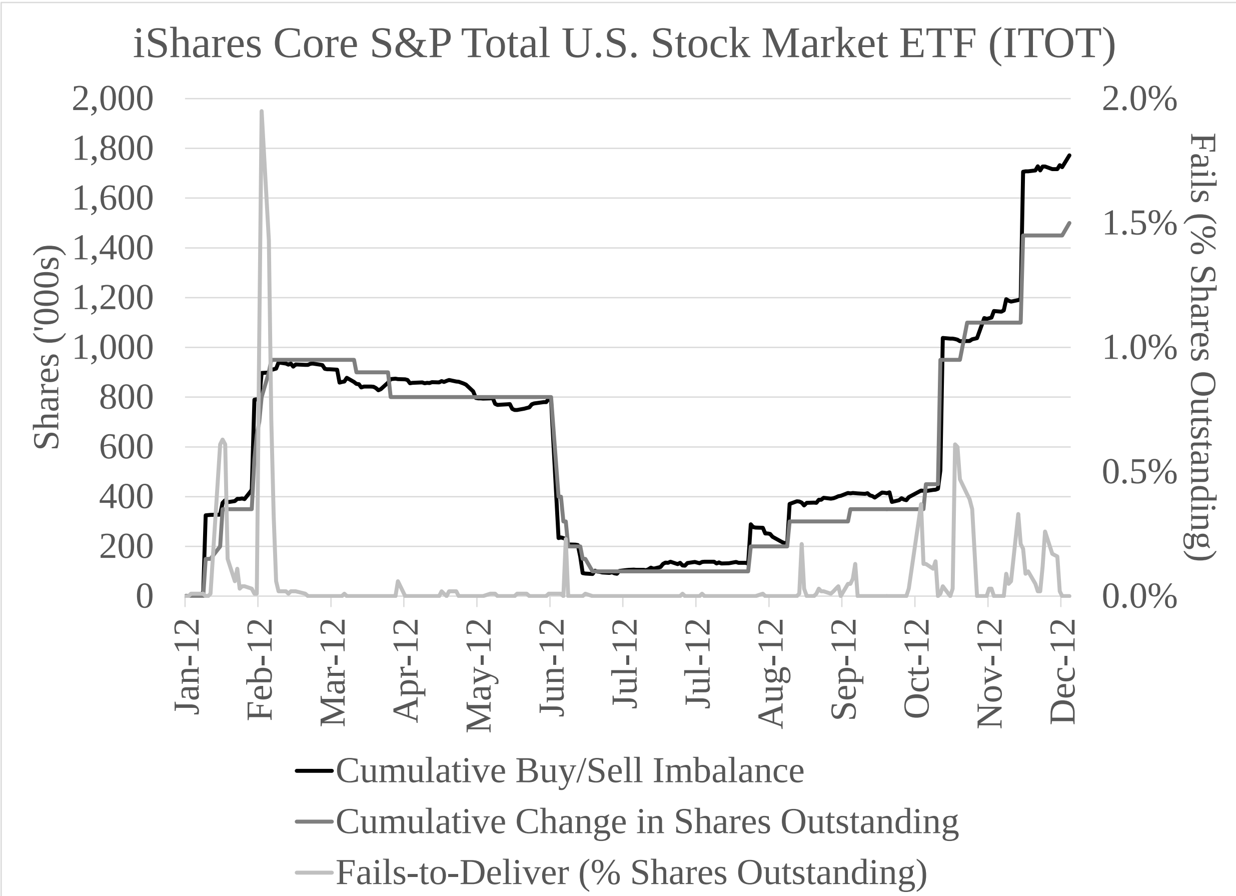


Figure 5 – An Example: ITOT – iShares Core S&P Total U.S. Stock Market ETF: This figure displays the cumulative buy-sell imbalance and the cumulative change in shares outstanding (in 1,000s of shares indexed by the left vertical axis) for the iShares Core S&P Total U.S. Stock Market ETF (ticker:ITOT) over the year 2012. Both the buy-sell imbalance and the change in shares outstanding values are set equal to 0 at the beginning of 2012 and are cumulative from that point forward. The figure also plots the ITOT failures-to-deliver as a percentage of total shares outstanding (in % indexed by the right vertical axis).



Appendix: Numerical Example of the Value of Waiting to Deliver ETF Shares

To illustrate the incentive a risk-neutral AP might have to wait and deliver shares at a later date (e.g., at T+6 days) rather than immediately creating new ETF shares to cover a short position related to an arbitrage opportunity, we have developed the following numerical example. In this model, we formulate estimates of the profit potential for two alternative strategies to cover a hypothetical short position of 100 shares: 1) sell ETF shares at time t=0 at the current market price, P_0, and then immediately place a creation unit order for 100 shares with the ETF plan sponsor by purchasing the underlying securities in the ETF basket at the current Net Asset Value (NAV_0), or 2) sell ETF shares at time t=0 at the current market price, P_0, and then enter a long futures position on the underlying ETF at t=0 with a futures price of F_0 to hedge and "lock in" an arbitrage profit today between the ETF's current market price (P_0) and the futures price, F_0. However, in this second strategy, the AP will then *wait* until t=6 to place a creation unit order for, ideally, *less than* 100 shares (thus avoiding some of the costs associated with creating these new ETF shares).[46] We refer to the first strategy as the "Short and Create" method and the second strategy as the "Short and Hedge, then Create" approach.

In order to formalize the payoffs to these two strategies, we present the following formulas:

Short and Create's profit: $$\pi' = \{(P_0 - NAV_0) - (f + \lambda)\}OIB_0 \qquad \text{(A1)}$$

Short and Hedge, then Create's profit: $$\pi = \{(P_0 - F_0) - (f + \lambda)(1 - \gamma) - c\}OIB_0 \qquad \text{(A2)}$$

where,

f = the creation unit fee (expressed as a dollar amount per ETF share),

c = the cost to hedge in the futures market (expressed as a dollar amount per ETF share),

OIB_0 = the number of shares the AP initially shorts to offset the positive buy-sell order imbalance caused by other traders' excess demand for the ETF's shares at t=0, and

λ = the "market impact" cost purchasing shares of the underlying basket of securities held by the ETF.

[46] In this set-up, we abstract away from fixed, minimum creation unit sizes and allow the AP to create ETF shares for whatever the exact amount of shares the AP has shorted. In addition, for simplicity, we assume that the explicit transaction cost for the AP to trade the ETF shares is zero (i.e., the AP does not incur any commission / brokerage costs to buy or sell the ETF).

This is also expressed as a dollar amount per ETF share and represents a linear cost for trading the underlying basket related to the AP's initial short position (OIB_0). One can view this as a cost paid to liquidity providers in the underlying securities to compensate them for their risk in trading with more informed traders, as in a Kyle (1985) model, or to cover inventory holding and order processing costs. For simplicity, we use a linear relation but a function that is convex in OIB_0 (e.g., a quadratic term) could also be used to increase the market impact costs for larger AP short positions. This alternative function would only favor waiting to deliver even further and thus we use the simpler, more conservative linear relation which allows the Short and Create strategy a better chance of out-performing the Short and Hedge, then Create strategy.

γ is the percentage of shares from the AP's short position that is expected to reverse over the 6-day waiting period. This "order reversal" parameter is a key determinant of the trade-off between the profit potentials for the two competing strategies. If $\gamma = 0$, then the AP will have to incur the market impact and creation costs on 100% of the short position and thus will cause the Short and Hedge, then Create strategy to be more costly than the Short and Create method. However, if $\gamma = 1.00$, then all of the order flow reverses over the 6-day period and the AP can simply purchase the ETF shares in the secondary market to cover the initial short position without having to incur the creation fee and market impact costs associated with creating some ETF shares by buying shares in the underlying basket of securities.

$F_0 = NAV_0 \cdot (1 + {}^{R}\!/_{365})^{(T)}$ is the futures price at t=0 which, for simplicity, is based solely on the ETF's NAV_0 and the daily risk-free rate ($R / 365$). This contract is assumed to expire exactly in T=6 days so that the futures price converges to the ETF's NAV at t=6 and the arbitrage opportunity disappears at that time as well (i.e., $F_6 = NAV_6 = P_6$ so that no arbitrage exists between the futures, NAV, and ETF prices).[47]

Since the AP is risk-neutral, the difference between the above two payoffs equals what we call the

[47] These assumptions about convergence to the same price at t=6 are made to simplify the calculations but the main insights of the model would remain unchanged if we were to allow for some divergence in these prices at the time of delivery.

"Value of Waiting."

$$\pi - \pi' = \{(NAV_0 - F_0) + (f + \lambda)\gamma - c\} \cdot OIB_0 \;=\; (\{NAV_0 - F_0 - c\} \cdot OIB_0) + (f + \lambda) \cdot OIB_0 \cdot \gamma$$

(A3)

The second equality in the above equation re-arranges the variables so that one can see that the Value of Waiting is a linear function with the first term representing a constant $((\{NAV_0 - F_0 - c\} \cdot OIB_0))$. The first term can be viewed as a constant because all of these parameters are known at t=0. The second term includes a slope $((f + \lambda) \cdot OIB_0)$ and a single independent variable (γ). Similarly, the slope term is also known at t=0. Thus, the only unknown variable in the above model is the percentage of shares which will reverse over the course of the 6-day waiting period (γ). Although this percentage could be forecasted by the AP with varying degrees of accuracy, it is not known with certainty at t=0 because market conditions and investor actions can cause γ to fluctuate over the 6-day window.

Based on Equation (A3) presented above, we create a numerical example by assuming specific values for the model's parameters and then varying the level of γ between 0 and 1.00.[48] Figure A1 displays the trade-off between the two trading strategies and shows that the Short and Create strategy is more profitable whenever γ is below 0.169 (i.e., less than 16.9% of the order flow reverses). In contrast, the Short and Hedge, then Create strategy is more profitable above this break-even value of γ. Thus, when γ is greater than 0.169, the AP will have an incentive to use a long futures position to hedge the initial short position and then wait to create ETF shares for only the portion that does not reverse (i.e., for $(1 - \gamma)$ of OIB_0). In effect, by waiting, the AP can avoid incurring the creation fee and market impact costs $(f + \lambda)$

[48] We assume the following values: P_0 = \$12.00 per share, NAV_0 = \$10.00 per share, F_0 = \$10.003 per share, λ = \$0.01 per share, c = \$0.0001 per share, f = \$.01 per share, R = .02 (i.e., 2% per year), and γ varies from 0.00 to 1.00. We also assume that the ETF's market price, NAV, and futures price all converge to \$11.00. For example, at γ = 0.40 (i.e., 40% of the order imbalance reverses), the Value of Waiting favors the Short and Hedge, then Create Strategy with a 6-day return of +0.23% in excess of the alternative Short and Create strategy. This gain is computed as a percentage of the Short and Create strategy's profit. On annualized basis, this represents a 15.20% return associated with waiting. As Figure A1 illustrates, the Value of Waiting varies greatly from -0.17% to +0.84% over the interval of γ = 0.00 to 1.00.

for that portion (γ) of the initial short position (OIB_0).

Figure A1 shows there is a clear trade-off between the two trading strategies and that the predictability of reversals in order imbalances can dictate which approach is most profitable for a specific ETF within a particular set of market conditions. Since we observe in our empirical results a large degree of operational shorting and FTD activity within ETF markets, one can surmise that the incentives to wait are more likely to outweigh the incentives to immediately create new shares to cover an AP's shorting activity. Thus, the "Value to Waiting" appears to be quite large for many APs in the U.S. ETF market. So, even though the numerical example presented here is fairly straightforward, it captures the main factors affecting the AP's decision-making process. Interestingly, our results are consistent with Nutz and Scheinkman's (2017) continuous-time model of trading among risk-neutral agents with heterogeneous beliefs when there are positive, convex costs of carrying a long position. In their model, the risky asset's supply and the associated carrying costs can interact to create situations where the "option to delay" (i.e., to wait and trade at a more favorable price in the future) affects the pricing of the asset.

One could also extend the above model in several ways. For example, although the trade-off outlined here is linear, the relationship could be nonlinear if AP's are assumed to be risk-averse and/or market impact costs are convex in the level of order imbalances. Also, another extension of the above model could incorporate order flow volatility as an alternative variable to describe the AP's uncertainty in terms of whether to choose to wait and deliver at T+6. For example, rather than use the order reversal parameter (γ), we could use the variance of order flow as another factor that affects the AP's choice between the two strategies discussed above. The above extensions are beyond the scope of the current analysis but, even if incorporated, the insights of basic model outlined here related to the trade-off between costs and benefits of the two strategies would remain intact.

Figure A1. The Value of Waiting

The chart below displays the trade-off between the payoffs to the Short and Create vs. the Short and Hedge, then Create trading strategies. The net payoff values are determined by Equation (A3) and the parameter assumptions described in the Appendix, as well as variations in the percentage of the initial order imbalance (OIB_0) that reverses over time (γ). Positive values indicate that there is an incentive for APs to wait and deliver ETF shares at the end of the 6-day trading window. Negative values represent levels of γ where the AP should not wait to deliver the shares and instead pursue the Short and Create strategy. The *% Profit Differential* is expressed as a percentage of the Short and Create strategy's profit level. The break-even point where the two strategies yield the same profit occurs when $\gamma = 0.169$ based on the model's parameter assumptions.