Security Lending Market, Secondary Market Arbitrageurs, and ETF Mispricing

Bochen Wu♣

This Version: September 2023

Abstract

This paper examines the effect of ETF short-sale costs on ETF pricing efficiency. I find that ETF premiums are positively associated with the costs of borrowing ETFs, which are primarily a friction for ETF secondary market arbitrageurs. Leveraging two exogenous variations in ETF borrowing costs, I establish a causal effect of borrowing costs on ETF mispricing. Furthermore, the sensitivity of ETF mispricing on borrowing costs depends on the activeness of primary market arbitrageurs. Collectively, empirical findings in the paper emphasize the role of the secondary market participants in the ETF arbitrage mechanism, and reveal an interdependence between the primary and secondary markets.

Keywords: ETF Mispricing, Security Lending Market, Short-Sale Constraints

JEL Classification: G12, G14, G23, G28.

♣Department of Finance, Faculty of Business and Economics, the University of Melbourne. Email Address: bochenw@student.unimelb.edu.au. I am sincerely grateful to my supervisory committee: Joachim Inkmann, Federico Nardari (Principal Supervisor), Qi Zeng, Zhuo Zhong, and Yichao Zhu, for their continuous guidance, support, and encouragement. I appreciate Richard Evans, David Frankel, Patrick Kelly, Nadia Massoud, Hedieh Rashidi Ranjbar, Gideon Saar, David Shin, Han Xiao, and seminar participants at the University of Melbourne for their helpful comments.

1 Introduction

Undoubtedly, Exchange Traded Funds (ETFs) have emerged as one of the most popular investment vehicles, with over $6.8 trillion in assets and accounting for approximately one-third of the secondary market dollar trading volume in the first quarter of 2023.[1] At the heart of ETFs' success lies a groundbreaking innovation – the arbitrage mechanism facilitated by Authorised Participants (APs), who have the ability to exchange a basket of securities for ETF shares, and vice versa, through the creation and redemption process. The novelty of APs and their crucial role in ETF pricing efficiency attract significant academic interest. The literature examining the ETF arbitrage mechanism focuses extensively on frictions in the primary market that affect the arbitrage trading of APs.[2]

However, in terms of trading volume, the ETF secondary market is way more important than the primary market. According to the 2022 Investment Company Institute (ICI) Fact Book, on average, around 85 percent of total ETF trading activities occurred on the secondary market.[3,4] Considering the importance of the ETF secondary market, it is surprising that there is a scarcity of research examining the impact of secondary market frictions on ETF pricing efficiency. One exception is Bae and Kim (2020), in which the authors document that ETF tracking errors are related to secondary market liquidity. As the authors emphasized, however, ETF secondary market liquidity

[1] The average daily trading volume exceeds $162 billion. For more information, please refer to Global ETF Market Facts in Q1 2023 provided by BlackRock, https://https://www.ishares.com/us/insights/global-etf-facts.

[2] Examples include Pan and Zeng (2019), Raddatz (2021), Gorbatikov and Sikorskaya (2022), and Zurowska (2022) among others.

[3] For some types of ETFs, the percentage of ETF secondary market trading relative to total trading activities is more than 90 percent. For instance, emerging market equity ETFs have 96 percent of trading occurred in the secondary market. For more detail, please see https://www.ici.org/system/files/2022-05/2022_factbook.pdf.

[4] The relative importance of the secondary market, compared to the primary market, is also discussed in the academic literature. For instance, Lettau and Madhavan (2018) document that, for US ETFs, the primary market is only a fraction of the secondary market. Similarly, with French data, Comerton-Forde and Marta (2021) find the trading volume of the ETF secondary market is much more than that of the primary market.

also heavily affects the willingness of APs to trade against mispricing because APs need transactions in both the primary and secondary markets to conduct ETF arbitrage.

In this paper, I examine the cost of borrowing ETF shares in the security lending market on ETF pricing efficiency. ETF borrowing costs are likely to be a friction that affects secondary market arbitrageurs rather than APs. The reason is that many APs are also ETF market makers, allowing them to establish short positions without entering the ETF lending market. Because of the exemptions from SEC delivery requirements (Rule 204), these APs can sell ETF shares that have not been created or borrowed, and postpone their creation and delivery, the ETF "operational" shorting strategy described in Evans et al. (2022). On the other hand, for the vast majority of secondary market participants, the cost of borrowing ETF shares matters because they do need to borrow ETFs before selling them short.

The intuition that the ETF borrowing costs have a negative impact on ETF pricing efficiency is straightforward. When an ETF is overpriced relative to its underlying NAV, arbitrageurs will simultaneously short the ETF and long the underlying securities. However, arbitrageurs will trade against this mispricing only if the price discrepancy is greater than the cost of arbitrage. Therefore, in equilibrium, the higher the cost of borrowing ETF shares (i.e., the arbitrage cost), the larger the ETF overpricing we will observe.[5] Given that the cost of borrowing ETFs is less likely to be a binding constraint for APs, whether ETF borrowing costs affect ETF pricing efficiency depends on whether secondary market participants actively engage in ETF arbitrage.

If secondary market participants do not trade against mispricing, they will not care about arbitrage cost, and hence any secondary market arbitrage cost is unlikely to af-

[5] This is the standard relationship between short-sale constraints and asset overvaluation. For theoretical consideration, please see, for example, Miller (1977), Jarrow (1980), and Diamond and Verrecchia (1987). For empirical evidence, please see Chen et al. (2002), Chang et al. (2007), Blocher et al. (2013) among others.

fect ETF mispricing. On the other hand, if they indeed engage in ETF arbitrage, arbitrage costs in the secondary market (ETF borrowing costs in this case) will affect their arbitrage trading, which in turn has an impact on ETF pricing efficiency. Therefore, examining the relationship between ETF borrowing costs and mispricing has a significant implication for a more fundamental question in understanding the ETF arbitrage mechanism: Do secondary market participants trade against ETF mispricing? Empirical evidence regarding this question is limited and inconclusive.[6] Therefore, whether the ETF pricing efficiency is affected by the cost of borrowing ETF is an empirical question and needs further exploration.

I provide empirical evidence supporting the hypothesis that the cost of borrowing ETF shares contributes to ETF overpricing: ETFs with the lowest borrowing costs exhibit an average mispricing of around zero, while those with the highest borrowing costs have an average mispricing of approximately 7 bps per day. After controlling for ETF- and day-fixed effects, the positive association remains highly statistically significant. These findings remain consistent after accounting for various ETF characteristics and the liquidity of underlying stocks. In terms of economic magnitude, a one-standard deviation increase in ETF borrowing costs corresponds to a one-basis point increase in ETF mispricing. Considering that the mean ETF mispricing is 1.1 bps per day, this magnitude holds economic significance.

Next, I verify that ETF borrowing costs are likely to be a friction that affects ETF secondary market arbitrageurs rather than APs. Specifically, I use the percentage change in ETF daily number of shares outstanding to measure APs' arbitrage activities.[7] I then

[6] Jain et al. (2021) show that High-Frequency Traders (HFTs) help reduce the price discrepancy between ETF market prices and underlying NAVs. On the other hand, Comerton-Forde and Marta (2021) find that the total activities of HFTs do not respond to ETF mispricing in the French ETF market, implying that the HFTs, one of the most significant players in the secondary market, do not trade against ETF mispricing.

[7] Kashner (2017) and Moussawi et al. (2022) show that ETFs rely on in-kind redemption exemption to achieve tax efficiency through "heartbeat" trades, which are characterized by a large outflow preceded by a large inflow a few days earlier. I follow descriptions in Moussawi et al. (2022) to identify "heartbeat" trades and set changes in the number of shares outstanding to zero on these days as these changes in

segment the full sample into two groups: days when APs create ETF shares (positive changes in the number of ETF shares outstanding) and days when APs do not move (no changes in the number of ETF shares outstanding). The effect of ETF short-sale cost on ETF premiums is much stronger on days when APs do not trade than on days when APs create ETF shares (trading against ETF relative overpricing), implying that ETF borrowing costs affect secondary market arbitrageurs more than APs. That ETF mispricing is positively correlated with short-sale costs on days when APs are inactive emphasizes the role of ETF secondary market investors in correcting ETF mispricing.

The observed positive relationship between ETF premiums and borrowing costs might be endogenous even though various control variables and fixed effects have been included. One plausible concern is that when ETF premiums are high, secondary market arbitrageurs heavily sell ETFs short, and drive up the borrowing costs (i.e., the reverse causality). To address this issue, I exploit two exogenous variations in the security lending market. The first exogenous variation is driven by the short-sale constraints of ETF constituent stocks. Several papers provide evidence that investors may use ETFs to bypass short-sale constraints at the stock level (e.g., Karmaziene and Sokolovski (2022) and Li and Zhu (2022) among others). To be more specific, for stocks that cannot be sold short, investors can short ETFs that contain these stocks and hedge other constituent stocks. Effectively, this strategy provides investors with similar negative exposures to these stocks.[8] In addition, Li and Zhu (2022) show that investors who have negative information about a stock and bet on stock price declines are more likely to use ETFs to circumvent underlying short-sale constraints. The borrowing demand by these investors is primarily driven by the bad news of the underlying stocks and hence is not related to ETF mispricing. On the other hand, however, this demand to borrow ETFs

shares outstanding are not related to AP arbitrage trading.

[8] A practical paper published in Barron's Magazine also mentioned this strategy and called it *Synthetic Shorting with ETFs*. The author pointed out that *Synthetic Shorting* strategy may frequently be used by hedge fund managers to circumvent short-sale constraints. For more information, please see https://www.barrons.com/amp/articles/synthetic-shorting-with-etfs-1488206009.

may push up ETF borrowing costs. Therefore, the short-sale constraints of underlying stocks provide an ideal exogenous variation in examining the effect of borrowing costs on ETF pricing efficiency.

To measure short-sale constraints at the stock level, I assume that stocks with missing lending borrowing cost data are not available for short-selling. I then utilize the percentage of constituent stocks with missing lending data as a proxy for the extent to which an ETF will be used to bypass underlying short-sale constraints. The empirical results indicate that ETF borrowing costs are positively related to short-sale constraints of underlying stocks in the cross-section, suggesting that investors using ETFs to bypass underlying short-sale constraints have a positive effect on ETF borrowing costs. More importantly, ETF borrowing costs projected by constituents' short-sale constraints contribute to ETF premiums, providing the first set of evidence that the effect of ETF borrowing costs on ETF mispricing is causal.

The second exogenous variation in the cost of borrowing ETFs is driven by ETF dividend events. A security loan transfers the legal right to receive dividends from the lender to the borrower. However, if any dividend payments occur during the loan period, the borrower is obligated to reimburse the lender. From the security lenders' point of view, while the dividend amounts are the same regardless of whether they are received directly from the firm or reimbursed by the borrower, the tax treatments differ significantly. Qualified dividends, issued to investors who have held the security for more than 60 days, are taxed as capital gains. On the other hand, reimbursed dividends are taxed at the ordinary income rate, which is higher than the capital gain tax rate for most US domestic investors. As a result, tax-sensitive security lenders have an incentive to avoid receiving reimbursed dividends. To achieve this, lenders may temporarily withdraw their shares from the security lending market or recall their loans on dividend record dates – dates when dividend ownership is established. This action reduces supply in the security lending market and drives up borrowing costs. Supporting this notion, both Thornock (2013) and Dixon et al. (2021) find lending fees abruptly spike on dividend record dates in the stock lending market. They further exploit this plausible exogenous shock in the stock lending market to test how short selling affects market quality.

Motivated by Thornock (2013) and Dixon et al. (2021), I investigate whether the ETF lending market experiences tightening on dividend record dates and, if so, whether the plausibly exogenous shifts in the ETF lending market contribute to ETF premiums. This setting is relatively clean to examine the effect of short-sale costs on ETF overpricing, as there is no additional information on dividend record dates. Any information related to ETF dividends should already be incorporated into ETF prices on dividend announcement dates. Empirically, in line with previous findings in the stock lending market, I observe a spike in ETF borrowing costs on ETF dividend record dates. Moreover, ETF premiums are significantly higher on these days. On average, ETF premiums are approximately three bps higher on dividend record dates compared to other days within a 30-day window around the dividend record day ($t - 15, t + 15$). These results remain consistent even after controlling for various fixed effects and accounting for the demand for dividends. I then employ a two-stage least squares (2SLS) regression using the dividend record day as an instrumental variable. The analysis confirms that elevated borrowing costs on dividend record dates lead to higher ETF overpricing.

To shed more light on how secondary market arbitrageurs trade against mispricing, in the last part of this paper, I investigate the interaction between the primary and secondary market arbitrageurs. I hypothesize that when APs are not active, secondary market arbitrageurs will not actively engage in correcting ETF mispricing either. The intuition is that secondary market arbitrageurs are essentially betting on the convergence of the ETF market prices and NAVs. On the other hand, the creation/redemption mechanism ensures that APs do not rely on price convergence, granting them an advantage in correcting ETF mispricing.[9] As a consequence, if the primary market is not active, secondary market arbitrageurs bear higher risks that the price-NAV divergence does not converge. Hence, secondary market arbitrageurs do not have sufficient incentives to correct mispricing. An extreme case is that, in the absence of APs, mispricing could become substantial and persistent (one example would be the closed-end discount puzzle discussed in Lee et al. (1991), Pontiff (1996) and many others). As APs become more active in correcting mispricing, secondary market arbitrageurs also become more engaged, as the convergence risk decreases.[10] However, when APs are highly active, fewer arbitrage opportunities remain for secondary market arbitrageurs, resulting in a reduced role for them in ETF pricing efficiency.

Empirically, I use two proxies to measure the primary market activeness: the number of APs and the standard deviation of percentage changes in the ETF daily number of shares outstanding. Then, I sort ETFs into three groups based on the level of primary market activeness.[11] For each group, I regress daily mispricing onto ETF borrowing costs (plus a variety of control variables and ETF- and Day-fixed effects). Empirical findings are consistent with my hypothesis. Specifically, when APs are operating at the lowest level of activeness, the sensitivity of ETF mispricing to borrowing costs is relatively low (if statistically significant), indicating that secondary market arbitrageurs are not active, resulting in a limited impact on ETF pricing efficiency. As APs reach a medium level of activeness, this sensitivity experiences a significant increase. This result implies that secondary market arbitrageurs are more active and bring more arbitrage capital, which in turn has a larger impact on ETF mispricing. However, as APs

[9] Observing that the ETF price is higher than the underlying NAV, APs will immediately short ETFs and purchase the underlying stocks to lock in the arbitrage profits within the trading day. At the end of the day, APs will convert the purchased basket of underlying stocks to ETFs and unwind their short positions in ETFs. Therefore, APs do not rely on the convergence of ETF price-NAV to realize arbitrage profits.

[10] In unreported tests, I confirm that the speed of ETF price-NAV convergence is indeed faster when APs are more active.

[11] Results strengthen when dividing ETFs into five groups based on primary market activeness.

become highly active, this sensitivity diminishes again, implying that secondary market arbitrageurs become less active and consequently have a small impact on ETF pricing efficiency.

This study provides several contributions to the literature investigating the ETF arbitrage mechanism. Prior research, both theoretical and empirical, has primarily focused on the role of APs in ensuring ETF pricing efficiency, largely overlooking the secondary market arbitrageurs. Theoretically, it is typically assumed that APs are the sole participants engaged in ETF arbitrage activities.[12] Empirical studies provide evidence suggesting constraints of APs negatively affect ETF pricing efficiency. Pan and Zeng (2019) show that shocks to the balance sheet of APs negatively impact the role of APs in correcting ETF mispricing. Similarly, Raddatz (2021) find that APs with high leverage respond less to ETF mispricing. Gorbatikov and Sikorskaya (2022) argue that arbitrage costs in the primary market are AP-specific and demonstrate that the diversity of APs helps improve ETF pricing efficiency. Zurowska (2022) argue that ETF creation unit size determines an AP's entry decision, which in turn has an impact on ETF pricing efficiency. This is the first paper providing causal evidence that the friction in the ETF secondary market (the cost of borrowing ETFs) negatively affects ETF pricing efficiency.

The literature on whether secondary market participants trade against ETF mispricing is limited and inconclusive. While Jain et al. (2021) document a positive relation between High-Frequency trading and ETF pricing efficiency in the US, Comerton-Forde and Marta (2021) use data from France and find that the total activity of High-Frequency traders, who are responsible for a significant portion of secondary market trading, does not respond to ETF mispricing. The result that the secondary market arbitrage cost (ETF borrowing cost) contributes to ETF mispricing implies that secondary

[12] Examples include Malamud (2016), Bhattacharya and O'Hara (2018), Pan and Zeng (2019), and Gorbatikov and Sikorskaya (2022).

market participants do trade against mispricing. Otherwise, they will not be affected by arbitrage costs. Therefore, I contribute to this literature by providing evidence (albeit indirect) that highlights the role of secondary market arbitrageurs in correcting ETF mispricing. More importantly, I provide novel insights into the the dynamics of ETF arbitrage mechanism by uncovering an interdependence between the primary and secondary markets. This finding suggests that future theoretical works may consider the interplay between APs and secondary market arbitrageurs to enhance our understanding of the ETF arbitrage mechanism.

More broadly, this study contributes to the literature examining short-sale constraints and asset prices. Theoretical discussion regarding the effect of short-sale constraints on asset prices dates back several decades (see, for example, Miller (1977), Jarrow (1980), and Diamond and Verrecchia (1987)). Recent empirical works in the stock market support the view that short-sale constraints are related to asset overvaluation (Chen et al. (2002), Chang et al. (2007), Blocher et al. (2013) among others). In this paper, I examine the relation between short-sale constraints and asset overvaluation using ETFs as a laboratory, in which overpricing can be accurately identified. By exploiting two exogenous variations in the ETF lending market, I provide causal evidence that short-sale constraints lead to asset overvaluation.

2 Data and Sample

2.1 Data Sources

To construct the sample used in this paper, I first extract U.S. Equity ETFs from the CRSP Mutual Fund Database.[13] Specifically, to find U.S. domestic mutual funds, I require that the first two characters in the CRSP Objective Codes (*crsp_obj_cd*) to be

[13] I omit bond ETFs from my analysis is to avoid the liquidity mismatch problem discussed in Pan and Zeng (2019).

"ED" and to identify ETFs, I use the variable *et_flag*.[14] ETF basic information such as expense ratio and daily NAV data are also from CRSP Mutual Fund Database. ETF holdings data is from Thompson Reuters Mutual Fund Holdings (S12). ETF and the underlying stock's borrowing cost data are from Markit Securities Finance (formerly known as Data Explorer) database. In addition, Daily prices and returns of ETFs and underlying stocks were retrieved from the CRSP Stock/Security files. The daily number of shares outstanding for ETFs is downloaded from Refinitiv DataScope.

To gather ETF primary market information, I accessed the SEC Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system to collect ETF N-CEN and N-CEN/A filings. Since 2018, ETF products have been required to file N-CEN Forms as part of their annual reports under rule 30a-1 under the Act (17 CFR 270.30a-1) by management companies.[15] ETF primary market information is found in Part E of the form, which includes names and legal entity identifiers (LEIs) for all registered authorized participants for each fund, even if these APs are not active (i.e., there is no creation/redemption transaction within a reporting period). In addition, for each AP within each ETF, the form also provides the dollar value of created and redeemed ETF shares. Lastly, for each ETF, creation/redemption units (the minimum number of ETF shares required for each creation/redemption transaction) and creation/redemption fees can also be observed. The N-CEN filing has also been used in Gorbatikov and Sikorskaya (2022) and Xiao (2022). Following these two papers, for each reporting period, I select the last available filing.

The final sample with ETF-day observations consists of 1,587 unique ETFs from July 2001 to September 2022.

[14] According to CRSP, "E" stands for Equity and "D" stands for Domestics. For more information about *crsp_obj_cd*, please refer to https://www.crsp.org/products/documentation/crsp-style-code-0.

[15] For a detailed description of N-CEN Form, please see https://www.sec.gov/files/formn-cen.pdf.

2.2 Variable Construction

Daily ETF relative mispricing (**Mispricing**), is defined as the percentage difference between ETF market price and NAV. An ETF premium (discount) indicates that the ETF price is higher (lower) than its NAV. Following Huang et al. (2021) and Li and Zhu (2022), the cost of borrowing securities is measured by Markit Security Finance Daily Cost of Borrow Score (**DCBS**), which is a number from 1 to 10. The DCBS reflects the rebate/fee charged by the agent lender, where 1 represents the cheapest borrowing cost and 10 represents the most expensive. For robustness, I also use two more borrowing cost variables from Markit: **vwafs1d** (value-weighted average fee for all new trades on the most recent day) and **vwafs7d** (value-weighted average fee for all new trades over the most recent 7 calendar days). In these variables, a value of 0 indicates the cheapest borrowing cost, while 5 represents the most expensive.[16]

In my empirical analysis, I incorporate various control variables to account for different ETF characteristics. **Expense Ratio** represents how much investors pay for the fund's operating expenses. **Bid-Ask Spread** is measured by the difference between ETF daily ask and bid prices, scaled by the mid-quote. **Ln(Turnover)** is the natural log of the number of ETF shares traded divided by the total number of shares outstanding. **Ln(Size)** is the natural log of ETF market capitalization, calculated as ETF price multiplied by the total number of shares outstanding. **Volatility** is the standard deviation of daily ETF returns over the past 22 trading days.[17]

I also control for several measures to capture the ETF primary market characteristics. **Number of APs** is the number of total registered APs for an ETF. **Number of Active APs** is the number of APs that have a non-zero dollar value of creation/redemption. **Creation Unit** is the minimum number of ETF shares that an AP has to exchange with

[16] Notebaly, in unreported tests, the results are qualitatively the same if I use either **vwafs1d** or **vwafs7d** to measure ETF borrowing cost, instead of **DCBS**. These results are available upon request.

[17] I require there should be at least 15 non-missing observations to compute ETF return volatility. Otherwise, it is set to be missing.

the ETF sponsor for each creation transaction. **Creation Fee** is the simple average of "in-kind" creation and cash creation fees.

2.3 Summary Statistics

[Insert Table 1]

Table 1 presents the summary statistics for the key variables used in the empirical analysis. The mean and median of ETF mispricing are both positive, approximately 1.1 and 0.81 basis points per day, respectively. However, there is a notable standard deviation in mispricing, which is around 37 basis points per day. In addition, the distribution of ETF mispricing is slightly skewed to the right. The average daily ETF borrowing cost (DCBS) is about 3. According to Blocher and Whaley (2015), stocks with a DCBS of 3 have a mean borrowing fee of 318 bps. Additionally, stocks with DCBS values of 1 and 10 correspond to mean loan fees of 36 and 5,278 basis points, respectively.[18] Furthermore, the distribution of ETF liquidity is positively skewed, with the mean bid-ask spread (25 bps) being more than twice the median spread (11 bps). In addition, ETF secondary turnover is highly right-skewed, with a mean of 100% of the number of shares outstanding and a median of only 0.6%. This is consistent with the findings in Bae and Kim (2020), in which the authors document that the most popular three ETFs account for almost half of the dollar trading volume in the ETF secondary market at the end of 2012. Not surprisingly, ETF size (market capitalization) is also highly positively skewed. The mean ETF size is around $3 billion whereas the median is approximately $264 million.

Turning to the ETF primary market, the mean (median) ETF has around 27 (26)

[18] For a more detailed comparison between DCBS and actual loan fees, please refer to the Table III in Blocher and Whaley (2015).

authorized participants but most of them are inactive. The median ETF has only 5 active APs. This observation is largely on par with what Gorbatikov and Sikorskaya (2022) has documented in their paper: the majority of ETF-AP links are inactive. In addition, the median ETF has a creation unit of 25,000 ETF shares. Lastly, creation fees appear to be relatively low, with both "in-kind" and cash creation fees averaging around 4.5 basis points per creation unit.

3 Empirical Results

Section 3.1 provides robust evidence suggesting that there is a positive relation between ETF borrowing cost and ETF mispricing. Section 3.2 demonstrates that the cost of borrowing ETFs is likely to be a friction that affects secondary market arbitrageurs more than APs. Leveraging on two exogenous variations in ETF borrowing costs, Section 3.3 shows that the effect of ETF borrowing costs on ETF mispricing is causal. Evidence in 3.4 indicates that there is an interaction between APs and secondary market arbitrageurs.

3.1 ETF Borrowing Cost and ETF Mispricing

Observing that an ETF is traded at a price higher than the underlying NAV, arbitrageurs would simultaneously short the ETF and purchase the underlying basket of stocks simultaneously until the profit is not significant enough to cover the corresponding cost. The profitability of this arbitrage trade is determined by both the degree of ETF price-NAV divergence and the costs incurred by the arbitrage strategy. Consequently, it should be expected that the higher the cost of borrowing ETFs, the wider the required ETF price-NAV deviation, resulting in higher ETF premiums.

Figure 1 illustrates the relationship between the proportion of ETF relative overvaluation and ETF borrowing costs, proxied by the Markit Security Finance Daily Cost of Borrow Score (DCBS). For each DCBS group (from 1 to 10), the proportion of ETF overpricing is computed as the number of ETFs traded at a price greater than their NAVs, divided by the total number of ETFs in that group. For ETFs with DCBS of 1 and 2, the proportion of overpricing is roughly 50%, indicating that both premium and discount are equally likely for low-shorting cost ETFs. Furthermore, as the borrowing cost (DCBS) increases, the proportion of ETF over-valuation monotonically increases. For ETFs with the most expensive borrowing costs, the probability of relative overpricing is approximately 65%, which is much greater than for ETFs with the cheapest cost of short selling.

[Insert Figure 1]

[Insert Figure 2]

Figure 2 plots the average ETF mispricing and confidence intervals by DCBS, the ETF borrowing costs. For each DCBS group, I compute the average daily ETF mispricing across all ETFs within that DCBS group. For ETFs with DCBS of 1 and 2, ETF premiums are approximately zero. On the other hand, as DCBS increases, ETF premiums increase dramatically and almost monotonically except for ETFs with a DCBS score of 6. For ETFs with a DCBS of 10, the premium is around 7 bps per day on average.

[Insert Table 2]

In what follows, I employ regressions to confirm the positive association between DCBS and ETF mispricing by accounting for a variety of control variables and fixed effects. Prior to presenting the regression results, I describe how DCBS is related to other variables discussed in section 2.2. Table 2 reports summary statistics by DCBS.

Consistent with Figure 2, ETF mispricing is positively associated with DCBS. In addition, ETFs with high borrowing costs tend to have higher expense ratios, although the relationship is not monotonic. In addition, ETFs with the lowest borrowing costs have disproportionately lower bid-ask spreads; and ETFs with higher short-sale costs have significantly higher bid-ask spreads or worse liquidity. The results are consistent with the idea that short-selling provides liquidity in the market (Comerton-Forde et al., 2016; Choi et al., 2009). If low short-sale costs facilitate short selling in the ETF market, it is not surprising to observe better liquidity for ETFs with low borrowing costs. Furthermore, the relationship between DCBS and ETF turnover in the secondary market is not linear. ETFs with moderate borrowing costs tend to exhibit higher turnover, whereas high-borrowing cost ETFs have disproportionately lower turnover. Additionally, ETF borrowing costs display a negative relationship with size (market capitalization). Regarding ETF primary market characteristics, low-borrowing cost ETFs tend to have more APs. More importantly, low-borrowing cost ETFs have more active APs on average. According to Gorbatikov and Sikorskaya (2022), ETFs with more active APs tend to have smaller mispricing in absolute value. Therefore, it is necessary to control for the number of active APs to rule out the possibility that the positive relation between DCBS and ETF mispricing is explained by the activeness of the ETF primary market. Lastly, the ETF creation unit seems to be negatively related to DCBS.

I then run the following regression to confirm the positive relationship between DCBS and ETF mispricing:

$$Mispricing_{i,t} = \beta_1 DCBS_{i,t} + \boldsymbol{X_{i,t}\beta} + \alpha_i + \gamma_t + \epsilon_{i,t} \tag{1}$$

where $DCBS_{i,t}$ is the measurement of ETF borrowing cost for ETF i on day t; $\boldsymbol{X_{i,t}}$ is a vector of control variables (including ETF expense ratio, ETF bid-ask spread, the natural log of ETF turnover in the secondary market, the natural log of ETF capitalization, and the volatility of daily ETF returns). In addition, α_i and γ_t represent the ETF and day fixed effect, respectively.

[Insert Table 3]

The results are reported in Table 3. Numbers in parentheses are t-statistics based on double-clustered standard errors at ETF and day levels. Across all specifications, estimated β_1s are positive and highly statistically significant at the one percent level. In terms of economic magnitude, a one-standard deviation increase in DCBS corresponds to approximately one-basis point increase in ETF mispricing. This effect holds economic significance, considering that the mean ETF mispricing is around 1.1 bps. In addition, both ETF bid-ask spread and turnover appear to have a positive effect on ETF mispricing although after controlling for ETF fixed-effect, the coefficient of bid-ask spread becomes statistically insignificant.

There are long and short legs in an arbitrage trading when ETFs are traded at a premium: short positions in ETF and long positions in the underlying securities. When the underlying stocks are difficult to purchase (perhaps because of the poor liquidity), the relative ETF overpricing would be hard to be corrected. To rule out this possibility, I control for the liquidity of the underlyings, which is measured by the average bid-ask spread of ETF holdings. Additionally, as shown by Gorbatikov and Sikorskaya (2022), ETFs with more active primary market have lower mispricing. As a consequence, I control for both the number of APs and the number of active APs. Lastly, I also control for creation unit and creation fees, which represent arbitrage cost in the primary market.

[Insert Table 4]

The results are presented in Table 4, demonstrating consistent results across all specifications: After controlling for both average underlying liquidity and ETF primary market characteristics, the coefficient of DCBS remains positive and highly statistically significant. Two primary market activeness measures are negatively related to ETF mispricing. In addition, creation fee, the direct measure of primary market arbitrage cost, is also positively correlated with ETF mispricing although the coefficient of creation fee becomes statistically insignificant when ETF fixed-effects are included. The interpretation of the positive relation between creation fee and ETF premium is straightforward. When ETFs are traded at a premium, APs short ETFs and buy the underlying securities. At the end of the trading day, the APs then use the purchased constituents to exchange for ETF shares (i.e., to create ETF shares). Therefore, the creation fees, which increase AP's break-even condition, should be positively related to ETF premium.

3.2 The Role of Secondary Market Arbitrageurs

Having established a robust relation between borrowing costs and ETF mispricing, in this section, I try to verify that the cost of borrowing ETFs mainly affects secondary market arbitrageurs (e.g., hedge funds and high-frequency traders), rather than the primary market arbitrageurs, the APs. Many authorized participants (but not all), players in the primary market, are also market makers. The exemption from SEC delivery requirements (Rule 204) allows market makers to sell ETFs to satisfy buy orders but postpone the creation and delivery of ETF shares. Therefore, APs who are also market makers possess the ability to establish short positions in ETF shares without the need to borrow ETF shares in the security lending market. As such, these APs are likely to be immune from changes in ETF short-selling costs. On the other hand, secondary market arbitrageurs always need to borrow ETF shares to make arbitrage profits when ETFs are overpriced relative to their NAVs. Consequently, secondary market arbitrageurs are expected to be more sensitive to the tightness in the ETF lending market than APs.

To formally test this idea, following Pan and Zeng (2019) and many others in this literature, I use the percentage changes in the daily number of ETF shares outstanding to measure the arbitrage trading of APs. Specifically, a positive (negative) change in the number of shares outstanding indicates that APs create (redeem) ETF shares. However, according to Moussawi et al. (2022), many APs help ETFs avoid distributing realized capital gains and reduce their tax overhang through "heartbeat" trades, which are characterized by a large outflow preceded by a large inflow a few days earlier. These large changes in the number of ETF shares outstanding thus do not correspond to APs' arbitrage trading. I follow Moussawi et al. (2022) to identify "heartbeat" trades and set these changes in the number of shares outstanding to zero. Recall that when observing ETF relative overpricing, APs short ETFs and buy the underlying basket of stocks within a trading day. Then, at the end of the day, APs use the purchased underlying stocks to exchange for ETF shares (create ETF shares) and unwind short positions in ETFs. Therefore, ETF share creation (redemption) corresponds to AP's trading against ETF overpricing (underpricing). As a result, I divide the sample into two sub-samples: days when there is an increase in the number of shares outstanding, and days when there is no change in the number of shares outstanding. The first (second) sub-sample then represents the one in which APs are active (inactive) in correct ETF relative overpricing.

[Insert Table 5]

Next, for each sub-sample, I re-run regression model (1) and the results are presented in Table 5. The left (right) two columns report results for the sub-sample in which APs create (do not create) ETF shares. Although the positive coefficients of DCBS are statistically significant in both sub-samples, the magnitude is much larger for the sub-sample when APs do not move than the one when APs create ETF shares.

In the sub-sample where APs create ETF shares, a one-standard deviation increase in DCBS corresponds to an approximate 0.44-bps increase in ETF mispricing. In sharp contrast, when APs do not move, a one-standard deviation increase in DCBS is associated with a 1.33-bps increase in ETF premiums. In terms of economic magnitude, the effect of DCBS on ETF mispricing is around three times larger when APs do not engage in ETF creation compared to when they do. The results suggest that the cost of borrowing ETFs affects secondary market arbitrageurs much more than APs.

In reality, however, it typically takes three days from AP's submission of creation or redemption order to the actual settlement (T+3).[19] This means that even though APs submit an order to create ETF shares, the actual change in the number of shares outstanding may not occur until a few days later. Furthermore, while most market participants must settle in T+3, market makers have up to T+6 days (Evans et al., 2022). Therefore, to rule out the possibility that the changes in the number of shares outstanding on day T reflect the stale creation order submitted a few days earlier, I lead changes in the number of shares outstanding from one day to six days, and conduct the same analysis. The results, as reported in Appendix Table A2, are consistent with the findings in Table 5: The effect of DCBS on ETF mispricing is more pronounced for secondary market arbitrageurs (i.e., when APs do not move). Both theoretically and empirically, prior research primarily focuses on the role of APs in the ETF arbitrage mechanism.[20] Indeed, the most significant innovation of ETF is AP's unique role in correcting price-NAV discrepancy. On the other hand, however, according to the Investment Company Institute estimates, approximately 85% of ETF trades correspond to secondary market activity.[21] That the arbitrage cost is associated with ETF mispricing when APs do not

[19] For more information regarding the mechanism of creation/redemption in the real market, please refer to Antoniewicz and Heinrichs (2014).

[20] Examples include Malamud (2016), Bhattacharya and O'Hara (2018), Pan and Zeng (2019), Gorbatikov and Sikorskaya (2022) for theoretical works; and Pan and Zeng (2019), Gorbatikov and Sikorskaya (2022), and Shim and Todorov (2022) for empirical works.

[21] For more detailed ETF market statistics, please see 2022 Investment Company Factbook: https://www.ici.org/doc-server/pdf%3A2022_factbook.pdf.

move highlights the importance of ETF secondary market participants in correcting ETF mispricing, which is largely ignored in the literature.

3.3 Causal Evidence

In the previous section, although I have controlled for both ETF- and day-fixed effects, time-varying unobservable factors that correlate with both ETF borrowing costs and ETF mispricing might be omitted. Furthermore, the observed positive correlation between ETF premiums and borrowing costs is also subject to reverse causality. That is, when ETFs experience large premiums, arbitrageurs would heavily sell ETFs short, pushing the borrowing costs high. In this section, I explore two exogenous variations in ETF borrowing costs and show that the effect of borrowing costs on ETF mispricing is causal.

3.3.1 Underlying Short-Sale Constraints

The first exogenous variation in ETF borrowing costs is driven by the investors who use ETFs to bypass short-sale constraints of underlying stocks. Several papers provide evidence that ETFs help investors circumvent short-sale constraints of underlying stocks through the *Synthetic Shorting* strategy (Karmaziene and Sokolovski, 2022; Li and Zhu, 2022). Specifically, if some stocks cannot be sold short, investors can short ETFs that contain these stocks and go long the other constituent stocks. Effectively, the *Synthetic Shorting* strategy provides investors with similar negative exposures on these constrained stocks. Li and Zhu (2022) show that investors who have bad news about stocks and bet on stock price declines are more likely to use ETFs to circumvent the underlying short-sale constraints. In particular, Li and Zhu (2022) show that stocks that are heavily shorted via their holding ETFs underperform lightly shorted ones. Furthermore, the predictability concentrates on stocks that face short-sale constraints.

The demand for borrowing ETFs driven by investors who would like to short underlying stocks originates from the bad news of these stocks and hence is not related to ETF mispricing. On the other hand, this demand may push up ETF borrowing costs. Therefore, if the demand to use ETFs to circumvent underlying short-sale constraints is related to ETF borrowing costs, then ETF borrowing costs that are related to underlying short-sale constraints are exogenous. I first examine whether ETF borrowing costs are indeed related to underlying short-sale constraints.

To quantify the level of short-sale constraints on underlying stocks, I adopt a measure based on the availability of borrowing costs (DCBS) in the Markit database. For each stock j on day t, if the borrowing cost is missing, I assume that the stock cannot be sold short on that day. Next, for each ETF i on day t, I calculate the proportion of ETF holdings that have missing DCBS. Therefore, ETFs with a higher proportion of missing underlying DCBS are expected to have more short-sale constrained stocks in their holdings; and these ETFs are more likely to be employed as a tool to short the underlying stocks.

[Insert Figure 3]

[Insert Figure 4]

I divide ETFs into five groups based on the proportion of missing underlying DCBS and then calculate the average ETF DCBS for each group. Figure 3 presents the relation between the percentage of missing underlying DCBS and ETF borrowing costs. The average ETF borrowing costs increase with the percentage of underlying stocks with missing DCBS. This positive relation is consistent with the notion that investors who use ETFs to circumvent underlying stocks' short-sale constraints push up ETF short-selling costs. If the cost of borrowing ETFs leads to ETF overpricing, it should be observed that sorting ETFs based on underlying short-sale constraints would generate spreads in ETF mispricing. The relation between the proportion of missing underlying DCBS and ETF mispricing is illustrated in Figure 4. ETFs with the lowest proportion of missing underlying DCBS have an average mispricing of roughly zero. On the contrary, those with the highest proportion of missing underlying DCBS on average are traded at a premium of 2 bps per day. This observation lends support to the conjecture that investors who overcome stock-level short-sale constraints through ETFs hinder ETF pricing efficiency.

[Insert Table 6]

[Insert Table 7]

Next, I use regressions to confirm the cost of borrowing ETFs is positively related to underlying short-sale constraints. Given that the within-ETF variation in the proportion of missing underlying DCBS is small, I examine this relation in the cross-section and run the following Fama and MacBeth (1973) regression:

$$DCBS_{i,t} = \alpha + \beta_1 Missing(\%)_{i,t} + \boldsymbol{X_{i,t}}\beta + \epsilon_{i,t} \tag{2}$$

where $DCBS_{i,t}$ is the borrowing cost for ETF i on day t, $Missing(\%)_{i,t}$ is the proportion of holdings that have missing borrowing cost data for ETF i on day t, and $\boldsymbol{X_{i,t}}$ is a set of control variables, which include ETF utilization, volatility, bid-ask spread, size, and turnover. The results are reported in Table 6. The left two columns represent the OLS benchmark and the right two columns represent Fama-Macbeth estimates. The coefficient of $Missing(\%)$ is positive and highly statistically significant. In terms of economic magnitude, a one-standard deviation increase in $Missing(\%)$ corresponds to a 0.23 increase in ETF's DCBS. Considering that the standard deviation of DCBS is 1.77, this effect is economically meaningful.

To investigate the potential impact of demand from investors who use ETFs to circumvent stock-level short-sale constraints on ETF pricing efficiency, I first project ETF DCBS onto the proportion of missing underlying DCBS day by day and collect fitted values. Then, I run the following Fama-Macbeth regression:

$$Mispricing_{i,t} = \alpha + \beta_1 \widehat{DCBS}_{i,t} + \boldsymbol{X_{i,t}}\beta + \epsilon_{i,t} \tag{3}$$

where $\widehat{DCBS}_{i,t}$ is the fitted value obtained from regressing ETF DCBS on the proportion of missing underlying DCBS, and $\boldsymbol{X_{i,t}}$ is a vector of control variables. The results are reported in Table 7. Again, the coefficient of $\widehat{DCBS}$ is positive and statistically significant at least at the 10 percent level. A one-standard deviation increase in $\widehat{DCBS}$ corresponds to roughly 6-bps increases in ETF premiums, demonstrating economic significance.

Collectively, I interpret the result that ETF borrowing costs projected by underlying short-sale constraints are positively related to ETF mispricing as causal evidence demonstrating that ETF borrowing costs lead to ETF overpricing.

3.3.2 ETF Dividend Events

The second exogenous variation in the ETF lending market is driven by ETF dividend events, which is first discussed in Thornock (2013). In the security lending market, a security loan transfers the right to receive dividends from the lender to the borrower. However, the borrower is obligated to reimburse the lender with the same amount as they received.[22] Although the total amount received by the security lender would be the same regardless of whether the dividends are received from the firm directly or reimbursed by the borrower, tax treatments differ significantly. Under the Jobs and Growth Tax Relief Reconciliation Act of 2003, qualified dividends, which are paid

[22]This reimbursed dividend payment is also known as a substitute dividend.

to investors who have held the securities for at least 60 days, are subject to a favorable capital gain tax rate of 15%. On the other hand, reimbursed dividends lose this qualification and are taxed at the investor's ordinary income tax rates, which can be as high as 35%. As a consequence, security lenders have an incentive to temporarily withdraw their shares from the lending market or recall existing loans around dividend record days if the tax levied on reimbursed dividends exceeds that on dividends directly paid by the firm. Therefore, the tax disadvantage of reimbursed dividend payments tightens the supply in security lending market and pushes the cost of borrowing securities high around dividend record dates, which establish the ownership of dividends.

Empirically, both Thornock (2013) and Dixon et al. (2021) confirm that lending supply experiences a dramatic drop around dividend record dates. In addition, Dixon et al. (2021) also document an increase in demand around dividend record dates because investors who have tax advantages may borrow shares to conduct dividend arbitrage.[23] By exploiting this exogenous shock in the security lending market, Thornock (2013) and Dixon et al. (2021) demonstrate that tightness in the lending market adversely affects market quality. In addition, with the same setting, Blocher et al. (2013) find that, around dividend record dates, there is a significant increase in abnormal returns for hard-to-borrow stocks.

[Insert Figure 5]

[Insert Figure 6]

Following these papers, I employ the tax-driven shock in the ETF lending market to explore the causal relationship between short-selling costs and ETF premiums.

[23] Investors with favorable tax status for reimbursed dividends may borrow shares, receive the dividends, and divide the tax savings with the original owner (typically foreign institutions). McDonald (2001) and Christoffersen et al. (2005) documents evidence of dividend arbitrage in Germany and Canada, respectively.

Firstly, I establish the presence of increased tightness in the ETF lending market around dividend record dates. Figure 5 plots the average DCBS (ETF borrowing costs) over a 30-day window $(t-15, t+15)$ surrounding ETF dividend record dates. This figure clearly depicts a notable surge in ETF borrowing costs on ETF dividend record dates. If the cost of borrowing ETF shares leads to ETF relative overpricing, we should observe a spike in ETF premiums on dividend record days. Figure 6 plots exactly the same pattern: ETF premiums sharply increase on dividend record dates.

[Insert Table 8]

Next, I run the following regression for the 30-day $(t-15, t+15)$ window surrounding ETF dividend record day:

$$Mispricing_{i,t} = \beta_1 \times RecordDay_{i,t} + \boldsymbol{X_{i,t}\beta} + \alpha_i + \gamma_{i,d} + \epsilon_{i,t} \tag{4}$$

where $Mispricing_{i,t}$ is the mispricing for ETF i on day t, $RecordDay_{i,t}$ is the dummy variable which is equal to one if day t is the dividend record day for ETF i and zero otherwise, $\boldsymbol{X_{i,t}}$ is a vector of control variables, α_i is the ETF fixed effects, and $\gamma_{i,d}$ is the ETF-Dividend fixed effects.

Table 8 presents OLS estimates for regression (4). On dividend record days, there is an approximately 3-bps increase in ETF premiums on average. The results are highly statistically significant for all specifications regardless of the inclusion of a battery of control variables and fixed effects. According to Gromb and Vayanos (2010), there are two building blocks to explain asset mis-valuation: one is the demand shock and the other one is the limits to arbitrage. What I am trying to demonstrate is that, on dividend record days, an exogenous increase in ETF borrowing costs drives ETF premiums (the limit-to-arbitrage channel). However, as observed in Figure 6, there is an increase in

ETF premium before dividend record dates. This probably represents the dividend capture trading. To rule out this demand shock channel, I introduce three proxies to measure the demand for ETF dividends: ETF returns on the cum-dividend day (the day before the ex-dividend day), the three-day, and the five-day buy-and-hold returns up to the cum-dividend day. The results (Model 3 and Model 6) suggest that the spike in ETF premium on dividend record day remains after controlling for the demand for ETF dividends.

The findings thus far reveal that both ETF borrowing costs and ETF premiums increase on dividend record days. However, the missing part is that I have not demonstrated that it is the increased borrowing costs that lead to heightened ETF premiums. To address this concern, I employ a Two-Stage Least Squares (2SLS) approach. Specifically, I first project DCBS onto the dividend record day dummy and obtain fitted values. Then, I re-run regression (4) by replacing the Dividend Record Day dummy with the projected DCBS.[24] The results, reported in Appendix Table A3, are consistent with those in Table 8: DCBS projected by the dividend record day is positively related to ETF premiums. Therefore, I conclude that the impact of ETF short-sale costs on ETF premiums is indeed causal.

3.4 Interaction between the Primary and Secondary Market

In this section, I attempt to shed more light on how secondary arbitrageurs trade by examining whether and how arbitrageurs in the primary and secondary markets interact with each other. The ETF price-NAV difference represents a textbook arbitrage opportunity for the primary market arbitrageurs (APs). Having observed an ETF premium, APs can immediately lock in the profits by selling ETF shares short and purchasing the underlying stocks. At the end of a trading day, APs use the underlying

[24] For consistency, this 2SLS regression is also run for the 30-day window around ETF dividend record days.

stocks to create ETF shares and unwind short positions in ETF shares. In this process, APs do not bear risks since the exchange of ETF shares always happens at NAVs and they do not rely on price convergence to make arbitrage profits. Secondary market arbitrageurs, on the other hand, are essentially betting on the convergence of ETF prices and underlying NAVs, and hence bear the risks that the price discrepancy does not converge or converge very slowly. Therefore, whether secondary market arbitrageurs correct ETF mispricing depends on how much risk they bear. I argue that this risk depends on the activeness of APs. When APs are inactive, mispricing persists and hence secondary market arbitrageurs are reluctant to trade against ETF mispricing because the risk of slow price convergence is high. One extreme example is the closed-end puzzle discussed in Lee et al. (1991) and Pontiff (1996) among others. As APs become more active, price convergence becomes faster, and secondary market arbitrageurs are more willing to trade against ETF mispricing. When APs are highly active, they become the marginal arbitrageur in the ETF market, and the arbitrage opportunity left for secondary market arbitrageurs is small.

To test this idea, I employ two proxies for ETF primary market activeness. Following Gorbatikov and Sikorskaya (2022), the first proxy is the number of active APs. Additionally, the second proxy is the standard deviation of percentage changes in the number of ETF shares outstanding. I then divide ETFs into three groups based on the activeness of the primary market. First, I confirm that the speed for price convergence is indeed faster when the primary market is more active. Following Madhavan (2014), I use ψ as an inverse measure of the speed of mispricing correction:

$$u_{i,t} = \psi_i u_{i,t-1} + \epsilon_{i,t}$$

where $u_{i,t}$ is the mispricing of ETF i at time t. In unreported test, I show that the average

ψ_i monotonically decreases with the primary market activeness.[25]

I then re-run regression (1) for each group, and the results are reported in Table 9. Panel A (B) represents results without (with) control variables. All reported results include ETF and day fixed effects, along with double-clustered standard errors at the ETF and day levels. The left (right) three columns represent the results with the number of active APs (standard deviation of changes in the number of ETF shares outstanding) as a measurement of the primary market activeness. When the primary market activeness is low, the sensitivity of ETF mispricing to DCBS is low (if significant). This is consistent with the conjecture that when the primary market is inactive, secondary market arbitrageurs bear more risks of persistent price divergence and are reluctant to correct mispricing. Therefore, their impact on ETF mispricing is small. In contrast, as the activeness of the primary market reaches a medium level, the coefficients of DCBS increase, approximately doubling in magnitude compared to the low primary market activeness scenario. This result indicates that secondary market arbitrageurs are more willing to trade against ETF relative overpricing and bring more arbitrage capital. Therefore, given the same level of changes in arbitrage costs, secondary market arbitrageurs exert a stronger influence on ETF mispricing. However, when the primary market activeness is high, the coefficient of DCBS decreases. This is again consistent with my hypothesis that when APs are highly active, the arbitrage opportunity left for the secondary market is small, and the marginal impact of secondary market arbitrageurs becomes smaller. The results are more pronounced if we include control variables (shown in Panel B of Table 9). In terms of economic magnitude, a one-standard deviation increase in DCBS corresponds to a 0.69-, 1.38-, and 0.39-bps increase in ETF mispricing, when the primary market activeness is low, medium, and high, respectively.[26]

[25] The results are available upon requests.

[26] In Appendix Table A4, for robustness, I divided ETFs into five groups based on the primary market activeness, and the results are slightly stronger.

4 Conclusion

In this paper, I examine whether the friction in the ETF secondary market hinders ETF pricing efficiency, which is largely overlooked in the literature. In particular, I show that the cost of borrowing ETFs is positively related to ETF mispricing. The cost of borrowing ETFs is likely to be a friction that mainly affects secondary market arbitrageurs because many APs are also ETF market makers who do not need to borrow ETFs before selling them short. Empirical results lend support to this conjecture. The effect of ETF borrowing costs on mispricing is much more pronounced on days when APs do trade, highlighting the importance of secondary market participants in correcting ETF mispricing. Leveraging on two exogenous variations in ETF borrowing fees, I establish a causal relationship between ETF borrowing costs and mispricing. Lastly, my empirical results provide novel insights into the dynamics of the ETF arbitrage mechanism by uncovering an interdependence between the primary and secondary markets. Future theoretical works may consider the interplay between APs and secondary market arbitrageurs when modeling the ETF arbitrage mechanism. Overall, this paper emphasizes the importance of the secondary market in ETF pricing efficiency.

References

Antoniewicz, R. S. and Heinrichs, J. (2014). Understanding exchange-traded funds: How ETFs work. *Jane, Understanding Exchange-Traded Funds: How ETFs Work (September 30, 2014)*.

Bae, K. and Kim, D. (2020). Liquidity risk and exchange-traded fund returns, variances, and tracking errors. *Journal of Financial Economics*, 138(1):222–253.

Bhattacharya, A. and O'Hara, M. (2018). Can etfs increase market fragility? effect of information linkages in ETF markets. *Effect of Information Linkages in ETF Markets (April 17, 2018)*.

Blocher, J., Reed, A. V., and Van Wesep, E. D. (2013). Connecting two markets: An equilibrium framework for shorts, longs, and stock loans. *Journal of Financial Economics*, 108(2):302–322.

Blocher, J. and Whaley, R. E. (2015). Passive investing: The role of securities lending. *Vanderbilt Owen Graduate School of Management Research Paper*, 2474904.

Chang, E. C., Cheng, J. W., and Yu, Y. (2007). Short-sales constraints and price discovery: Evidence from the hong kong market. *The Journal of Finance*, 62(5):2097–2121.

Chen, J., Hong, H., and Stein, J. C. (2002). Breadth of ownership and stock returns. *Journal of financial Economics*, 66(2-3):171–205.

Choi, D., Getmansky, M., and Tookes, H. (2009). Convertible bond arbitrage, liquidity externalities, and stock prices. *Journal of Financial Economics*, 91(2):227–251.

Christoffersen, S. E., Geczy, C. C., Musto, D. K., and Reed, A. V. (2005). Crossborder dividend taxation and the preferences of taxable and nontaxable investors: Evidence from canada. *Journal of Financial Economics*, 78(1):121–144.

Comerton-Forde, C., Jones, C. M., and Putniņš, T. J. (2016). Shorting at close range: A tale of two types. *Journal of Financial Economics*, 121(3):546–568.

Comerton-Forde, C. and Marta, T. (2021). ETF effects: The role of primary versus secondary market activities. *Available at SSRN 3980678*.

Diamond, D. W. and Verrecchia, R. E. (1987). Constraints on short-selling and asset price adjustment to private information. *Journal of Financial Economics*, 18(2):277–311.

Dixon, P. N., Fox, C. A., and Kelley, E. K. (2021). To own or not to own: Stock loans around dividend payments. *Journal of Financial Economics*, 140(2):539–559.

Evans, R. B., Moussawi, R., Pagano, M. S., and Sedunov, J. (2022). Operational shorting and ETF liquidity provision. *Available at SSRN*.

Fama, E. F. and MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of political economy*, 81(3):607–636.

Gorbatikov, E. and Sikorskaya, T. (2022). Two aps are better than one: ETF mispricing and primary market participation. *Available at SSRN 3923503*.

Gromb, D. and Vayanos, D. (2010). Limits of arbitrage. *Annu. Rev. Financ. Econ.*, 2(1):251–275.

Huang, S., OHara, M., and Zhong, Z. (2021). Innovation and informed trading: Evidence from industry etfs. *The Review of Financial Studies*, 34(3):1280–1316.

Jain, A., Jain, C., and Jiang, C. X. (2021). Active trading in ETFs: The role of high-frequency algorithmic trading. *Financial Analysts Journal*, 77(2):66–82.

Jarrow, R. (1980). Heterogeneous expectations, restrictions on short sales, and equilibrium asset prices. *The Journal of Finance*, 35(5):1105–1113.

Karmaziene, E. and Sokolovski, V. (2022). Short-selling equity exchange traded funds and its effect on stock market liquidity. *Journal of Financial and Quantitative Analysis*, 57(3):923–956.

Kashner, E. (2017). The heartbeat of ETF tax efficiency. Retrived from FactSet Insight: https://insight.factset.com/the-heartbeat-of-etf-tax-efficiency.

Lee, C. M., Shleifer, A., and Thaler, R. H. (1991). Investor sentiment and the closed-end fund puzzle. *The Journal of Finance*, 46(1):75–109.

Lettau, M. and Madhavan, A. (2018). Exchange-traded funds 101 for economists. *Journal of Economic Perspectives*, 32(1):135–154.

Li, F. W. and Zhu, Q. (2022). Short selling ETFs. *Review of Asset Pricing Studies*, 12(4):960–998.

Madhavan, A. (2014). Exchange-traded funds: an overview of institutions, trading, and impacts. *Annu. Rev. Financ. Econ.*, 6(1):311–341.

Malamud, S. (2016). A dynamic equilibrium model of ETFs. *CEPR Discussion Paper No. DP11469*.

McDonald, R. L. (2001). Cross-border investing with tax arbitrage: The case of german dividend tax credits. *The Review of Financial Studies*, 14(3):617–657.

Miller, E. M. (1977). Risk, uncertainty, and divergence of opinion. *The Journal of finance*, 32(4):1151–1168.

Moussawi, R., Shen, K., and Velthuis, R. (2022). ETF heartbeat trades, tax efficiencies, and clienteles: The role of taxes in the flow migration from active mutual funds to etfs. *Available at SSRN 3744519*.

Pan, K. and Zeng, Y. (2019). ETF arbitrage under liquidity mismatch. *Available at SSRN 3723406*.

Pontiff, J. (1996). Costly arbitrage: Evidence from closed-end funds. *The Quarterly Journal of Economics*, 111(4):1135–1151.

Raddatz, C. E. (2021). Does the diversity and solvency of authorized participants matter for bond etf arbitrage? evidence from the dash for cash episode. *Evidence from the Dash for Cash Episode (June 4, 2021)*.

Shim, J. J. and Todorov, K. (2022). ETFs, illiquid assets, and fire sales. *Available at SSRN 3886881*.

Thornock, J. (2013). The effects of dividend taxation on short selling and market quality. *The Accounting Review*, 88(5):1833–1856.

Xiao, H. (2022). The economics of etf redemptions: Liquidity and fire sale risks. *Available at SSRN 4096222*.

Zurowska, E. A. (2022). *ETF primary market structure and its efficiency*. PhD thesis, University of Pennsylvania.

Figures and Tables

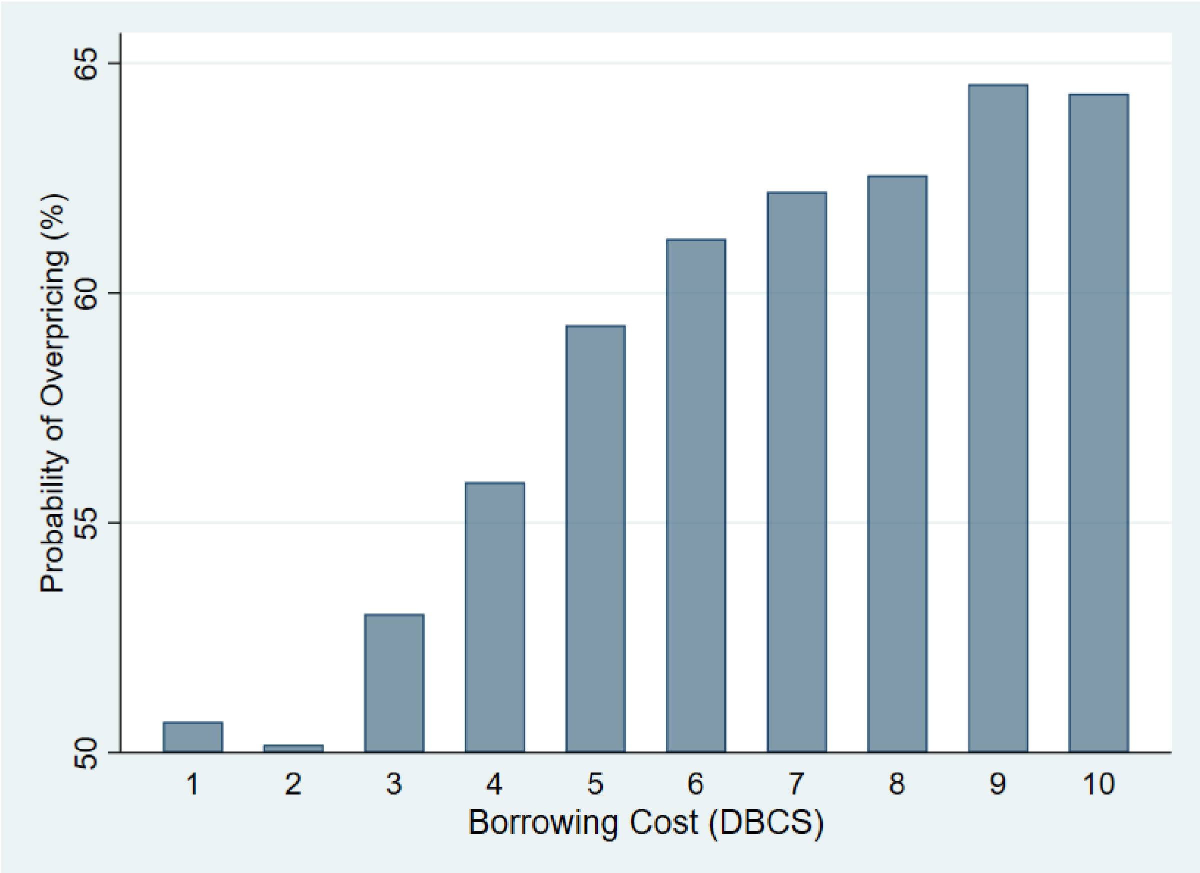

Figure 1 Probability of ETF Premium by Borrowing Cost

This figure plots the relationship between ETF borrowing cost and the proportion of overpriced ETFs. ETF borrowing costs are proxied by Markit Security Finance Daily Cost of Borrow Score (DCBS) with a score of 1 (10) representing the cheapest (most expensive) borrowing cost. For each DCBS group (from 1 to 10), the proportion of ETF overpricing is computed as the number of ETFs traded at a price greater than the underlying NAV, divided by the total number of ETFs within that group. The sample period is from July 2001 to September 2022.

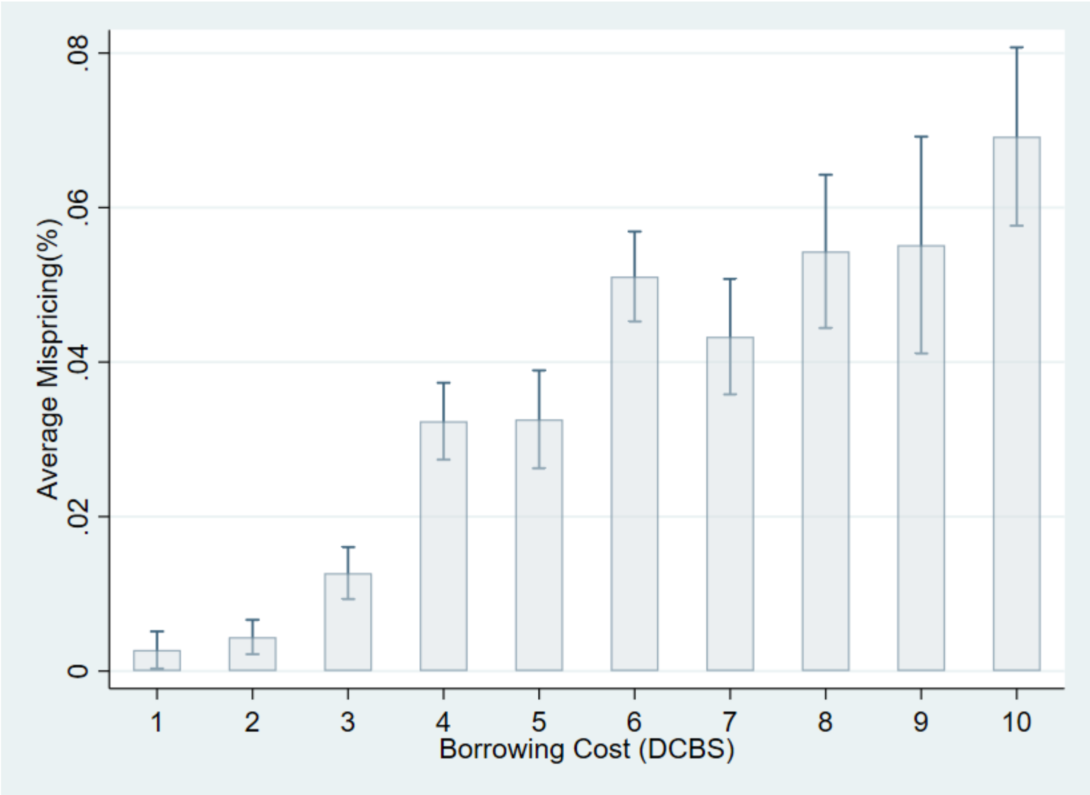

FIGURE 2 ETF MISPRICING BY BORROWING COST

This figure plots the relationship between ETF borrowing cost and ETF mispricing, which is defined as the percentage difference between the ETF market price and the underlying NAV. A positive (negative) suggests that an ETF is traded at a premium (discount). ETF borrowing costs are proxied by Markit Security Finance Daily Cost of Borrow Score (DCBS) with a score of 1 (10) representing the cheapest (most expensive) borrowing cost. For each DCBS group (from 1 to 10), mispricing is computed as the equally-weighted average mispricing within that group. The sample period is from July 2001 to September 2022.

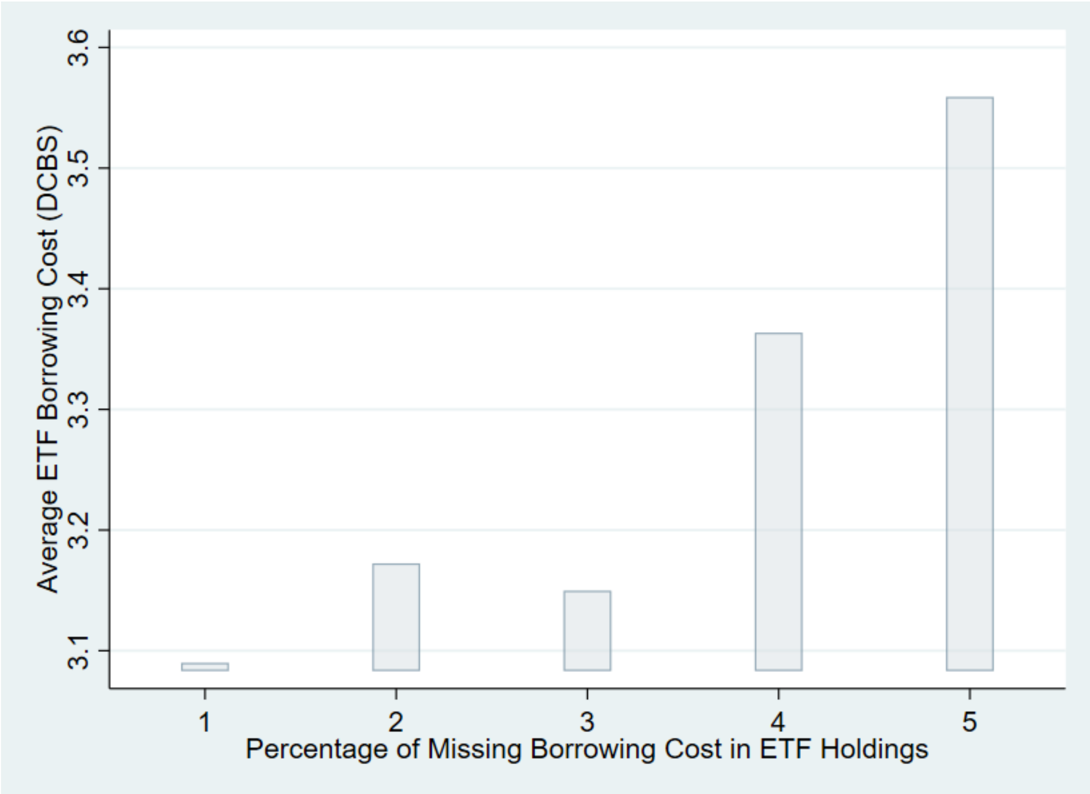

Figure 3 Underlying Short-Sale Constraints and ETF Lending Cost

This figure plots the relationship between ETF borrowing cost and the short-sale constraints of underlying stocks. Both ETF and underlying stock borrowing costs are proxied by Markit Security Finance Daily Cost of Borrow Score (DCBS) with a score of 1 (10) representing the cheapest (most expensive) borrowing cost. Short-sale constraints of underlying stocks are measured by the percentage of ETF holdings with missing borrowing cost data. The full sample is divided into five groups based on the underlying short-sale constraints. Group 1 (5) represents the group of ETFs with the lowest (highest) level of underlying short-sale constraints. Each bar then represents the average ETF borrowing cost for each underlying short-sale constraint group. The sample period is from July 2001 to September 2022.

Figure 4 Underlying Short-Sale Constraints and ETF Mispricing

This figure plots the relationship between ETF mispricing and the short-sale constraints of underlying stocks. ETF mispricing is defined as the percentage difference between the ETF market price and the underlying NAV. A positive (negative) suggests that an ETF is traded at a premium (discount). Both ETF and underlying stock borrowing costs are proxied by Markit Security Finance Daily Cost of Borrow Score (DCBS) with a score of 1 (10) representing the cheapest (most expensive) borrowing cost. Short-sale constraints of underlying stocks are measured by the percentage of ETF holdings with missing borrowing cost data. The full sample is divided into five groups based on the underlying short-sale constraints. Group 1 (5) represents the group of ETFs with the lowest (highest) level of underlying short-sale constraints. Each bar then represents the average ETF mispricing for each underlying short-sale constraint group. The sample period is from July 2001 to September 2022.

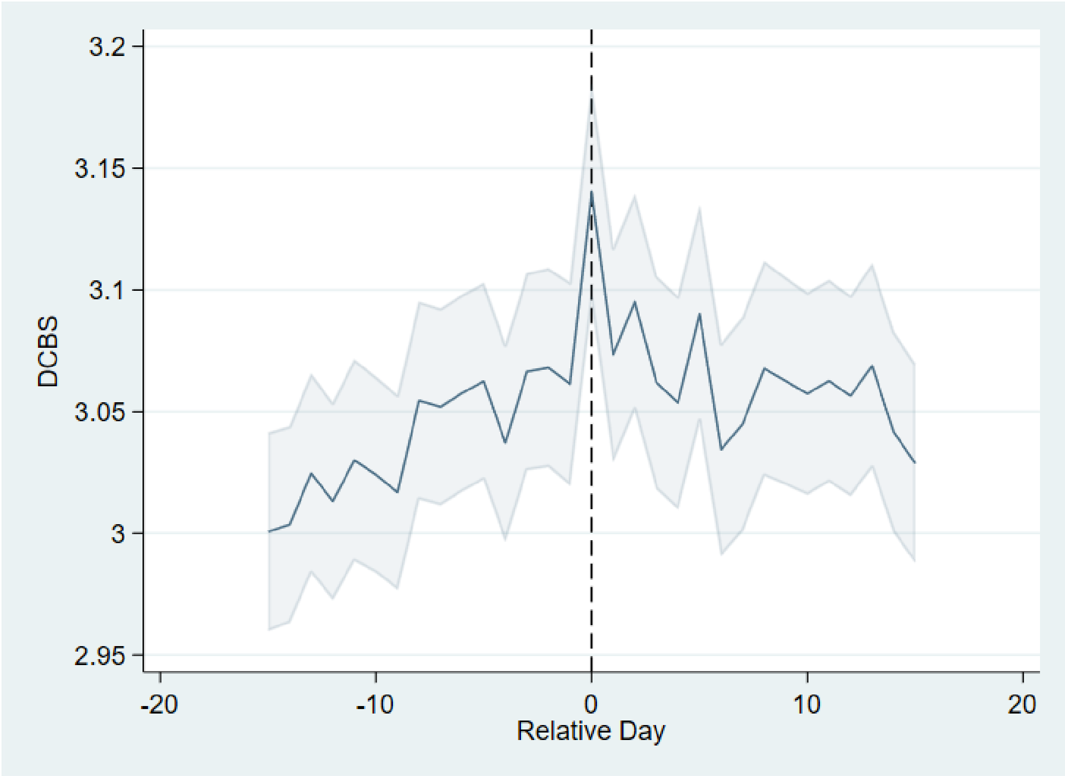

Figure 5 Borrowing Cost around Dividend Record Day

This figure plots average ETF borrowing costs and the corresponding confidence interval within a 30-day window ($t-15$, $t+15$) around ETF dividend record date. ETF borrowing costs are proxied by Markit Security Finance Daily Cost of Borrow Score (DCBS) with a score of 1 (10) representing the cheapest (most expensive) borrowing cost. The sample period is from July 2001 to September 2022.

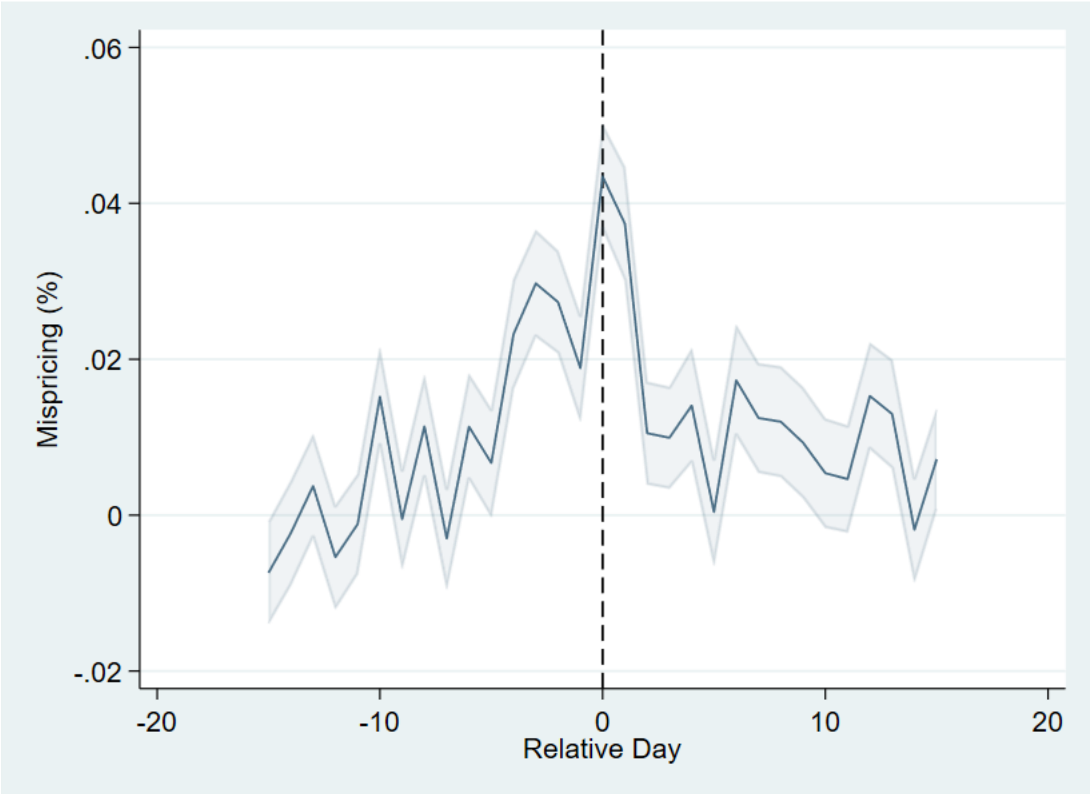

Figure 6 ETF Mispricing around Dividend Record Day

This figure plots average ETF mispricing and the corresponding confidence interval within a 30-day window $(t-15, t+15)$ around ETF dividend record date. ETF mispricing is defined as the percentage difference between the ETF market price and the underlying NAV. A positive (negative) suggests that an ETF is traded at a premium (discount). The sample period is from July 2001 to September 2022.

Table 1 Summary Statistics

This table reports summary statistics of variables used in this study. **Mispricing** is computed as the percentage difference between the ETF market price and the underlying NAV. A positive (negative) suggests that an ETF is traded at a premium (discount). **DCBS** is the Markit Security Finance Daily Cost of Borrow Score, which is a number from 1 to 10 indicating the rebate/fee charged by the agent lender, where 1 (10) indicates the cheapest (most expensive) borrowing cost. **vwafs1d** is the value-weighted average fee for all new trades on the most recent day. **vwafs7d** is the value-weighted average fee for all new trades over the most recent 7 calendar days. Both **vwafs1d** and **vwafs7d** are a number from 0 to 5 with 0(5) indicating the cheapest (most expensive) borrowing cost. **Expense Ratio** represents how much investors pay for the fund's operating expenses. **Bid-Ask Spread** is measured by the difference between ETF daily ask and bid prices, scaled by the mid-quote. **Turnover** is the number of ETF shares traded divided by the total number of shares outstanding. **Size** is the ETF market capitalization, calculated as ETF price multiplied by the total number of shares outstanding. **Volatility** is the standard deviation of daily ETF returns over the past 22 trading days. **Number of APs** is the number of total registered APs for an ETF. **Number of Active APs** is the number of APs that have a non-zero dollar value of creation/redemption. **Creation Unit** is the minimum number of ETF shares that an AP has to exchange with the ETF sponsor for each creation transaction. **Creation Fee** is the simple average of "in-kind" creation and cash creation fees. The sample period is from July 2001 to September 2022.

	N	Mean	Std. Dev.	p5	p25	Median	p75	p95
Mispricing (bps)	2,467,256	1.100	36.913	-41.317	-6.515	0.807	8.186	44.886
DCBS	1,434,663	3.359	1.777	1.000	2.000	3.000	4.000	7.000
vwafs1d	957,164	2.864	1.716	0.000	1.000	3.000	4.000	5.000
vwafs7d	1,411,216	2.984	1.675	0.000	2.000	3.000	4.000	5.000
Expense Ratio	2,163,103	0.434	0.256	0.090	0.250	0.400	0.600	0.850
Bid-Ask Spread (bps)	2,466,618	25.189	106.405	1.708	5.632	11.744	24.051	69.510
Turnover	1,617,456	1.055	4.791	0.001	0.003	0.006	0.018	5.366
Size (Million)	1,617,337	3,224.010	15,280.733	1.394	51.129	264.352	1,245.231	13,669.742
Volatility(%)	2,450,544	1.311	1.028	0.470	0.741	1.052	1.535	3.001
Number of APs	835,098	26.766	15.685	3.000	13.000	26.000	41.000	50.000
Number of Active APs	826,144	6.314	4.297	1.000	3.000	5.000	8.000	16.000
Creation Unit	835,098	42,053.067	72,215.601	10,000	25,000	50,000	50,000	50,000
Creation Fee - In Kind (%)	830,997	0.045	0.070	0.000	0.011	0.028	0.052	0.144
Creation Fee - Cash (%)	828,983	0.042	0.069	0.000	0.008	0.026	0.050	0.137

Table 2 Summary Statistics by Borrowing Cost

This table reports summary statistics by ETF borrowing cost, which is proxied by DCBS **Mispricing** is computed as the percentage difference between the ETF market price and the underlying NAV. A positive (negative) suggests that an ETF is traded at a premium (discount). **DCBS** is the Markit Security Finance Daily Cost of Borrow Score, which is a number from 1 to 10 indicating the rebate/fee charged by the agent lender, where 1 (10) indicates the cheapest (most expensive) borrowing cost. **Expense Ratio** represents how much investors pay for the fund's operating expenses. **Bid-Ask Spread** is measured by the difference between ETF daily ask and bid prices, scaled by the mid-quote. **Turnover** is the number of ETF shares traded divided by the total number of shares outstanding. **Size** is the ETF market capitalization, calculated as ETF price multiplied by the total number of shares outstanding. **Volatility** is the standard deviation of daily ETF returns over the past 22 trading days. **Number of APs** is the number of total registered APs for an ETF. **Number of Active APs** is the number of APs that have a non-zero dollar value of creation/redemption. **Creation Unit** is the minimum number of ETF shares that an AP has to exchange with the ETF sponsor for each creation transaction. **Creation Fee** is the simple average of "in-kind" creation and cash creation fees. The sample period is from July 2001 to September 2022.

DCBS	Mispricing(bps)	Expense Ratio	Spread(bps)	Turnover	Size (Million)	Volatility	No. APs	No. Active APs	Creation Unit
1	0.273	0.255	7.554	17.506	16,424.873	1.305	36.206	13.234	47,338.410
2	0.441	0.384	14.501	11.242	3,142.623	1.398	29.953	8.806	44,059.932
3	1.268	0.445	17.408	7.443	1,132.743	1.441	27.596	6.849	49,315.326
4	3.234	0.472	18.662	18.797	738.532	1.453	27.374	6.317	43,183.997
5	3.258	0.452	19.424	25.270	640.776	1.320	26.152	5.910	42,174.305
6	5.107	0.411	19.771	14.633	645.299	1.222	24.829	5.685	41,100.127
7	4.329	0.367	18.123	2.343	742.293	1.215	26.845	5.449	42,484.246
8	5.433	0.364	18.948	0.900	973.500	1.135	26.977	5.565	47,992.056
9	5.515	0.438	18.146	0.013	563.094	1.122	20.298	4.833	41,262.092
10	6.919	0.404	16.025	0.012	750.660	1.054	23.419	4.830	41,824.252
10 - 1	6.100	0.156	9.487	-0.392	-26,701.882	-0.141	-12.816	-7.932	-4,987.498
p-value	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)

Table 3 ETF Borrowing Cost and Mispricing

This table reports regression estimates of ETF mispricing on ETF borrowing cost, which is proxied by DCBS **Mispricing** is computed as the percentage difference between ETF market prices and the underlying NAVs. A positive (negative) suggests that an ETF is traded at a premium (discount). **DCBS** is the Markit Security Finance Daily Cost of Borrow Score, which is a number from 1 to 10 indicating the rebate/fee charged by the agent lender, where 1 (10) indicates the cheapest (most expensive) borrowing cost. **Expense Ratio** represents how much investors pay for the fund's operating expenses. **Bid-Ask Spread** is measured by the difference between ETF daily ask and bid prices, scaled by the mid-quote. **Ln(Turnover)** is the natural log of the number of ETF shares traded divided by the total number of shares outstanding. **Ln(Size)** is the natural log of ETF market capitalization, calculated as ETF price multiplied by the total number of shares outstanding. **Volatility** is the standard deviation of daily ETF returns over the past 22 trading days. Numbers in parentheses are *t*-statistics based on double-clustered standard errors at the ETF and the day level. The sample period is from July 2001 to September 2022.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
DCBS	0.813***	0.617***	0.650***	0.512***	0.771***	0.553***	0.660***	0.490***
	(9.666)	(6.490)	(8.241)	(5.593)	(9.120)	(5.140)	(8.408)	(5.504)
Expense Ratio		-1.783*		-4.984*		-2.477**		-8.456**
		(-1.737)		(-1.787)		(-2.234)		(-2.491)
Bid-Ask Spread		0.042**		0.013		0.042**		0.013
		(2.164)		(1.106)		(2.105)		(1.131)
Ln(Turnover)		0.208*		0.453***		0.344***		0.676***
		(1.946)		(2.677)		(2.821)		(4.013)
Ln(Size)		-0.318**		-0.187		-0.416***		-0.127
		(-2.222)		(-0.851)		(-2.619)		(-0.488)
Volatility		-0.769***		-0.809***		-0.438		-0.044
		(-3.987)		(-3.770)		(-1.220)		(-0.084)
ETF FE	✗	✗	✓	✓	✗	✗	✓	✓
Day FE	✗	✗	✗	✗	✓	✓	✓	✓
No. of Obs	1,434,564	976,020	1,434,554	976,018	1,434,542	975,999	1,434,532	975,997
$Adj.\ R^2$	0.002	0.005	0.067	0.045	0.030	0.030	0.094	0.070

Table 4 ETF Borrowing Cost and Mispricing (Additional Controls)

This table reports regression estimates of ETF mispricing on ETF borrowing cost with additional control variables. ETF borrowing cost is proxied by DCBS. **Mispricing** is computed as the percentage difference between the ETF market price and the underlying NAV. A positive (negative) suggests that an ETF is traded at a premium (discount). **DCBS** is the Markit Security Finance Daily Cost of Borrow Score, which is a number from 1 to 10 indicating the rebate/fee charged by the agent lender, where 1 (10) indicates the cheapest (most expensive) borrowing cost. **Expense Ratio** represents how much investors pay for the fund's operating expenses. **Bid-Ask Spread** is measured by the difference between ETF daily ask and bid prices, scaled by the mid-quote. **Ln(Turnover)** is the natural log of the number of ETF shares traded divided by the total number of shares outstanding. **Ln(Size)** is the natural log of ETF market capitalization, calculated as ETF price multiplied by the total number of shares outstanding. **Volatility** is the standard deviation of daily ETF returns over the past 22 trading days. **Mean Holding Spreads** is the average bid-ask spread of stocks in ETF holdings. **Number of APs** is the number of total registered APs for an ETF. **Number of Active APs** is the number of APs that have a non-zero dollar value of creation/redemption. **Creation Unit** is the minimum number of ETF shares that an AP has to exchange with the ETF sponsor for each creation transaction. **Creation Fee** is the simple average of "in-kind" creation and cash creation fees. Numbers in parentheses are *t*-statistics based on double-clustered standard errors at the ETF and the day level. The sample period is from 2018 to 2022.

	Model 1	Model 2	Model 3	Model 4
DCBS	0.402***	0.323***	0.397***	0.307***
	(4.183)	(3.543)	(3.984)	(3.372)
Expense Ratio	-1.552	-16.777**	-2.334**	-11.709*
	(-1.497)	(-2.441)	(-2.311)	(-1.858)
Bid-Ask Spread	0.023	0.001	0.028	0.004
	(1.238)	(0.053)	(1.488)	(0.331)
Ln(Turnover)	0.942***	0.874***	1.042***	1.259***
	(6.007)	(5.485)	(6.581)	(7.386)
Ln(Size)	0.033	0.350	-0.001	0.133
	(0.268)	(1.405)	(-0.011)	(0.485)
Volatility	-1.203***	-0.866***	-0.634	0.103
	(-6.691)	(-4.055)	(-1.636)	(0.202)
Mean Holding Spread	-0.046	-0.202**	-0.050*	-0.185*
	(-1.627)	(-2.236)	(-1.697)	(-1.934)
Number of APs	-0.067***	-0.046**	-0.067***	0.000
	(-5.442)	(-2.538)	(-4.965)	(0.007)
Number of Active APs	-0.115**	-0.208***	-0.131***	-0.225***
	(-2.528)	(-3.479)	(-2.632)	(-3.375)
Creation Unit	0.000	-0.000**	0.000*	-0.000
	(0.681)	(-2.470)	(1.699)	(-1.159)
Creation Fee	0.464***	0.208	0.433***	0.168
	(3.306)	(0.976)	(3.079)	(0.856)
ETF FE	✗	✓	✗	✓
Day FE	✗	✗	✓	✓
No. of Obs	327,205	327,202	327,205	327,202
$Adj.\ R^2$	0.021	0.070	0.053	0.101

TABLE 5 AP CREATION *V.S.* NO AP CREATION

This table reports regression estimates of ETF mispricing on ETF borrowing cost, which is proxied by DCBS. The left two columns report results for days when there is an increase in the number of ETF shares outstanding (APs create ETF shares) whereas the right two columns report results for days when there is no change in the number of ETF shares outstanding. **Mispricing** is computed as the percentage difference between the ETF market price and the underlying NAV. A positive (negative) suggests that an ETF is traded at a premium (discount). **DCBS** is the Markit Security Finance Daily Cost of Borrow Score, which is a number from 1 to 10 indicating the rebate/fee charged by the agent lender, where 1 (10) indicates the cheapest (most expensive) borrowing cost. **Expense Ratio** represents how much investors pay for the fund's operating expenses. **Bid-Ask Spread** is measured by the difference between ETF daily ask and bid prices, scaled by the mid-quote. **Ln(Turnover)** is the natural log of the number of ETF shares traded divided by the total number of shares outstanding. **Ln(Size)** is the natural log of ETF market capitalization, calculated as ETF price multiplied by the total number of shares outstanding. **Volatility** is the standard deviation of daily ETF returns over the past 22 trading days. Numbers in parentheses are *t*-statistics based on double-clustered standard errors at the ETF and the day level. The sample period is from July 2001 to September 2022.

	Creation		No Creation	
	Mispricing	Mispricing	Mispricing	Mispricing
DCBS	0.253***	0.166*	0.701***	0.508***
	(2.902)	(1.850)	(8.074)	(5.059)
Expense Ratio		-11.091		-8.000**
		(-1.641)		(-2.255)
Bid-Ask Spread		0.039		0.012
		(1.487)		(0.908)
Ln(Turnover)		0.095		0.820***
		(0.629)		(4.386)
Ln(Size)		-0.308		-0.471
		(-1.539)		(-1.312)
Volatility		-0.462		0.059
		(-0.673)		(0.126)
ETF FE	✓	✓	✓	✓
Day FE	✓	✓	✓	✓
No. of Obs	222,867	200,585	1,073,672	651,897
$Adj.\ R^2$	0.150	0.159	0.105	0.078

TABLE 6 UNDERLYING SHORT-SALE CONSTRAINTS AND ETF LENDING COST

This table reports regression estimates of ETF DCBS on underlying short-sale constraints. The left (right) two columns represent OLS (Fama-Macbeth regression) results. **Missing(%)** is the percentage of ETF underlying stocks with missing DCBS data. **DCBS** is the Markit Security Finance Daily Cost of Borrow Score, which is a number from 1 to 10 indicating the rebate/fee charged by the agent lender, where 1 (10) indicates the cheapest (most expensive) borrowing cost. **Utilization** is the number of shares on loan, divided by the number of shares available for lending. **Volatility** is the standard deviation of daily ETF returns over the past 22 trading days. **Bid-Ask Spread** is measured by the difference between ETF daily ask and bid prices, scaled by the mid-quote. **Ln(Size)** is the natural log of ETF market capitalization, calculated as ETF price multiplied by the total number of shares outstanding. **Ln(Turnover)** is the natural log of the number of ETF shares traded divided by the total number of shares outstanding. Numbers in parentheses are *t*-statistics based on Newey-West standard errors. The sample period is from July 2001 to September 2022.

	OLS		FMB	
	(1) DCBS	(2) DCBS	(3) DCBS	(4) DCBS
Missing(%)	0.010***	0.007***	0.007***	0.014***
	(3.314)	(3.646)	(12.064)	(3.119)
Utilization		0.006***		0.003***
		(10.825)		(10.486)
Volatility		-0.074***		-0.103**
		(-3.832)		(-2.384)
Bid-Ask Spread		-14.091***		21.570
		(-2.775)		(1.616)
Ln(Size)		-0.286***		-0.285***
		(-17.949)		(-43.259)
Ln(Turnover)		-0.302***		-0.168***
		(-19.113)		(-36.563)
No. of Obs	1,147,943	893,664	1,147,943	893,664
Adj. R^2	0.009	0.219	0.015	0.308

TABLE 7 UNDERLYING SHORT-SALE CONSTRAINTS AND ETF MISPRICING

This table reports regression estimates of ETF mispricing on projected DCBS by underlying short-sale constraints. The left (right) two columns represent OLS (Fama-Macbeth regression) results. $\widehat{DCBS}$ is the borrowing costs projected by the underlying short-sale constraint, which is proxied by the percentage of ETF underlying stocks with missing DCBS data. **Utilization** is the number of shares on loan, divided by the number of shares available for lending. **Volatility** is the standard deviation of daily ETF returns over the past 22 trading days. **Bid-Ask Spread** is measured by the difference between ETF daily ask and bid prices, scaled by the mid-quote. **Ln(Size)** is the natural log of ETF market capitalization, calculated as ETF price multiplied by the total number of shares outstanding. **Ln(Turnover)** is the natural log of the number of ETF shares traded divided by the total number of shares outstanding. Numbers in parentheses are *t*-statistics based on Newey-West standard errors. The sample period is from July 2001 to September 2022.

	OLS		FMB	
	(1) Mispricing	(2) Mispricing	(3) Mispricing	(4) Mispricing
$\widehat{DCBS}$	0.017***	0.011***	0.219*	0.564**
	(8.042)	(5.068)	(1.740)	(2.375)
Utilization		0.000***		0.000***
		(6.417)		(2.607)
Volatility		-0.010***		-0.014
		(-5.961)		(-1.460)
Bid-Ask Spread		4.698***		26.837***
		(3.135)		(4.759)
Ln(Size)		-0.003***		0.002**
		(-4.561)		(2.384)
Ln(Turnover)		0.002**		0.005***
		(2.330)		(8.823)
No. of Obs	1,147,866	893,617	1,147,866	893,617
$Adj.\ R^2$	0.003	0.012	0.058	0.230

TABLE 8 ETF MISPRICING ON DIVIDEND RECORD DAY

This table reports regression estimates of ETF mispricing on dividend record day. The regression is run within a 30-day $(t-15, t+15)$ window around the ETF dividend record day. **Mispricing** is computed as the percentage difference between the ETF market price and the underlying NAV. A positive (negative) suggests that an ETF is traded at a premium (discount). **Record Day** is a dummy variable which is equal to 1 if day t is the dividend record day for ETF i. **Expense Ratio** represents how much investors pay for the fund's operating expenses. **Bid-Ask Spread** is measured by the difference between ETF daily ask and bid prices, scaled by the mid-quote. **Ln(Turnover)** is the natural log of the number of ETF shares traded divided by the total number of shares outstanding. **Ln(Size)** is the natural log of ETF market capitalization, calculated as ETF price multiplied by the total number of shares outstanding. **Volatility** is the standard deviation of daily ETF returns over the past 22 trading days. **Mean Holding Spreads** is the average bid-ask spread of stocks in ETF holdings. $\mathbf{Ret_t}$ is the return on the cum-dividend day (the day before the ex-dividend day). $\mathbf{Ret_{t-2,t}}$ is the three-day buy-and-hold return up to the cum-dividend day. $\mathbf{Ret_{t-4,t}}$ is the five-day buy-and-hold return up to the cum-dividend day. Numbers in parentheses are t-statistics based on standard errors clustered at the ETF-Dividend Event level. The sample period is from 2018 to 2022.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Record Day	0.0337***	0.0278***	0.0278***	0.0338***	0.0286***	0.0274***
	(11.693)	(10.309)	(10.308)	(11.730)	(10.591)	(10.105)
Expense Ratio		-0.0227*	-0.0222*		0.0680	0.3440***
		(-1.936)	(-1.906)		(0.981)	(5.377)
Bid-Ask Spread		0.0002***	0.0002***		0.0001***	0.0002***
		(3.437)	(3.458)		(3.090)	(3.537)
Ln(Turnover)		-0.0073***	-0.0072***		0.0005	-0.0093***
		(-5.691)	(-5.625)		(0.526)	(-4.128)
Ln(Size)		0.0002	0.0003		-0.0154***	0.0016
		(0.222)	(0.270)		(-2.606)	(0.772)
Volatility		-0.0108***	-0.0108***		0.0079	-0.0119***
		(-3.776)	(-3.848)		(1.521)	(-3.863)
Ret_t			0.2257			0.2674
			(1.292)			(1.638)
$Ret_{t-2,t}$			0.0019			0.0016
			(1.293)			(1.153)
$Ret_{t-4,t}$			-0.0021*			-0.0030***
			(-1.914)			(-3.116)
ETF FE	✗	✗	✗	✗	✗	✓
ETF-Dividend FE	✗	✗	✗	✓	✓	✗
No. of Obs	327,358	217,271	217,271	327,358	217,271	217,271
$Adj.\ R^2$	0.000	0.011	0.011	0.357	0.425	0.124

Table 9 ETF Borrowing Cost and Mispring by Primary Market Activeness

This table reports regression estimates of ETF mispricing on ETF borrowing cost, conditional on different levels of ETF primary market activeness, which are proxied by two variables: the number of active APs (results shown in the left three columns) and the standard deviation of changes in daily number of ETF shares outstanding (results shown in the right three columns). **Mispricing** is computed as the percentage difference between the ETF market price and the underlying NAV. A positive (negative) suggests that an ETF is traded at a premium (discount). **DCBS** is the Markit Security Finance Daily Cost of Borrow Score, which is a number from 1 to 10 indicating the rebate/fee charged by the agent lender, where 1 (10) indicates the cheapest (most expensive) borrowing cost. **Expense Ratio** represents how much investors pay for the fund's operating expenses. **Bid-Ask Spread** is measured by the difference between ETF daily ask and bid prices, scaled by the mid-quote. **Ln(Turnover)** is the natural log of the number of ETF shares traded divided by the total number of shares outstanding. **Ln(Size)** is the natural log of ETF market capitalization, calculated as ETF price multiplied by the total number of shares outstanding. **Volatility** is the standard deviation of daily ETF returns over the past 22 trading days. Numbers in parentheses are *t*-statistics based on double-clustered standard errors at the ETF and the day level.

Panel A: Without Control Variables						
	Number of Active AP			S.D. of Changes in ShrOut		
	Low	Medium	High	Low	Medium	High
DCBS	0.317	0.580***	0.412***	0.445***	0.767***	0.423***
	(1.327)	(4.401)	(3.914)	(3.621)	(5.417)	(3.890)
No. of Obs	87,585	214,307	270,945	409,036	374,731	301,547
$Adj.\ R^2$	0.154	0.145	0.086	0.079	0.074	0.134
Panel B: With Control Variables						
DCBS	0.476	0.721***	0.416***	0.383***	0.768***	0.228**
	(1.392)	(3.058)	(3.343)	(2.747)	(5.007)	(1.994)
Expense Ratio	0.564	-70.401	-5.338	-4.368	-8.518	-10.131***
	(0.040)	(-1.206)	(-0.837)	(-1.037)	(-1.570)	(-2.601)
Bid-Ask Spread (bps)	0.018	0.008	-0.037	-0.014	0.030*	0.014
	(0.943)	(0.642)	(-1.244)	(-0.776)	(1.789)	(1.062)
Ln(Turnover)	2.061***	1.724***	0.488***	0.781***	0.800***	0.713***
	(6.373)	(5.048)	(6.187)	(2.857)	(6.128)	(3.904)
Ln(Size)	0.502	-2.624	-0.757***	-0.179	0.388**	-0.130
	(0.426)	(-1.103)	(-2.823)	(-0.469)	(1.968)	(-0.462)
Volatility	-2.436	0.470	-0.563**	-0.013	-0.949*	0.298
	(-1.582)	(0.265)	(-2.283)	(-0.013)	(-1.925)	(0.318)
No. of Obs	38,043	125,597	212,947	362,965	337,006	273,807
$Adj.\ R^2$	0.168	0.166	0.101	0.084	0.077	0.140

Appendix

Table A1 Variable Definitions

Variable	Definition
Mispricing	The percentage difference between ETF market prices and the underlying NAVs. A positive (negative) suggests that an ETF is traded at a premium (discount).
DCBS	The Markit Security Finance Daily Cost of Borrow Score, which is a number from 1 to 10 indicating the rebate/fee charged by the agent lender, where 1 (10) indicates the cheapest (most expensive) borrowing cost.
Expense Ratio	The amount that investors pay for the fund's operating expenses, expressed as a percentage of an ETF's average net assets.
Bid-Ask Spread	The difference between ETF daily ask and bid prices, scaled by the mid-quote.
Ln(Turnover)	The natural log of ETF shares traded divided by the total number of shares outstanding.
Ln(Size)	The natural log of market capitalization, calculated as ETF price multiplied by the total number of shares outstanding.
Volatility	The standard deviation of daily ETF returns over the past 22 trading days.
Number of APs	The number of total registered APs for an ETF.
Number of Active APs	The number of APs that have a non-zero dollar value of creation/redemption.
Creation Unit	The minimum number of ETF shares that an AP has to exchange with the ETF sponsor for each creation transaction.
Creation Fee	The simple average of "in-kind" creation and cash creation fees.
Ret_t	The return on the cum-dividend day (the day before ex-dividend day).
$Ret_{t-2,t}$	The three-day buy-and-hold return up to the cum-dividend day (the day before the ex-dividend day).
$Ret_{t-4,t}$	The five-day buy-and-hold return up to the cum-dividend day (the day before the ex-dividend day).
Missing(%)	The percentage of ETF underlying stocks with missing DCBS data.
Utilization	The number of shares on loan, divided by the number of shares available for lending.

Table A2 AP Creation v.s. No AP Creation (Robustness)

This table reports regression estimates of ETF mispricing on ETF borrowing cost, which is proxied by DCBS. "Creation" columns refer to days when there is an increase in the number of shares outstanding led n days. "No Creation" columns refer to days when there is no change in the number of shares outstanding led n days. **Mispricing** is computed as the percentage difference between the ETF market price and the underlying NAV. A positive (negative) suggests that an ETF is traded at a premium (discount). **DCBS** is the Markit Security Finance Daily Cost of Borrow Score, which is a number from 1 to 10 indicating the rebate/fee charged by the agent lender, where 1 (10) indicates the cheapest (most expensive) borrowing cost. **Expense Ratio** represents how much investors pay for the fund's operating expenses. **Bid-Ask Spread** is measured by the difference between ETF daily ask and bid prices, scaled by the mid-quote. **Ln(Turnover)** is the natural log of the number of ETF shares traded divided by the total number of shares outstanding. **Ln(Size)** is the natural log of ETF market capitalization, calculated as ETF price multiplied by the total number of shares outstanding. **Volatility** is the standard deviation of daily ETF returns over the past 22 trading days. Numbers in parentheses are t-statistics based on double-clustered standard errors at the ETF and the day level. The sample period is from July 2001 to September 2022.

	Lead 1 Day		Lead 2 Days		Lead 3 Days		Lead 4 Days		Lead 5 Days		Lead 6 Days	
	Creation	No Creation	Creation	No Creation	Creation	No Creation	Creation	No Creation	Creation	No Creation	Creation	No Creation
DCBS	0.189**	0.503***	0.182**	0.499***	0.169*	0.528***	0.179**	0.515***	0.201**	0.519***	0.244***	0.517***
	(2.121)	(5.083)	(1.966)	(4.885)	(1.861)	(5.199)	(1.996)	(5.156)	(2.150)	(5.109)	(2.628)	(5.080)
Expense Ratio	-11.292*	-8.270**	-8.562	-8.147**	-10.914	-7.615**	-9.456**	-7.849**	-5.685	-8.079**	-7.987	-7.362**
	(-1.719)	(-2.363)	(-1.174)	(-2.235)	(-1.606)	(-2.188)	(-1.988)	(-2.302)	(-0.817)	(-2.180)	(-1.181)	(-2.104)
Bid-Ask Spread	0.044	0.009	0.062***	0.009	0.085**	0.010	0.094***	0.009	0.074**	0.008	0.094***	0.014
	(1.333)	(0.742)	(3.266)	(0.692)	(2.554)	(0.839)	(4.412)	(0.708)	(2.050)	(0.702)	(4.119)	(1.089)
Ln(Turnover)	0.258	0.766***	0.431***	0.703***	0.266*	0.729***	0.455***	0.699***	0.232	0.750***	0.340**	0.769***
	(1.628)	(4.065)	(2.721)	(3.972)	(1.655)	(4.253)	(3.091)	(3.963)	(1.601)	(4.261)	(2.570)	(4.370)
Ln(Size)	-0.048	-0.484	0.118	-0.526	-0.033	-0.496	-0.004	-0.505	-0.077	-0.494	0.009	-0.484
	(-0.254)	(-1.333)	(0.624)	(-1.472)	(-0.169)	(-1.407)	(-0.022)	(-1.427)	(-0.435)	(-1.393)	(0.055)	(-1.360)
Volatility	0.086	-0.252	-0.349	0.013	-0.608	0.135	-0.397	0.005	-0.244	0.118	-0.142	-0.064
	(0.105)	(-0.513)	(-0.356)	(0.029)	(-0.602)	(0.289)	(-0.486)	(0.010)	(-0.291)	(0.237)	(-0.180)	(-0.133)
ETF FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Day FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
No. of Obs	200,325	652,139	199,999	652,278	199,105	652,772	198,723	652,971	198,668	652,872	198,575	652,932
Adj. R^2	0.177	0.077	0.205	0.076	0.188	0.076	0.178	0.075	0.174	0.075	0.165	0.076

Table A3 ETF Mispricing on Dividend Record Day (2SLS)

This table reports regression estimates of ETF mispricing on DCBS projected by dividend record day. The regression is run within a 30-day $(t-15, t+15)$ window around ETF dividend record day. **Mispricing** is computed as the percentage difference between the ETF market price and the underlying NAV. A positive (negative) suggests that an ETF is traded at a premium (discount). **Record Day** is a dummy variable which is equal to 1 if day t is the dividend record day for ETF i. **Expense Ratio** represents how much investors pay for the fund's operating expenses. **Bid-Ask Spread** is measured by the difference between ETF daily ask and bid prices, scaled by the mid-quote. **Ln(Turnover)** is the natural log of the number of ETF shares traded divided by the total number of shares outstanding. **Ln(Size)** is the natural log of ETF market capitalization, calculated as ETF price multiplied by the total number of shares outstanding. **Volatility** is the standard deviation of daily ETF returns over the past 22 trading days. **Mean Holding Spreads** is the average bid-ask spread of stocks in ETF holdings. $\mathbf{Ret_t}$ is the return on the cum-dividend day (the day before the ex-dividend day). $\mathbf{Ret_{t-2,t}}$ is the three-day buy-and-hold return up to the cum-dividend day. $\mathbf{Ret_{t-4,t}}$ is the three-day buy-and-hold return up to the cum-dividend day. Numbers in parentheses are t-statistics based on standard errors clustered at the ETF-Dividend Event level. The sample period is from 2018 to 2022.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
$\widehat{DCBS}$	0.2429***	0.1860***	0.1858***	0.2462***	0.1549***	0.1480***
	(5.403)	(4.790)	(4.790)	(5.844)	(5.602)	(5.487)
Expense Ratio		0.0717**	0.0713**		0.0074	0.8366***
		(2.414)	(2.413)		(0.047)	(4.428)
Bid-Ask Spread		0.0013***	0.0012***		0.0006**	0.0012***
		(3.760)	(3.712)		(2.486)	(4.231)
Ln(Turnover)		0.0538***	0.0537***		-0.0026**	-0.0137***
		(4.778)	(4.793)		(-2.423)	(-3.655)
Ln(Size)		0.0567***	0.0564***		-0.0257***	-0.0107***
		(4.745)	(4.764)		(-3.789)	(-3.181)
Volatility		0.0137	0.0142		-0.0005	-0.0037
		(1.544)	(1.617)		(-0.088)	(-1.060)
Ret_t			-0.4501			-0.1622
			(-1.238)			(-1.057)
$Ret_{t-2,t}$			-0.0040			-0.0006
			(-1.175)			(-0.390)
$Ret_{t-4,t}$			0.0036			0.0016
			(1.417)			(1.195)
ETF FE	✗	✗	✗	✗	✗	✓
ETF-Dividend FE	✗	✗	✗	✓	✓	✗
No. of Obs	183,935	143,369	143,369	183,720	143,248	143,366

Table A4 ETF Borrowing Cost and Mispricing by Primary Market Activeness

This table reports regression estimates of ETF mispricing on ETF borrowing cost, conditional on different levels of ETF primary market activeness, which are proxied by two variables: the number of active APs (results shown in the left three columns) and the standard deviation of changes in daily number of ETF shares outstanding (results shown in the right three columns). **Mispricing** is computed as the percentage difference between the ETF market price and the underlying NAV. A positive (negative) suggests that an ETF is traded at a premium (discount). **DCBS** is the Markit Security Finance Daily Cost of Borrow Score, which is a number from 1 to 10 indicating the rebate/fee charged by the agent lender, where 1 (10) indicates the cheapest (most expensive) borrowing cost. **Expense Ratio** represents how much investors pay for the fund's operating expenses. **Bid-Ask Spread** is measured by the difference between ETF daily ask and bid prices, scaled by the mid-quote. **Ln(Turnover)** is the natural log of the number of ETF shares traded divided by the total number of shares outstanding. **Ln(Size)** is the natural log of ETF market capitalization, calculated as ETF price multiplied by the total number of shares outstanding. **Volatility** is the standard deviation of daily ETF returns over the past 22 trading days. Numbers in parentheses are *t*-statistics based on double-clustered standard errors at the ETF and the day level.

	Number of Active AP					S.D. of Changes in ShrOut				
	1	2	3	4	5	1	2	3	4	5
Panel A: Without Control Variables										
DCBS	0.133	0.336**	0.628***	0.663***	0.134***	0.374**	0.556***	0.837***	0.420***	0.453***
	(0.328)	(2.055)	(3.170)	(4.881)	(2.698)	(2.408)	(4.212)	(4.516)	(3.919)	(3.115)
No. of Obs	40,780	106,132	80,473	179,662	167,922	246,760	244,606	226,030	201,332	171,721
Adj. R^2	0.188	0.196	0.164	0.106	0.110	0.102	0.074	0.091	0.103	0.175
Panel B: With Control Variables										
DCBS	0.038	0.536**	0.631**	0.632***	0.165***	0.343*	0.528***	0.816***	0.364***	0.255*
	(0.061)	(2.111)	(2.488)	(4.142)	(3.032)	(1.757)	(3.813)	(4.271)	(3.062)	(1.733)
Expense Ratio	24.770	-79.701**	4.319	-18.525*	9.596***	-6.698	-3.645**	-21.769**	-11.054	-8.597*
	(1.174)	(-2.386)	(0.096)	(-1.789)	(2.608)	(-0.764)	(-2.119)	(-2.042)	(-1.490)	(-1.829)
Bid-Ask Spread (bps)	0.038	0.028	0.071**	-0.039	-0.017**	-0.092*	0.037	0.061**	0.010	0.083**
	(0.570)	(0.678)	(2.232)	(-1.285)	(-2.461)	(-1.927)	(1.518)	(2.116)	(0.213)	(2.105)
Ln(Turnover)	2.545***	1.650***	1.481***	1.276***	0.218***	0.699*	0.720***	0.800***	0.956***	0.634***
	(4.749)	(7.056)	(4.661)	(4.714)	(3.282)	(1.878)	(3.231)	(5.211)	(6.470)	(3.401)
Ln(Size)	4.390	0.626	-1.929*	-0.642	0.078	-0.768	0.434*	0.468**	0.436**	-0.237
	(1.516)	(0.535)	(-1.679)	(-0.968)	(0.458)	(-1.055)	(1.815)	(1.991)	(2.302)	(-0.562)
Volatility	-3.677*	0.918	-3.659***	0.007	-0.163	0.260	-0.209	-1.379**	-0.357	0.077
	(-1.878)	(0.580)	(-2.843)	(0.008)	(-0.888)	(0.154)	(-0.252)	(-2.113)	(-0.489)	(0.066)
No. of Obs	16,649	51,400	44,740	127,397	138,203	216,398	218,954	204,475	183,060	155,572
Adj. R^2	0.194	0.244	0.187	0.117	0.126	0.109	0.078	0.095	0.092	0.179