Underlying Valuation Uncertainty, Strategic ETF Creation and Pricing Efficiency*

Alan Guoming Huang
University of Waterloo

Russ Wermers
University of Maryland

Jinming Xue
Southern Methodist University

Xing (Alex) Zhou
Southern Methodist University

November 2, 2025

*Author emails are, respectively, aghuang@uwaterloo.ca, wermers@umd.edu, jinmingx@smu.edu, and axzhou@smu.edu. The authors gratefully acknowledge helpful comments from Pouya Behmaram, Nicole Boyson, Johan Cassel, Yong Chen, Jaewon Choi, Katie Gouinlock, Riste Ichev, Chotibhak (Pab) Jotikasthira, Jimmy Karam, Jiacui Li, Robert Macrae, Ananth Madhavan, Marco Rossi, Ke Shen, John Shim, Laura Starks, Karamfil Todorov, Kumar Venkataraman, David Xiaoyu Xu, Joey Zhang, Haibei Zhao, and seminar and conference participants at the 2025 NBER Financial Market Frictions and Systemic Risks Meeting, 2025 European Finance Association Conference, 2025 Inquire Europe Autumn Seminar, the International Research Seminar (IRS) 2025, 2025 MRS International Risk Conference, Four Corners Academic Meeting, Shanghai University of Finance and Economics, University of Cincinnati, University of Waterloo, and the Office of Financial Research at the Department of the Treasury. All errors are ours.

Underlying Valuation Uncertainty, Strategic ETF Creation and Pricing Efficiency

ABSTRACT

This paper identifies a novel mechanism through which uncertainty in corporate bond valuation shapes the creation process of corporate bond ETFs. Although multiple ETFs can potentially accept the same bond during the "in-kind" creation process, they often value it differently. Such valuation dispersion is driven by fund pricing conventions and external pricing sources, but can also reflect ETFs' efforts to minimize tracking error. Authorized Participants (APs) exploit these differences by delivering a bond to create the ETF that values it most highly. While this strategic behavior increases APs' profits, it dilutes holdings of incumbent investors and undermines ETF pricing efficiency.

JEL Classification: G12, G23

Keywords: ETF, corporate bond, creation basket, authorized participant, liquidity, valuation dispersion

1 Introduction

Exchange traded funds (ETFs) have seen an extraordinary expansion over the past decade, emerging as a major asset class in global markets. As of July 2025, the global ETF industry reached a record $17.34 trillion in assets under management (AUM).[1] Growth has been especially pronounced in products that hold illiquid underlying assets (traditionally corporate bonds and international equities, and more recently, bank loans, private debt, and private equity), whose values are difficult to assess due to the opacity or absence of active trading markets.[2]

Valuation uncertainty in illiquid securities creates distinct challenges for the ETF structure. A unique feature of ETFs is their primary market, where authorized participants (APs) create or redeem shares in exchange for baskets of underlying assets, thereby aligning ETF share market prices with the pricing of the underlying securities (the net-asset-value, or NAV, of the ETF). When asset values are difficult to estimate, different ETFs may assign different prices to the same security, leading to variation in the exchange ratios between ETF shares and the underlying securities. Because ETFs backed by illiquid assets often permit flexibility in how APs design creation baskets,[3] these valuation discrepancies across ETFs raise an important question: how does valuation uncertainty influence APs' creation decisions, and what are the implications for ETF investors and pricing efficiency?

In this paper, we use corporate bond ETFs as a laboratory to examine how uncertainty in the valuation of underlying assets affects ETF outcomes. We start by showing that an AP can maximize its expected profitability from the creation process by delivering a bond to create the ETF that has the highest pricing of the bond. To illustrate, Table I reports the recorded prices on September 28, 2022 of a bond issued by Oracle that was held by fifteen

[1] See the press release by ETFGI on August 27, 2025: `https://etfgi.com/news/press-releases/2025/08/etfgi-reports-assets-invested-etfs-industry-globally-reached-new-0`.

[2] In 2024, global bond ETF assets rose 20% to $2.6 trillion, the fastest pace of growth across all ETF categories. BlackRock projects that global bond ETF assets will reach $6 trillion by 2030. See `https://www.etf.com/sections/news/fixed-income-etf-assets-hit-6t-2030-blackrock`.

[3] See, for example, Shim and Todorov (2021), and Koont, Ma, Pastor, and Zeng (2025).

different bond ETFs on that day, along with information on whether the bond was included in the creation basket for each ETF that day. These ETFs assigned different values to the bond, with prices ranging from 67.90 to 68.52, a dispersion of almost one percent. Among the fifteen ETFs, six experienced creations on that day. Remarkably, however, the Oracle bond was included in only one ETF's creation basket: that of the iShares iBoxx $ Investment Grade Corporate Bond ETF (LQD), which happened to assign it the highest value among all 15 ETFs.

The case of the Oracle bond on that September day is not an isolated incident. Rather, it illustrates a broader mechanism through which APs allocate underlying bonds when confronted with cross-ETF valuation discrepancies during the creation of new ETF shares. Over our sample period, a bond is, on average, held by roughly 10 ETFs at any given time. The prices reported by different ETFs for a given underlying bond on the same day vary considerably, with the interquartile range and standard deviation of cross-ETF price deviations averaging about 19 and 16 basis points (bps) across all bond-days, respectively. As we show, these large price discrepancies across ETFs are related to the (il)liquidity of the underlying bonds; that is, ETFs are more likely to have differential bond prices when a bond trades infrequently (or not at all). Valuation discrepancies tend to be larger for older, lower-rated, smaller-issue, and longer-maturity bonds. In the time series, cross-ETF bond valuation dispersion spiked during the COVID-19 pandemic, when the corporate bond market experienced a liquidity shock.[4] Nevertheless, significant valuation dispersion is evident throughout the period from 2017 to the present.

The cross-ETF dispersion in valuation of the same underlying bond that we document creates an opportunity that APs can exploit during the creation process. For a given underlying bond, all else equal, an AP can deliver it (as part of the creation basket) for a

[4]While valuation dispersion for investment-grade (IG) bonds is generally lower than that for high-yield (HY) bonds, IG valuation dispersion skyrocketed at the peak of the COVID liquidity crisis, surpassing that observed for HY bonds. This aligns with the greater deterioration of liquidity in the IG sector during that period.

greater number of ETF shares if the ETF issuer assigns a higher value to the bond, relative to another ETF issuer's valuation of the bond. It is important to note that the AP's potential profitability from selecting a particular ETF with which a particular bond is exchanged for ETF shares also depends on the overall valuation approach of the selected ETF. The advantage of creating ETF shares by delivering bonds to the ETF that most highly values them (relative to other ETFs) diminishes, or may even vanish, if that ETF inflates the prices of all bonds in its portfolio by the same proportion, resulting in a NAV that with a price inflated by the same proportion.[5] The key is having a *subset* of bond holdings available that are overvalued, relative to other holdings, which can then be selectively included in a creation basket by the AP—with the expectation that a subsequent redemption basket will not contain the same subset at the same inflated value. Using a simple example, we illustrate that APs gain the most from delivering a bond to an ETF that assigns it the highest value while maintaining a lower relative (to other ETFs) valuation of other bonds in its portfolio.

Guided by this prediction, we construct a fund-bond-day panel restricted to ETFs that experience positive inflows (and, therefore, APs deliver creation baskets to each ETF sponsor on that day) to examine whether APs direct bonds toward the ETFs that value them most highly. Our identification strategy relies on several stringent fixed effects. One concern is that the likelihood of a bond's inclusion in an ETF's creation basket may be driven by deviations between the ETF's market price and its NAV. When an ETF trades at a premium, to the extent that the NAV reflects the market price of the underlying portfolio, the premium creates an arbitrage opportunity for APs, increasing the probability (with the size of the

[5]Take, for example, an ETF that uniformly prices all underlying holdings at double their respective market values. That ETF's reported NAV would correspondingly rise by 100 percent, and creation or redemption would take place at that higher NAV without affecting the number of shares created or redeemed. However, sophisticated market participants would recognize that this inflation in valuations in the primary market is purely mechanical. As a result, the ETF's secondary-market price would remain unaffected, even though the inflated NAV would make the ETF appear to trade at a substantial discount. Because both the ETF and its underlying securities are scaled upward proportionally, no profitable opportunity would emerge through the creation or redemption mechanism; this fact holds even if all underlying bond prices are (proportionately) scaled upward upon delivery for ETF creation, and downward upon ETF redemption. This outcome contrasts with mutual funds, where investors can redeem (purchase) shares at NAV for cash if valuations are inflated (depressed) relative to the market value of underlying portfolio holdings.

premium) that the AP will deliver a particular bond as part of a creation basket in return for ETF shares on that day. To absorb such influences, as well as other time-varying ETF-level characteristics such as ETF size, fees, and bid-ask spreads, we include fund-day fixed effects in our regressions. Another potential concern is that shocks to individual bonds, such as fire sales triggered by credit events or liquidity disruptions, may alter dealer inventories and, thereby, increase the likelihood of those bonds being delivered to any ETF. To control for such bond-specific, time-varying shocks, we include bond–day fixed effects. Finally, to account for persistent ETF preferences for certain bonds, we incorporate fund–bond fixed effects. Together, these fixed effects allow us to isolate the impact of valuation dispersion on APs' creation decisions from other influences at both the fund and bond levels.

With this design, we find strong support for the mechanism that we document. A bond is significantly more likely to be included in the creation basket of the ETF that assigns it the highest value, relative to the ETF's overall valuation tendency. Quantitatively, bonds priced above the cross-ETF median are 6.72 percentage points more likely to be delivered than those priced below, compared to the unconditional probability of inclusion of a bond in a creation basket of only 9.6%. This economically significant effect underscores the role of valuation dispersion in shaping AP creation decisions.

Beyond the robust fixed-effects models noted above, we conduct a series of tests to ensure the robustness of our findings. First, a fund-by-fund analysis shows that the positive relation between valuation dispersion and creation basket membership holds across the vast majority of ETFs.[6] Second, we assess the robustness of our results with respect to the valuation dispersion measure. We find that our conclusions remain largely unchanged regardless of whether ETF holdings are converted to bid-based valuations, mid-based valuations, or left under their own reporting convention. Third, although our main analysis focuses on fund–days with net inflows, which are most relevant for creation activity, we also include

[6]We further validate our valuation dispersion measure using supplementary data from the iShares website, and show that our results remain robust when restricting the sample to ETFs issued by BlackRock.

fund–days with net outflows and even those with no flows—since creations may still occur on such days—and find consistent results. Fourth, we show that APs' strategic behavior arises only in the creation process, but not in the redemption process where ETF issuers exercise greater control. Finally, our results remain robust when we employ alternative measures of cross-ETF price differences, and when we examine the relationship over time.

What drives the dispersion in valuations across ETFs? We find that both ETF issuers and, more importantly, tracking index providers play a role.[7] A key factor is the practice of tracking index providers, who typically supply indicative or end-of-day prices for the bonds that constitute the underlying indexes. However, the bond valuations used by an ETF may not always align with those provided by the index provider. For example, we compare bond prices reported by LQD with those provided by Markit for the iBoxx Liquid Investment Grade $ Index, the benchmark tracked by LQD. We find that LQD's prices generally align closely with the bid prices from the index. Yet, when discrepancies occur, even small deviations can have a substantial impact on the bonds included in LQD's creation basket.[8]

We explore the implications of APs' strategic behavior in the context of cross-ETF valuation dispersion for both ETF pricing and investor welfare. Our findings show that gains from such strategic behavior by APs during the creation process contribute meaningfully to APs' overall arbitrage profits, but at the cost of potentially undermining ETF price efficiency and harming existing investors. APs' gains come directly at the expense of existing investors: when an ETF issuer accepts bonds at valuations higher than those of its peers, existing shareholders experience dilution. Moreover, we find that when an ETF trades at a discount, typically in the presence of selling pressure, the incentives created by bond price

[7] For instance, Table I shows that all iShares ETFs and State Street (SPDR) ETFs tracking Bloomberg indexes value the Oracle bond similarly. In contrast, WisdomTree, Schwab, and Hartford ETFs that also track Bloomberg indexes, as well as iShares ETFs tracking ICE and Markit iBoxx indexes, assign noticeably different values to the same bond.

[8] Our analysis suggests that these price adjustments likely aim to minimize tracking error by selectively attracting certain underweighted bonds into LQD's portfolio; nevertheless, this tracking-error concern provides a profitable opportunity for APs.

dispersion persist, encouraging share creations despite the oversupply of ETF shares. As a result, ETF prices deviate further from fundamentals, exacerbating pricing inefficiencies.

Our paper advances the understanding of APs, a unique class of intermediaries that play a central role in the functioning of ETF markets. APs are commonly viewed as arbitrageurs whose primary incentive is to profit from discrepancies between an ETF's market price and the value of its underlying holdings. These arbitrage incentives underpin a large body of research examining how ETFs influence their underlying markets. Ben-David, Franzoni, and Moussawi (2018) show that liquidity shocks can propagate to the underlying securities through the ETF arbitrage mechanism, thereby amplifying volatility in the underlying stocks. Additional studies have also documented that ETFs can affect the underlying assets, including in bond markets.[9] However, this view was challenged by Box, Davis, Evans, and Lynch (2021) who use intraday data to show that ETF trading has limited impact on the underlying stocks. Instead, shocks originating in the underlying market drive adjustments in ETF prices.

A growing strand of research examines how APs' behavior is shaped by liquidity conditions, transaction costs, and inventory management considerations, reflecting their dual roles as arbitrageurs and market makers in the underlying securities. Much of this work centers on bond ETFs, where liquidity frictions are more pronounced. Pan and Zeng (2019) show that APs' balance sheet constraints can limit their arbitrage capacity when shocks arise in the underlying corporate bond market. Shim and Todorov (2023) document that APs act as a stabilizing buffer between ETFs and their underlying bonds, absorbing redemption pressure by holding bonds to prevent fire sales. Koont et al. (2025) argue that bond ETFs facilitate liquidity transformation by actively adjusting their creation and redemption baskets to minimize the transaction costs incurred by APs when trading the underlying assets. Du (2023) and Dannhauser and Karmaziene (2023) show that APs use ETFs to manage inventories in

[9]For additional studies on equity ETFs, see, for example, Bhattacharya and O'Hara (2018), Da and Shive (2018), Sağlam, Tuzun, and Wermers (2025), and Glosten, Nallareddy, and Zou (2021). For studies on bond ETFs, see, for example, Dannhauser (2017), and Dannhauser and Hoseinzade (2022).

corporate bond markets. For equity ETFs, Evans, Moussawi, Pagano, and Sedunov (2022) demonstrate that APs' ability to short sell ETF shares prior to creation plays a central role in facilitating ETF liquidity provision. Related studies further highlight that the liquidity of the underlying assets affects the effectiveness of the ETF arbitrage mechanism.[10]

Building on these insights, our study examines how valuation uncertainty, a pervasive feature of illiquid and opaque asset markets, creates distinctive incentives for strategic AP behavior. We show that such behavior can materially affect ETF pricing and investor welfare. While prior research has largely focused on how ETFs influence the underlying assets, our analysis emphasizes the reverse channel, i.e., how conditions in the underlying market shape ETF dynamics through AP behavior. In this sense, our study complements Box et al. (2021) and Reilly (2022) and contributes to a deeper understanding of the interconnections between ETFs, their underlying asset markets, and the broader ETF ecosystem.[11]

More broadly, our work also relates to the literature on bond valuation disagreement and its implications for asset pricing and market efficiency. Hu, Pan, and Wang (2013) show that price deviations in U.S. Treasury bonds reflect the availability of arbitrage capital and contain valuable information about market liquidity. Jankowitsch, Nashikkar, and Subrahmanyam (2011) document that illiquidity in the corporate bond market generates price dispersions for the same bond on the same day. Feldhütter (2012) demonstrates that prices differ across trader types and uses these dispersions to identify selling pressure in the corporate bond market. O'Hara and Zhou (2021) examine same-bond, same-day dispersions across dealers to assess execution quality in corporate bond transactions. Eisfeldt, Herskovic, and Liu (2024) show that interdealer price dispersion stems from frictions within the dealer sector, leading to a misallocation of credit risk among intermediaries and constituting a priced risk factor in corporate bonds.

[10]See, for example, Holden and Nam (2024), Finnerty, Reisel, and Zhong (2025), Brogaard, Heath, and Huang (2025).

[11]For additional studies on the relationship between ETFs and their underlying markets, see, for example, Ben-David, Franzoni, Kim, and Moussawi (2023), Andrews, Henderson, and Reed (2024), and Evans, Karakaş, Moussawi, and Young (2025).

We build on this literature by focusing on the implications of bond valuation dispersion for ETFs, a rapidly growing segment of financial markets. We show that same-day, same-bond valuation differences across ETFs give rise to strategic behavior by APs, linking micro-level pricing frictions in bond markets to ETF-level outcomes. In this sense, our paper is also related to research examining how valuation dispersion affects the behavior of mutual fund participants.[12] Cici, Gibson, and Merrick Jr (2011) show that mutual fund managers strategically mark their bond holdings, and thus their fund NAVs, to smooth reported returns. Choi, Kronlund, and Oh (2025) argue that bond mutual funds adjust the prices of their holdings, and consequently their NAVs, in response to investor flows, often without explicit disclosure. Our study differs by focusing on how valuation discrepancies at the holdings level influence AP behavior, a feature unique to the ETF market's in-kind creation and representative sampling mechanism. More importantly, we examine how the strategic behavior of APs impacts ETF pricing efficiency and investor welfare, an aspect not addressed in previous research.

This paper is organized as follows: Section 2 provides the institutional background for our study. Section 3 introduces a simple illustration of the mechanism by which APs allocate underlying bonds to different ETFs with varying valuations during the creation process. Section 4 describes the data and sample construction. Section 5 presents an empirical analysis of the impact of cross-ETF bond valuation dispersions on APs' creation baskets. Section 6 explores the drivers behind bond valuation dispersions across ETFs. Section 7 examines the impact of APs' strategic behavior on ETF flows and price efficiency. Finally, Section 8 concludes.

[12] Broadly related to this literature is research that examines the impact of mutual fund liquidity transformation on information collection in the underlying security. For example, Sialm and Xu (2025) find that during trading suspensions, illiquid Chinese stocks held by mutual funds with stale valuations experienced greater investor information acquisition and attracted fund flows.

2 Institutional Background

The rapid growth of corporate bond ETFs underscores their increasingly important role in bond markets, facilitating liquidity transformation and serving both investors and dealers. For institutional and retail investors, direct participation in the corporate bond market is often costly due to its fragmentation, limited transparency, and low liquidity. ETFs mitigate these frictions by providing low-cost, diversified exposure to broad segments of the bond market, as well as intraday liquidity and transparent pricing. At the same time, corporate bond ETFs have become an important liquidity management tool for dealers. The illiquid nature of corporate bonds has long posed challenges for dealers engaged in market-making, particularly during periods of stress. By exchanging illiquid bonds for liquid ETF shares, dealers can offload large inventories, reduce balance sheet risk, and sustain their intermediation capacity.

A distinctive feature of ETFs is the in-kind creation process, which allows APs, typically broker-dealers or large institutional investors, to exchange a basket of underlying securities for ETF shares. In the primary market, when new shares of an ETF are created, an AP delivers a specific basket of securities to the ETF issuer. In return, the AP receives ETF shares, which can be sold or traded on an exchange. This process is referred to as "in-kind" because the AP receives, instead of cash, ETF shares that represent a pro-rata stake in the ETF's underlying portfolio.

Corporate bond ETFs often employ representative sampling in their creation baskets, which allows the APs to create shares of the ETF without needing to deliver all the bonds held by the ETF. Instead, the creation basket typically contains a smaller, negotiated subset of the ETF's underlying bonds, along with a cash balancing amount.[13] This "fractional

[13] This is because a pro-rata replication of the underlying portfolio of a bond ETF is not efficient or even infeasible. For instance, as of Nov 15, 2024, the iShares Core US Aggregate Bond ETF has 12,101 holdings. It is nearly impossible for an AP to source all these bonds in proportional amounts to create a basket. The cash balancing amount accounts for the difference between the value of the bond basket and the NAV of creation units.

basket" feature enables flexibility in how the underlying bonds are selected and delivered during the creation process. Such flexibility fosters AP arbitrage activities, which help minimize premiums and discounts in the ETF's secondary market price. It also offers ETF issuers an efficient means for portfolio rebalancing. For APs who also serve as bond dealers, this system presents an opportunity to overweight bonds that they wish to offload.[14]

Since many corporate bond ETFs track similar indices or focus on similar segments of the bond market (e.g., investment-grade or high-yield corporate bonds), a dealer can choose to deliver the same bonds to multiple ETFs when creating new shares. Moreover, given the relatively small number of APs active in the corporate bond ETF market, dealers often serve as APs for several competing ETFs.[15] This market concentration means that the same bond could be accepted into the creation baskets of multiple ETFs, which may influence the AP's decisions regarding which ETF to use for the creation process.

One of the critical dynamics in the corporate bond ETF market is the valuation of bonds. While ETFs may hold similar bonds, the valuation of those bonds can differ significantly across ETFs. This is due to various factors, including differences in the data sources and pricing models used by ETF issuers, as well as the inherent illiquidity of corporate bonds, which are not continuously traded. As a result, different ETFs may assign different prices to the same bond, resulting in valuation dispersion across ETFs.

[14] For example, when an ETF faces strong investor demand, an AP can include more of the bonds it needs to dispose of in the creation basket, thus optimizing the ETF creation process to meet its liquidity needs. This is especially valuable for dealers holding large inventories of bonds that are difficult to sell in the secondary market but can be included in the ETF's creation basket as part of the representative samples.

[15] See, for example, a staff analytical note released by the Bank of Canada in 2020: `https://www.bankofcanada.ca/2020/11/staff-analytical-note-2020-27/`. Du (2023) documents that top three APs account for 82% of creations and redemptions for corporate bond ETFs. See Gorbatikov and Sikorskaya (2022) for an analysis for primary market concentration in ETFs across different asset classes.

3 A Simple Example of the Strategic Creation Mechanism under Bond Valuation Uncertainty

Differences in bond valuations play a critical role in the decisions made by APs when selecting an ETF for the creation process. APs are incentivized to deliver bonds within creation baskets to ETFs that assign relatively higher values to those bonds. Table II illustrates how cross-ETF dispersion in bond valuations influences the composition of custom creation baskets selected by APs, and serves to guide the structure of our empirical tests.

Consider two bonds, Bond 1 and Bond 2, held by four ETFs (A, B, C, and D), which all maintain identical portfolios consisting of one unit of each bond. However, the reported prices differ across the four ETFs: for Bond 1, the prices are 102, 105, 105, and 105, respectively; while for Bond 2, the prices are 110, 110, 113, and 107, respectively.

Fixed income ETFs typically allow for custom creation baskets that need not replicate the exact portfolio holdings and may include a cash component to reconcile any differences between the basket's value and the ETF's NAV. Assume that an AP decides to create one ETF share using only two units of Bond 1. Given the variation in reported bond prices across the ETFs, the required cash adjustment in the creation basket will differ depending on which ETF is selected. Consider ETFs A and B, both of which assign a value of 110 to Bond 2. However, ETF A values Bond 1 at 102, while ETF B assigns it a higher value of 105. As a result, the NAVs of the two ETFs differ: 212 for ETF A and 215 for ETF B. Suppose an AP submits a custom creation basket consisting solely of two units of Bond 1. Under ETF A's pricing, the basket is valued at 204, requiring an additional cash component of 8 to reach the NAV. In contrast, the same basket is valued at 210 for ETF B, requiring only 5 in cash. Since the two ETFs hold identical portfolios, the AP would prefer to create shares in ETF B, as it minimizes the required cash contribution.

It is important to note that an AP's decision about which ETF to use for share creation also depends on how the ETF values *other* bonds in its portfolio. This is because the valuations of all portfolio holdings determine the ETF's NAV, which in turn affects the size of the cash component required in the creation basket. For example, ETF C assigns values of 105 to Bond 1 and 113 to Bond 2, both higher than those reported by ETF A. Nonetheless, when an AP creates shares using two units of Bond 1, the required cash component for ETF C remains 8, the same as for ETF A. This is because the higher valuation of Bond 2 increases ETF C's NAV to 218, fully offsetting the higher valuation it gives to Bond 1. As a result, despite its higher bond prices, ETF C is less favorable than ETF B, which requires only 5 in cash.

This example underscores that what ultimately matters is not an ETF's relative valuation of a particular bond across funds, but its relative valuation of that bond *within* the ETF's own portfolio. That is, how favorably the ETF prices the bond relative to its peers, compared to how it values its other holdings relative to its peers. This concept is further reinforced by the case of ETF D, which values Bond 1 at 105 and Bond 2 at 107, resulting in a NAV of 212. When an AP contributes two units of Bond 1 (valued at 210 by ETF D), the required cash component is only 2, making ETF D the most favorable option among the four.[16]

This mechanism depends on the relative valuation dispersion of bonds within each fund. Specifically, the extent to which an ETF's valuation of a given bond deviates from the valuations assigned by its peers, adjusted for the ETF's overall pricing tendency (i.e., whether it systematically prices bonds higher or lower than other ETFs). Accordingly, we hypothesize that APs are more likely to include a bond in an ETF's creation basket when that bond is valued relatively higher by the ETF compared to its other holdings.

To quantify an ETF's within-fund relative valuation of a bond, we first calculate the ETF's valuation of the bond relative to its peers and then adjust this measure by subtracting the ETF's average relative valuation across all bonds in its portfolio. We then link this

[16]A symmetric analysis using Bond 2 to create ETF shares suggests that ETF D is the least favorable option, while ETFs A and C are the most favorable. This is because A and C assign relatively higher valuations to Bond 2 compared to the other bonds in their respective portfolios.

adjusted relative valuation to the size of the required cash component in the creation basket. Specifically, for each bond held by an ETF, we first calculate the price deviation (PD) as the difference between the bond price reported by the ETF and the median price of that bond reported by all ETFs. We then adjust the PD at the bond-fund level by subtracting the average PD for Bonds 1 and 2 within the given fund. For example, ETF B reports prices of 105 for Bond 1 and 110 for Bond 2, which match the respective medians across all ETFs. Consequently, both the PDs and the ETF-level average PD are zero, yielding an adjusted PD of zero for Bond 1 by ETF B. In contrast, although the PD for Bond 1 is also 0 in ETFs C and D, their ETF-level adjustments are 1.5 and −1.5, respectively, because ETF C and ETF D value Bond 2, respectively, at 113 and 107. Consequently, the adjusted PD for Bond 1 is −1.5 in ETF C and 1.5 in ETF D. Similarly, the adjusted PD for Bond 1 in ETF A is −1.5, as its reported value for Bond 1 is below the median across the four ETFs.

These adjusted PDs correspond to the required cash component when using two units of Bond 1 to create one ETF share. Bond 1 in ETF D has the highest adjusted PD (1.5) and corresponds to the smallest required cash component (2), followed by ETF B, which has the second-highest adjusted PD and a cash component of 5. In contrast, ETFs A and C have the lowest adjusted PDs (−1.5), each requiring the largest cash component of 8.[17] Taken together, Table II demonstrates the ETF strategic creation mechanism under bond valuation uncertainty: the adjusted PD of a bond held by an ETF captures the relative attractiveness of including that bond in the creation basket. A higher adjusted PD reflects a more favorable within-fund relative valuation, which reduces the required cash component, making the bond more appealing and more likely to be included in primary market creation activity.

[17]Similarly, when Bond 2 is used to create new shares, it is optimal to do so with ETFs A and C, which assign relatively higher valuations to Bond 2 compared to their other holdings.

4 Data and Sample Construction

Our study primarily utilizes a combination of security-level holdings data and fund-level flow data for corporate bond ETFs. We construct our sample by initially selecting all U.S.-domiciled ETFs, including those that have been delisted, and narrowing our focus to "U.S. Fixed Income" funds in Morningstar with substantial investments in corporate bonds.[18] Target date ETFs are excluded, as these funds tend to increase cash holdings as they approach their closure dates (Koont et al. (2025)). We also exclude Vanguard ETFs, as they do not report daily holdings to Morningstar, and because some of these ETFs are structured as share classes of existing mutual funds.[19] For each corporate bond ETF, we gather historical daily portfolio holdings from Morningstar covering the period from 2017 to 2023.[20] Along with holding par amounts, the data also include each ETF's estimated valuation for every bond it holds. Our sample comprises 90 corporate bond ETFs holding a total of 29,001 unique bonds, resulting in 157,143,077 fund-bond-day observations. See Internet Appendix B for a detailed description of the ETF selection process.

Because actual creation baskets are not directly observable, we adopt a method consistent with the literature by assuming that creations occur on days with net flows (e.g., Shim and Todorov (2023), Koont et al. (2025)).[21] We identify these flow days using changes in the daily number of shares outstanding, as reported by Bloomberg. If creations were to occur on a day with zero net flow, it would suggest that creations and redemptions perfectly offset each other, which is unlikely. While creations may occur on days with net outflows, we

[18]Specifically, we choose the following seven Morningstar categories of funds: "US Fund Corporate Bond," "US Fund High Yield Bond," "US Fund Long-Term Bond," "US Fund Intermediate Core Bond," "US Fund Intermediate Core-Plus Bond," "US Fund Multisector Bond," and "US Fund Short-Term Bond."

[19]For instance, on February 29, 2024, Vanguard Total Bond Market Index Fund, a mutual fund with an AUM of $314 billion, has five mutual fund share classes and one ETF share class—Vanguard Total Bond Market Index Fund ETF (ticker: BND) with an AUM of $105 billion.

[20]Our sample starts in 2017, as there are few corporate bond ETFs prior to 2017.

[21]Depending on whether an ETF reports its end-of-day holdings before or after that day's creation/redemption activity, discrepancies may arise between Morningstar's holdings data and Bloomberg's shares outstanding data. Following Koont et al. (2025), we adjust for these timing differences. See Internet Appendix C for a detailed description of our correction procedure.

focus on net inflow days, where at least one creation must have taken place. This procedure yields a creation-only sample comprising 42,508,900 fund–bond–day observations.[22]

For bonds held by the ETFs in our sample, we obtain detailed information on each bond's characteristics from the Mergent Fixed Income Securities Database (FISD), including issuance and maturity dates and credit rating history. The FISD provides a comprehensive record of rating changes from the three major rating agencies: Standard & Poor's (S&P), Moody's, and Fitch. We use this data to calculate for each bond a daily composite rating that integrates the ratings from these agencies.[23]

Table III provides summary information about the ETFs and their bond holdings. On average, the ETFs in our sample have a size of approximately $3 billion. However, the distribution is highly skewed, with a median size of only $395 million, indicating the presence of a few large ETFs. During the sample period, ETFs trade at a slight premium: the average (median) ETF premium (Price/NAV$-$1) is 0.10% (0.08%). Notably, price deviations from NAV occur, with the bottom and top deciles reflecting a 0.15% discount and a 0.39% premium, respectively. Panel B shows that over three-quarters of the bonds are investment-grade, with an average (median) remaining time to maturity of 10.37 (6.59) years and an average (median) age of 4.23 (3.21) years.

5 Bond Valuation Dispersion and Creation Basket

In this section, we empirically test the hypothesis developed in Section 3. We demonstrate that the mechanism by which adjusted bond valuation dispersion across ETFs influences APs' creation basket decisions is broadly evident in our sample. We start by estimating the

[22]As shown in Section 5.3.1, our results are robust to the inclusion of days with net outflows as well as days without any net flows.

[23]Specifically, we assign a numeric value to each grade of an S&P rating, where 1, 2, 3, 4, ... represent AAA, AA, A, BBB, ..., respectively. We apply the same method to convert Moody's and Fitch ratings into numeric values. If a bond has only one rating, that rating serves as its composite rating. For bonds rated by two agencies, we use the lower of the two ratings as the composite. For bonds rated by all three agencies, the composite rating is determined by the median of the three ratings.

bond valuation dispersions across ETFs and analyzing the factors influencing these dispersions. We then adjust each ETF's bond-level valuations to account for its overall valuation tendency and examine how these adjusted bond valuations affect the composition of APs' ETF creation baskets. We conclude the section with a series of robustness checks and additional analyses.

5.1 Dispersion of bond valuations across ETFs and bond liquidity

To measure how an ETF's valuation of a specific bond differs from that of other ETFs holding the same bond on the same day, we begin by inferring each ETF's bond-level valuation from security-level holdings data. Morningstar reports both the market value and par amount of each bond position held by each ETF. We calculate the implied price assigned by an ETF to a bond by dividing the reported market value by the par amount. Although primary market transactions are based on dirty prices (which include accrued interest), some ETFs report holdings using clean prices. We identify each ETF's pricing convention and adjust clean prices by adding accrued interest to ensure comparability. Internet Appendix D provides a detailed description of our methodology.

Using the prices inferred from each ETF's daily holdings, we then estimate a fund-bond-date level price deviation measure ($PD_{f,b,t}$):

$$PD_{f,b,t} = P_{f,b,t} - \widetilde{P_{b,t}} \tag{1}$$

where $P_{f,b,t}$ is the price for bond b reported by ETF f on day t, and $\widetilde{P_{b,t}}$ is the median price of bond b reported by all ETFs that hold the bond on day t.

On average, a bond is held by ten ETFs on any given day (Table IV). For each bond on each day, we compute a price dispersion measure, $\textit{PD Dispersion}_{b,t}$, defined as either the standard deviation or the interquartile range of price deviations ($PD_{f,b,t}$) reported by all ETFs holding bond b on day t. As shown in Table IV, the average $\textit{PD Dispersion}_{b,t}$ is

19.16 bps when measured using the interquartile range and 16.01 bps when measured using the standard deviation. Price dispersion varies substantially across bonds and days, with standard deviations of 36.94 bps and 21.93 bps, respectively, for the interquartile-based and standard deviation-based measures.

Bond illiquidity appears to significantly contribute to price dispersion across ETFs. Figure 1 presents the weekly time series of the median of $\textit{PD Dispersion}_{b,t}$ throughout our sample period. Panel (a) of Figure 1 indicates that the average price dispersion across ETFs typically fluctuates around 20 basis points, with the exception of the weeks of March 14 and March 21, 2020, when the corporate bond market faced a liquidity crisis triggered by the COVID-19 pandemic.[24] During this period, price dispersion spiked to over 190 basis points. In addition, in line with the greater deterioration of liquidity in the IG sector,[25] Panel (b) shows that although price dispersion for IG bonds is typically lower than that for HY bonds during normal market conditions, it experienced a sudden surge, significantly exceeding the price dispersion for HY bonds at the peak of the COVID liquidity crisis.

We also investigate the relationship between price dispersion and bond illiquidity in the cross section. A key factor contributing to divergence in bond valuation is the lack of continuous trading. To capture the impact of trading activity on price dispersion, we obtain transaction data for each bond in our sample from the Trade Reporting and Compliance Engine (TRACE), provided by the Financial Industry Regulatory Authority (FINRA). The TRACE data offers detailed trade-level information for each corporate bond transaction in the secondary market, including bond identifiers, trade execution dates and times, prices, quantities, and other relevant details. Using the TRACE data with ETF holdings data, we

[24]See, for example, O'Hara and Zhou (2021), Kargar, Lester, Lindsay, Liu, Weill, and Zúñiga (2021), and Boyarchenko, Kovner, and Shachar (2022).

[25]See, for example, Haddad, Moreira, and Muir (2021), and Ma, Xiao, and Zeng (2022).

construct a bond-day level sample and estimate the following empirical model:

$$\begin{aligned} PD\,Dispersion_{b,t} &= \beta_1 \times \text{Traded Dummy}_{b,t} + \beta_2 \times \text{Age}_{b,t} + \beta_3 \times \text{Issue Size}_{b,t} \\ &+ \beta_4 \times \text{Credit Rating}_{b,t} + \beta_5 \times \text{Time to Maturity}_{b,t} + \mu_i + \mu_t + \varepsilon_{b,t} \end{aligned} \tag{2}$$

where Traded Dummy$_{b,t}$ is a dummy variable that indicates whether bond b is traded on day t based on the TRACE data. Additionally, we incorporate liquidity measures based on bond characteristics, such as Age and Issue Size, as newer bonds and larger issues are generally more liquid.[26] We also control for a bond's credit rating and time to maturity. Finally, we include issuer and day fixed effects to account for potential influences from unobservable issuer characteristics and macroeconomic conditions. Standard errors are double-clustered at the bond and date levels.

Consistent with the notion that greater liquidity reduces price dispersions, Column (1) of Panel A, Table V shows that price dispersion among ETFs declines when a bond is more heavily traded, and is smaller for new bonds and for larger issues. In addition, lower rated and long-term bonds face larger price dispersions. We replace the issuer and day fixed effects with issuer-day fixed effects to control for time-varying issuer heterogeneities, and obtain similar results (Column (2)). Our results are qualitatively the same when we replace the bond trading dummy with the logarithm of the number of trades for bond b on day t. Columns (3) and (4) show that price dispersion across ETFs decreases as trading frequency increases. Our findings continue to hold when we use the standard deviation of prices across ETFs to capture price dispersion (Panel B, Table V).

5.2 ETF bond valuation and AP creation baskets

To examine how adjusted valuation differences across ETFs affect the composition of APs' creation baskets, we follow established methods in the literature to infer the composition of

[26] See, for example, Hotchkiss and Jostova (2017).

an ETF's creation basket based on changes in its holdings. Specifically, on days with net ETF inflows, an increase in a bond's position likely results from the creation of ETF shares. We then create a dummy variable at the fund-bond-day level to indicate whether a bond is part of the creation basket for a given ETF on a specific day. $BondCreation_{f,b,t}$ equals one if the ETF's position (par amount) in bond b increases on day t, and zero otherwise. As shown in Table VI, consistent with prior findings, we observe that on days when ETFs experience net inflows, the probability of a bond being included in the creation basket is 9.6%. In contrast, on days with net outflows, the probability of a bond being included in the redemption basket rises to 18.4%, suggesting that redemption baskets typically contain twice the number of bonds compared to creation baskets.[27]

We then construct a fund-bond-day level sample and estimate the following model:

$$BondCreation_{f,b,t} = \beta \times (PD_{f,b,t} - \overline{PD_{f,t}}) + \varepsilon_{f,b,t} \tag{3}$$

where $PD_{f,b,t}$ is defined as in model (1), and $\overline{PD_{f,t}}$ is the average $\text{PD}_{f,b,t}$ estimated at the fund-day level and captures the ETF adjustment as illustrated in the economic mechanism in Section 3. The difference between $PD_{f,b,t}$ and $\overline{PD_{f,t}}$ indicates how a fund values a bond on a given day relative to its peers, after controlling for the fund's overall valuation tendency. Standard errors are double-clustered at the date and fund levels.

As hypothesized in Section 3, conditional on ETF creation, a bond that receives a higher relative valuation by a given ETF, after adjusting for the ETF's overall valuation tendency, should be more likely to be included in that ETF's creation basket by APs. This hypothesis implies a positive coefficient β in model (3). The results in Table VII strongly support this hypothesis. Specifically, Column (1) shows that the coefficient for the adjusted PD measure

[27] We note that while our data affords only net flows on the fund level, creations and redemptions in fact can both occur in a given day. In our sample, ETFs frequently experience same-day bond positional changes in both directions, for which the only plausible reason is that there are simultaneous creations and redemptions in the day (by different APs with different baskets). Table VI shows a 1.7% (2.9%) probability that a bond being included in the redemption (creation) basket on a net-creation (net-redemption) day, suggesting that ETFs likely receive creation and redemption requests on the same day.

is 6.437, both positive and highly significant, indicating that a higher adjusted valuation of a bond by an ETF increases the probability of its inclusion in the ETF's creation basket.

We next include fund-date fixed effects in our analysis, which are crucial for addressing concerns about unobservable, time-varying fund-level characteristics and shocks that could influence the number and the size of create baskets received on a given day. For instance, on trading days with excessive investor demand in the secondary market and a positive ETF premium, there will be many sizable AP creation orders and various APs could submit different bonds for creation baskets, which may significantly increase the probability of creation basket inclusion for all bonds in the ETF portfolio. By incorporating fund-date fixed effects, we can account for these potential biases along with other fund level characteristics such as fund age, total net asset, secondary market liquidity, and the number of APs associated with an ETF. Moreover, adding fund-date fixed effects allows us to control for the impact of a fund's valuation on a particular bond from the fund's overall valuation tendency. Thus, we re-estimate model (3) by replacing $(PD_{f,b,t} - \overline{PD_{f,t}})$ with $PD_{f,b,t}$, while incorporating fund-date fixed effects. This replacement is made because $\overline{PD_{f,t}}$ becomes redundant when fund-date fixed effects are added. As shown in Column (2), the coefficient of $PD_{f,b,t}$ is positive and highly significant, indicating that, after controlling for time varying fund characteristics (including the fund's overall valuation tendency), a fund is more likely to receive a bond in its creation basket when it assigns a higher value to that bond relative to its peers.

It is important to note that the economic mechanism through which bond valuation affects ETF creation decisions operates through cross-ETF valuation dispersion for a given bond on a given day. Therefore, bond-level characteristics, even if time-varying, should not attenuate the effect of $PD_{f,b,t}$ on $BondCreation$ in model (3). Column (3) confirms this prediction, showing that our results remain robust when we further include bond-date fixed effects. To further account for potential ETF-specific preferences, such as when a bond is a top constituent in an ETF's benchmark index, we add fund-bond fixed effects. As shown in Column (4), the coefficient on $PD_{f,b,t}$ is 7.915 and is statistically significant, suggesting that a one standard deviation increase in $PD_{f,b,t}$ (20.1 bps) is associated with a 1.6 percentage point increase in the likelihood that a bond is included in the creation basket, conditional on an AP initiating a creation.

To further assess the economic magnitude of the effect of ETFs' adjusted bond valuation on APs' creation basket decisions, we construct a rank-based measure derived from the PD measure. Specifically, for each bond b on day t, we assign a percentile rank to each ETF based on the price it assigns to bond b. In Column (5), we define a dummy variable equal to one if a bond's rank is above the 50th percentile and zero otherwise. We then re-estimate model (3) by replacing the continuous PD measure with this binary indicator. The coefficient on the dummy is 0.0672 and statistically significant, indicating that bonds priced above the median are 6.72 percentage points more likely to be included in the creation basket, conditional on an AP conducting a creation. This effect has economic significance, given that the baseline probability of inclusion in a creation basket is only 9.6%.

In Column (6), we create four dummy variables for ETFs falling into the following rank intervals: below the 20th percentile, 20th-40th percentile, 60th-80th percentile, and above the 80th percentile, whereas the 40-60th percentile range is the baseline group. We then re-estimate model (3) by replacing the PD measure with these four dummy variables. Column (6) of Table VII reveals a monotonically increasing relationship between the price deviation rank and the likelihood of a bond being included in a creation basket. The coefficients on the four dummies are -0.051, -0.023, 0.0418, and 0.0718, respectively, all statistically significant. These results indicate that, relative to the baseline group, bonds priced in the 60th–80th and above-80th percentile ranges are 4.18 and 7.18 percentage points more likely to be included in a creation basket, respectively. In contrast, bonds priced in the 20th-40th and below-20th percentile ranges are 2.3 and 5.1 percentage points less likely to be included.

To further investigate the relationship between adjusted valuation differences and inclusion in creation baskets, we conduct a fund-by-fund analysis and re-estimate model (3) for each ETF in our sample. Panel A, Table VIII reports the distribution of coefficients for price deviation across our sample of 90 ETFs. Of these, 81 ETFs exhibit a positive coefficient, indicating a positive association between the ETF price deviation and the likelihood of creation basket inclusion. Of the 81 ETFs, 63 show statistical significance at the 10% level, and 18 are statistically insignificant. Conversely, 9 ETFs have a negative coefficient, with, untabulated, only 3 (1) showing statistical significance at the 10% (1%) level and the remaining 6 being statistically insignificant. Moreover, the magnitude of the positive coefficient is substantially larger than that of the negative coefficient: the mean for the positive-coefficient ETFs is 5.36, while the mean for the negative-coefficient ETFs is only -0.83. These results suggest that the relationship between ETFs' adjusted bond valuation differences and the inclusion in creation baskets, as documented earlier, is evident across most ETFs.

Panel B reports the number of ETFs with positive and statistically significant coefficients, disaggregated by fund family. iShares, the largest corporate bond ETF issuer in our sample, manages 27 ETFs, all of which exhibit positive coefficients. Among these, 26 are statistically significant at the 10% level, and 25 are significant at the 1% level. State Street follows with 8 ETFs, 6 of which have positive coefficients, and 4 are statistically significant at the 1% level. Other major ETF issuers, including Invesco, Xtrackers, American Century, Goldman Sachs, and PIMCO, also show a high proportion of positive and significant coefficients. Overall, the results suggest that our main findings are robust and broadly exist across fund families.

5.3 Robustness checks and additional analyses

We perform robustness checks and supplementary analyses to better understand bond valuation dispersion across ETFs and to validate the reliability of our main findings.

5.3.1 Extended samples

Our main analysis focuses on the subsample of fund-days with net inflows, as these are the most relevant for creation activity. However, in practice, ETF creations can also occur on days with net outflows or no net flows. To test the robustness of our findings, we expand the sample to include these fund-days. We replicate the analyses from Table VII using a broader sample that includes i) all fund-days with non-zero flows (both inflows and outflows), or ii) all fund-days (the full sample), whether or not there are net flows. As shown in Panels A and B of Table A2 in the Internet Appendix, our main results remain robust.

5.3.2 Holdings-based prices and ETF creation

We infer each ETF's bond valuation by dividing the reported market value of its bond holdings by the corresponding par amount. A potential concern is that these inferred prices may not perfectly align with the actual prices used by ETFs to calculate NAVs and, by extension, to determine creation baskets. Although we expect ETFs to follow consistent pricing conventions across reporting and operational activities, we conduct two robustness tests to address this concern.

First, we re-estimate the PD measure by imposing uniform pricing conventions across all ETFs, either using bid or mid prices, and replicate our main empirical analysis (Table VII). Specifically, for each month, we infer whether an ETF tends to report its holdings using bid or mid prices by comparing its inferred prices with bid and ask quotes at daily closes in MarketAxess CP+ data. CP+ provides AI-driven price estimates from a leading electronic trading platform for corporate bonds.[28] Our analysis is restricted to the 2020–2023 period due to the availability of CP+ data. We then use the bid-ask spreads from CP+ to adjust reported ETF prices so that all ETFs are assumed to use either bid or mid pricing. As shown in Table A3 of the Internet Appendix, these adjustments have negligible impact on our results. Regardless of whether all ETFs are assumed to use bid prices, mid prices, or their own reporting convention, the PD coefficient remains positive, highly significant, and similar in magnitude.

Second, we conduct a validation exercise using LQD as a case study. We show that

[28]See Shin, Zhou, and Zhu (2025) for a detailed description of the CP+ data.

prices inferred from LQD's daily holdings closely match those used in its NAV calculation and in share creation. See Internet Appendix E for a detailed illustration of this comparison. Moreover, as shown in Table A4 in the Internet Appendix, our main results remain robust when restricted to iShares ETFs issued by Blackrock, the largest ETF issuer in our sample, further reinforcing the validity of our approach.

5.3.3 ETF bond valuation and redemption baskets

We focus on the strategic behavior of APs during the creation process, as they have the flexibility to negotiate customized creation baskets with ETF issuers. This flexibility, however, does not necessarily extend to the redemption process, where redemption baskets are largely determined by ETF issuers.[29] One interesting question is whether ETF issuers selectively include bonds that they assign higher values in redemption baskets, since by doing that, ETF issuers can deliver fewer bonds (or less cash) in exchange for a given number of ETF shares.

To test this, we re-estimate model (3) using a new dependent variable, $BondRedemption_{f,b,t}$, a dummy variable that equals one if the number of shares of bond b held by ETF f decreases on day t, and zero otherwise. While the coefficient on PD remains positive and statistically significant, its economic magnitude is substantially smaller. Specifically, the coefficient on $PD_{f,b,t}$ is 0.890, compared to 7.915 in the main results where $BondCreation_{f,b,t}$ is used as the dependent variable (see Column (1) of Table A5 and Column (4) of Table VII). Column (2) of Table A5 further shows that the implied economic effect ranges from –0.178% to 0.063% for various PD percentiles, which is negligible given that the unconditional probability of a bond being included in a redemption basket is 18.4% during net negative flow days. The overall effect of the PD measure on $BondRedemption_{f,b,t}$ weakens further when we expand the sample to include positive flow days (Columns (3) and (4)) and largely disappears in

[29]For further details, see Dannhauser and Hoseinzade (2022) and Koont et al. (2025), as well as the commentary released by BlackRock, the largest provider of corporate bond ETFs (`https://www.blackrock.com/corporate/literature/publication/sec-proposed-rule-exchange-traded-funds-092618.pdf`).

the full sample covering all fund-bond-day observations (Columns (5) and (6)). This muted relationship is consistent with the notion that selectively including higher PD bonds in redemptions may distort the portfolio's composition and introduce tracking error, making such a strategy suboptimal from the perspective of ETF issuers.

5.3.4 Alternative approach to capture adjusted price deviation

As illustrated in Section 5.2, the construction of our main independent variable ($PD_{f,b,t} - \overline{PD_{f,t}}$) standardizes bond prices reported by multiple ETFs in two steps: we subtract first the cross-sectional median price of bond b on day t across all ETFs with $PD_{f,b,t} = P_{f,b,t} - \widetilde{P_{b,t}}$, and then subtract the average PD for fund f on day t ($\overline{PD_{f,t}}$) across all bonds in the ETF. The order of subtraction should not affect the estimation results, as the inclusion of both fund $\times$ date and bond $\times$ date fixed effects effectively demeans $P_{f,b,t}$ across both dimensions. Specifically, any fund-day level adjustment (i.e., subtracting $\overline{PD_{f,t}}$) is absorbed by fund $\times$ date fixed effects, while any bond-day level adjustment (i.e., subtracting $\widetilde{P_{b,t}}$) is absorbed by bond $\times$ date fixed effects.

In Table A6, we demonstrate that the order of these transformations is inconsequential by re-estimating model (3) using the raw bond price $P_{f,b,t}$. Because the model includes both sets of fixed effects, raw bond prices are effectively demeaned across all ETFs holding the same bond on the same day, and across all bonds within the same fund on the same day. In Columns (1), (3), and (5), we confirm that the results are robust to this specification. The coefficient on raw bond price is smaller in magnitude due to the presence of outliers in $P_{f,b,t}$. Directly winsorizing $P_{f,b,t}$ is problematic, as it curates bonds with extreme prices rather than addressing large within-bond price deviations. To better capture such outliers, we instead winsorize the price deviation. Specifically, we compute $\text{PD}_{f,b,t} = P_{f,b,t} - \widetilde{P_{b,t}}$, winsorize $PD_{f,b,t}$ at 1% and 99%, and finally reconstruct *Bond Price (winsorized)* as $\widetilde{P_{b,t}} + PD_{f,b,t}$(winsorized). In Columns (2), (4), and (6), we re-estimate model (3) using *Bond Price (winsorized)*. The resulting coefficients match exactly those reported in Table VII and Table A2, further confirming the robustness of our approach.[30]

5.3.5 The impact of bond valuation on ETF creation over time

Both the corporate bond ETF market and the underlying bond market have experienced significant evolution over our sample period. Changes in pricing conventions by index providers have reshaped trading dynamics and liquidity conditions in the corporate bond market (Shin et al. (2025)). In parallel, innovations such as portfolio trading, which more tightly link the ETF and bond markets, have grown substantially over time (Li, O'Hara, Rapp, and Zhou (2025)).

A natural question is whether the impact of bond valuation on ETF creation has changed over time. To explore this, we divide our sample by year and re-estimate the main specification of model (3) separately for each year. As shown in Table A7, while the strength of the relationship varies across years, the effect remains statistically significant throughout the sample period. This suggests that our main results are not driven by transitory market conditions or time-specific structural changes.

6 Sources of Cross-ETF Price Dispersion

What drives valuation dispersions across ETFs? One key source of variation stems from ETF issuers, who may adopt different pricing conventions when reporting the value of their bond holdings. For instance, some issuers consistently use bid prices, while others rely on mid prices.[31] They may also differ in their choice of pricing sources, such as third-party vendors, dealer quotes, or proprietary valuation models. These issuer-level differences

[30] In our main analysis, we adjust for each ETF's overall valuation tendency by computing $\overline{PD_{f,t}}$, the simple average of $PD_{f,b,t}$ across all bonds held by fund f on day t. Our main results will remain unchanged if $\overline{PD_{f,t}}$ is instead calculated as a value-weighted average, as long as both fund × date and bond × date fixed effects are included in the regression; these fixed effects absorb any fund-level or bond-level variation, rendering the choice between simple and weighted averages inconsequential to the estimation.

[31] See, for example, Cici et al. (2011).

contribute to price dispersion across ETFs. Another source of variation is the index provider. Each corporate bond ETF in our sample tracks a specific index, and the index provider typically supplies indicative or end-of-day pricing for the bonds in that index. These prices are commonly used to calculate ETF NAVs and to determine the value of ETF creation and redemption baskets. Since index providers also vary in their data sources and pricing methodologies, even ETFs from the same issuer may report different bond values due to index-level differences to help minimize tracking errors relative to the benchmark index. Finally, cross-ETF valuation dispersion may arise from idiosyncratic factors. ETF issuers may exercise discretion adjusting bond values for specific funds due to market conditions or fund-specific considerations.

6.1 Decomposition of price deviation: The roles of ETF issuers and bond index providers

To better understand the contributions of these sources to observed valuation dispersion, we re-estimate cross-ETF price dispersion while controlling for ETF issuer and index provider identities. We begin by re-estimating the dispersion of the fund-bond-date level PD for each bond on each day using ETFs from the same issuer. Panel A of Table IX shows that ETF issuer-level differences account for a substantial portion of the observed dispersion: the average interquartile range declines from 19.16 bps to 16.40 bps, while the average standard deviation drops from 16.01 bps to 11.34 bps.[32] Panel B shows that index providers play an even larger role. When we re-estimate PD dispersion using ETFs following the same index provider, the average falls more significantly to 11.16 bps using the interquartile range and to 11.18 bps using the standard deviation. Lastly, when we jointly control for both ETF issuer and index provider, restricting the estimation of PD dispersion to ETFs from the same issuer and using the same index provider, the average price dispersion falls below 1 bps for

[32]See Table IV for the full-sample values.

both measures (Panel C). This indicates that issuer- and index-level differences account for the vast majority of cross-ETF valuation dispersion.[33] Nevertheless, in rare cases we still observe sizable dispersion, suggesting that discretionary pricing decisions by ETF issuers may contribute to specific instances.

6.2 Decomposition of Price Deviation: A Case Study of LQD

We use the iShares iBoxx $ Investment Grade Corporate Bond ETF (ticker: LQD) as a case study to further investigate how ETF issuers and index providers contribute to valuation dispersion across ETFs. LQD is the most actively traded corporate bond ETF by volume and offers broad exposure to liquid, USD-denominated investment-grade corporate bonds. Specifically, we decompose each bond's price deviation reported by LQD from the cross-ETF median into two components:

$$PD_{\mathrm{LQD},b,t} = \underbrace{(P_{\mathrm{LQD},b,t} - P_{\mathrm{iBoxx},b,t})}_{\text{Active component}} + \underbrace{\left(P_{\mathrm{iBoxx},b,t} - \widetilde{P_{b,t}}\right)}_{\text{Passive component}}, \tag{4}$$

where $P_{\mathrm{iBoxx},b,t}$ refers to the price reported by the tracking index provider (iBoxx) for bond b on day t, and $\widetilde{P_{b,t}}$ is the median price reported by all ETFs. The first term reflects how an ETF issuer adjusts the price reported by the tracking index provider (i.e., an active component), while the second term captures how the index provider's valuation of a bond deviates from the median of reported bond prices across all ETFs (i.e., a passive component). If the ETF issuer strictly follows the bond valuations provided by the index provider, the observed price deviations of a bond in the ETF would be entirely driven by the passive component, i.e., the index provider's reported price bias relative to other ETFs.[34]

[33] This is consistent with the example in Table I, where all six iShares ETFs tracking Bloomberg indices value the bond at 67.90.

[34] As a robustness check, we also construct the passive component using the median price reported by all ETFs other than LQD. The results are virtually unchanged, indicating that our decomposition is not sensitive to whether the ETF under study is included in the benchmark median.

We obtain daily constituent and pricing information for the iBoxx Liquid Investment Grade (USD) Index (hereafter, iBoxx IG Index)—the benchmark index tracked by LQD—from Markit Indices GmbH, a subsidiary of S&P Dow Jones Indices LLC.[35] For each bond on each day, Markit provides both bid and ask prices for all bonds included in the index. We then combine the daily security-level holdings data for LQD with the daily constituent data for the iBoxx IG index. On average, LQD holds approximately 2,000-3,000 bonds and closely tracks the iBoxx index, with the weight difference having a mean of just 0.06 bps and a standard deviation of 0.95 bps. LQD's portfolio follows a barbell strategy, with the majority of its bond holdings maturing in the 1-10 year and 20-30 year ranges. Although all bonds in LQD's portfolio are investment-grade, more than half are rated BBB, the lowest rating within the investment-grade category (see Panel A of Table X).

LQD, on average, values bonds slightly lower than its peers, with the mean (median) price deviation being -9 (-4) bps (Panel B of Table X). This pattern is consistent with LQD's use of bid prices to mark bond valuations, whereas some other ETFs may use mid prices. However, there are considerable variations in its price deviation. The standard deviation of LQD's price deviation is 33 bps, which is greater than the full-sample price dispersion (see Table IV). A potential explanation for this larger dispersion is that LQD tracks an index provided by Markit, whereas most of the other ETFs in our sample track indexes constructed by Bloomberg. Differences in pricing methodologies by index providers may contribute to the observed cross-sectional variation.

Indeed, most of LQD's valuation deviation from its peers is primarily driven by the passive component, which is estimated as the difference between iBoxx bid prices and the cross-ETF median prices.[36] This component has a mean of -9 bps and a standard deviation of 32 bps, which align closely with the overall price deviation. In 36% of cases, iBoxx bond

[35]We note that daily constituent and pricing information for other indices is not available to us, despite our best efforts.

[36]We focus on bonds that are both in the iBoxx IG Index and the LQD portfolio. There are on average 54 (2.4%) bond issues in the iBoxx IG index but not in the LQD portfolio, and on average 75 (3.4%) bond issues in the LQD portfolio but not in the iBoxx IG index.

prices are higher than the median reported prices across ETFs. In contrast, the active deviation component, defined as the difference between LQD's reported bond prices and the corresponding iBoxx bid prices, is generally close to zero, with a mean of just 0.5 bps, and it remains within a narrow range of 0 to 0.1 bps from the 1st to the 75th percentile. This finding also confirms that iShares, the issuer of LQD, relies on bid prices when reporting the market value of its bond holdings. Interestingly, when LQD adjusts the prices reported by iBoxx, it consistently increases them. There is a 10% (7%) chance that LQD's reported prices exceed those of iBoxx by more than 2 bps (3 bps). The relatively liquid nature of the bonds held by LQD may limit the scope for significant active price adjustments. Nevertheless, such deviations suggest discretionary pricing behavior by iShares, a possibility we explore further below.

We next examine the impact of the two components of price deviation on the likelihood of a bond's inclusion in LQD's creation baskets. As a large and liquid ETF, LQD experiences frequent creations and redemptions, with flows occurring on 95.6% of the days in our sample period. Panel C of Table X shows that, on net-creation days, an average of 25.4% of the bonds are included in the creation baskets, while on net-redemption days, an average of 40.5% are in the redemption baskets. These figures are notably higher than the overall sample averages, reflecting LQD's greater liquidity due to its frequent creation and redemption activities.

We re-estimate our primary results (model (3) of Table VII) for LQD. Since our analysis now focuses on a single ETF, we use date fixed effects to capture any time-varying valuation trends of LQD relative to other ETFs, and bond fixed effects to control for unobserved heterogeneity across bonds. Additionally, we account for the potential time-varying impacts of credit risk and interest rate risk by including fixed effects for i) credit ratings (categorized into broad rating groups: AAA, AA, A, and BBB), and ii) time to maturity (categorized into three groups: short-term (1–10 years), medium-term (10–20 years), and long-term (20–30 years)). Column (1) of Table XI confirms that our main finding from the full sample of ETFs holds for LQD— specifically, LQD's *BondCreation* is positively related to its price deviation measure.

We then separately assess the impact of the distinct components of the price deviation measure on *BondCreation*. Specifically, we replace $PD_{\mathrm{LQD},b,t}$ with either component and re-estimate model (3). Columns (2) and (3) show that the coefficients for both components are positive and highly significant, indicating that LQD's bond deviation in either component increases the likelihood of the bond being included in its creation baskets. Our results remain consistent when both components are included in the regression (Column (4)). Notably, the active component appears to have much stronger explanatory power for *BondCreation* than the passive component, as evidenced by the adjusted R^2 increasing from 14.1% in Column (3) to 39.0% in Column (4).

We also create dummy variables for each of the two components to better assess their individual impact on *BondCreation*. Specifically, the *Passive Dummy* variable is set to one if the price reported by the iBoxx IG Index exceeds the median price reported by other ETFs, and zero otherwise. The *Active Dummy* variable is set to one if the price reported by LQD exceeds the iBoxx IG Index price by more than two basis points, and zero otherwise. We then replace the two components of the price deviation measure with these dummy variables and re-estimate the model. Column (5) shows that a positive passive price deviation is associated with a 3.98 percentage point increase in *BondCreation*. In contrast, an active price deviation has a much larger effect, with a deviation greater than 2 basis points linked to a 70 percentage point increase in *BondCreation*. This result suggests that while LQD does not often report prices substantially above the index, but when it does adjust pricing beyond a small threshold, the impact on the likelihood of the bond being included in its creation basket is significant.

An interesting question is what motivates ETF issuers to actively adjust their prices. One possible explanation is their incentive to minimize tracking error. Since ETFs typically track indexes, bonds that help reduce tracking error are more valuable to them. To test this hypothesis, we compare the portfolio weight difference between the ETF and the tracked index and examine its relationship with the probability of a bond being included in creation baskets. A bond is more likely to be included in the creation basket if the ETF underweights it relative to the index, as the ETF will need to acquire more of that bond to correct the discrepancy. The opposite holds when the ETF overweights a bond.

Specifically, we estimate the following model for the active component of price deviation:

$$P_{\text{LQD},b,t} - P_{\text{iBoxx},b,t} = \beta \times (w_{\text{iBoxx},b,t-1} - w_{\text{LQD},b,t-1}) + \mu_b + \mu_t + \mu_r + \mu_m + \varepsilon_{\text{LQD},b,t} \quad (5)$$

where $w_{\text{LQD},b,t-1}$ ($w_{\text{iBoxx},b,t-1}$) is the weight of bond b in LQD (iBoxx) on day $t-1$. The model is estimated with bond fixed effects (μ_b), day fixed effects (μ_t), rating fixed effects (μ_r), and maturity fixed effects (μ_m).

The results in Table XII support our hypothesis. Column (1) shows that the coefficient for the weight difference between iBoxx and LQD from the previous day is positively related to the active component of price deviation. This suggests that LQD tends to increase the price of a bond when it is underweighted relative to the iBoxx IG Index, potentially to encourage a custom creation basket to tilt towards these underweighted bonds.

iBoxx IG Index exclusively includes bonds with at least three years expected remaining life.[37] Anticipating a bond's imminent removal from the index, the LQD manager could proactively underweight it. As a result, even if these bonds are underweighted in the LQD portfolio, a positive price deviation is unlikely to occur. We re-estimate model (5) for bond subsamples defined by a 3.5-year maturity cutoff, which serves as a buffer to account for rebalancing timing under the index's $\geq$ 3-year inclusion rule. Columns (2) and (3) show that the positive association between bond underweighting and the next day active component of the price deviation is not significant for bonds with a remaining time to maturity of less than 3.5 years. In contrast, the negative relationship is even stronger for bonds with more than 3.5 years to mature.

[37]See the factsheet of iBoxx® USD Liquid Investment Grade Index: `https://cdn.ihs.com/www/pdf/MKT-iBoxx-USD-Liquid-Investment-Grade-Index-factsheet.pdf`.

To assess the economic significance of the above, we create a dummy variable, *Underweight Dummy*, that takes the value of one if $w_{\text{LQD},b,t-1}$ is less than $w_{\text{iBoxx},b,t-1}$, and zero otherwise. We then re-estimate model (5), replacing $P_{\text{LQD},b,t} - P_{\text{iBoxx},b,t}$ with *Active Dummy* (the dummy of active price deviation $>$ 2 bps), and $w_{\text{LQD},b,t-1} - w_{\text{iBoxx},b,t-1}$ with the *Underweight Dummy.* Column (4) shows that when a bond is underweighted by LQD relative to iBoxx, the likelihood of LQD adjusting the bond's price upward by 2 basis points the next day increases by 2.11 percentage points. This represents 20% of the unconditional probability that the active price deviation exceeds 2 basis points. Columns (5) and (6) confirm differential effects based on bond maturity. For bonds with at least 3.5 years of remaining life, Column (6) shows that the probability of an active upward price adjustment $>$ 2 bps increases by 2.54 percentage points. Overall, these findings support the hypothesis that an ETF is more likely to increase the price of a bond when it is underweighted in its portfolio.

7 Impact of Valuation Dispersions on Bond ETFs

Thus far, we have outlined a novel mechanism by which APs benefit from bond valuation dispersion across ETFs. Specifically, we have shown that APs can gain by strategically delivering bonds to ETFs that value them most highly. A natural question arises: How does this strategic behavior by APs impact the ETF industry, if at all? In this section, we conduct a series of tests to examine the impact of cross-ETF valuation dispersion on AP profitability, investor wealth, and ETF pricing efficiency.

7.1 AP profitability and investor wealth

A typical fixed income ETF holds thousands of corporate bonds. For instance, LQD typically holds between 2,000 and 3,000 bonds, while AGG holds more than 10,000. In practice, a fractional creation basket can include over 50 bonds and sometimes several hundred bonds. How much can an AP benefit by selecting bonds with a higher PD measure, compared to selecting bonds based on criteria unrelated to the PD measure? Quantifying this benefit would help assess the economic significance of the observed behavior and shed light on the incentives driving APs' decisions in the primary market.

Importantly, APs' profits from such strategic selection in fractional baskets represent a direct wealth transfer from existing ETF investors. When an ETF accepts a creation basket disproportionately concentrated in bonds it values more highly, the value of outstanding ETF shares is diluted; in other words, incumbent investors receive securities that are overpriced relative to the rest of the portfolio. Thus, our estimates of AP gains from strategic selection can be equivalently interpreted as dilution costs borne by ETF investors.

A back-of-the-envelope calculation: To gauge potential gains from strategic bond selection, we draw on the estimates from Column (5) of Table VII, which indicate that a bond priced above the median valuation across ETFs is 6.72 percentage points more likely to be included in a creation basket. In the absence of our documented APs' strategic behavior, one would expect bonds priced above and below the median to be included with equal probability of 9.6%, the baseline inclusion probability on net-inflow days. Our findings imply that the inclusion probabilities are 12.96% and 6.24% for bonds with valuations above and below the cross-sectional median, respectively. This translates to an average creation basket composition of 67.5% above-median bonds and 32.5% below-median bonds, rather than a balanced 50-50 split.

To assess the economic value of this skewed selection, we use the ETF-adjusted price deviation dispersion (*PD Dispersion*) reported in Table IV. The average interquartile range is 19.16 basis points, implying an average gain or loss of 9.58 basis points depending on whether a bond is priced above or below the median. Applying this to the observed 67.5-32.5 composition above, the average gain from strategic selection is approximately 3.35 basis points relative to a neutral 50-50 selection strategy.

This gain is economically meaningful for APs. For instance, Table III shows that the half interquartile range of ETF premiums or discounts, the main driver of ETF arbitrage, is about 11.6 basis points. If an AP creates 100,000 ETF shares with a NAV of \$100 each, the arbitrage profit from exploiting secondary market mispricing would be roughly \$11,600. By strategically selecting higher-PD bonds, the AP can earn an additional \$3,353, representing a 29% increase in overall profitability.

Estimating actual gains realized through creation baskets: To further assess the impact of APs' strategic bond selection on their profitability and investor wealth, we analyze changes in Morningstar ETF holdings alongside cross-sectional differences in ETF bond valuations. Our analysis focuses on fund-bond-day observations where the ETF experiences non-zero flows and the change in bond holdings $\Delta Shares_{f,b,t}$ is positive. For a given ETF f on day t, we first calculate the fund's overall valuation tendency, $\overline{PD_{f,t}}$, defined as the average price deviation $PD_{f,b,t}$ across all bonds held by the fund. We then compute the ETF-adjusted bond price as

$$P^{Adj}_{f,b,t} = P_{f,b,t} - \overline{PD_{f,t}}. \tag{6}$$

Next, we benchmark this adjusted bond price against the median ETF-adjusted price across all ETFs that hold bond b on day t, denoted $\widetilde{P^{Adj}_{b,t}}$. The dollar profit ($P\&L^{C,\$}_{f,b,t}$) for the AP (equivalently, the loss to ETF investors) from including bond b in ETF f's creation basket is then calculated as:

$$P\&L^{C,\$}_{f,b,t} = \Delta Shares^{+}_{f,b,t} \times (P^{Adj}_{f,b,t} - \widetilde{P^{Adj}_{b,t}}), \tag{7}$$

where $\Delta Shares^{+}_{f,b,t} = \max(\Delta Shares_{f,b,t}, 0)$. It is important to note that we do not take a position on whether the bond is misvalued by the aggregate market, nor do we infer whether inclusion in a creation basket implies lower future bond returns. Instead, we focus on relative valuation by comparing each ETF's pricing to the median pricing among peer ETFs holding the same bond. This measure captures both gains and losses. If an AP delivers a bond that the ETF prices below the peer median, then $P^{Adj}_{f,b,t} - \widetilde{P^{Adj}_{b,t}}$ would be negative, leading to negative values of $P\&L^{C,\$}_{f,b,t}$, or losses for APs.

To estimate the percentage profit to APs for conducting a creation basket at fund-day level, we aggregate $P\&L^{C,\$}_{f,b,t}$ across all bonds, normalized by the size of the creation basket measured as $\sum_b \left(\Delta Shares^{+}_{f,b,t} \times P_{f,b,t}\right)$:

$$\overline{P\&L^{C,\%}_{f,t}} = \frac{\sum_b P\&L^{C,\$}_{f,b,t}}{\sum_b \left(\Delta Shares^{+}_{f,b,t} \times P_{f,b,t}\right)}. \tag{8}$$

Panel A of Table XIII reports the average $\overline{P\&L^{C,\%}_{f,t}}$ across creation baskets, value-weighted at the fund-day level by the size of the creation basket. Column (1) presents the summary statistics for the full sample of 90 ETFs. The average gain for an AP from conducting a creation basket is 1.6 bps. Columns (2) and (3) focus on subsamples of ETFs with positive and positive and statistically significant β estimates, as reported in Table VIII. Column (4) further restricts the sample to ETFs with β estimates that are statistically significant at the 1% level and have values greater than 7, corresponding to the 75th percentile among positive β estimates. This final sample consists of 20 ETFs, which tend to be larger funds, with a median total net assets (TNA) over \$1 billion, compared to the overall sample median of \$409 million. In this restricted subsample, we find that the average (median) gain for APs is approximately 3.2 (3.2) bps , with the 75th percentile at 6.1 bps and the 90th percentile at 10.5 bps. Panel B of Table XIII reports the equal-weighted summary statistics, which are consistent with the value-weighted results. Overall, the estimated average gains for APs range from 1.6 to 3.2 bps, aligning with the back-of-the-envelope calculations. These gains come at the expense of existing ETF investors, representing a material dilution of their holdings.

7.2 Implications for ETF pricing efficiency

As discussed earlier, an ETF is more likely to acquire a bond through the creation process when it values the bond more highly than its peers, while still maintaining a relatively conservative overall valuation of its portfolio. This means that, relative to its peers, the

ETF prices some bonds lower and others higher, leading to substantial dispersion in the PD measure across bonds within its holdings. Put differently, the greater the cross-bond dispersion in an ETF's PD measures, the more likely it is to attract creations from APs. By contrast, because bond valuation dispersion does not materially influence the composition of redemption baskets (see Section 5.3.3), cross-bond dispersion in an ETF's PD measures has little effect on its redemptions. Consistent with this prediction, our data (Table A8) show that an ETF's probability of receiving creations, but not redemptions, is positively and significantly related to its cross-bond PD dispersion.

Given the asymmetric impact of an ETF's cross-bond PD dispersion on the primary market, an important question arises: how does this asymmetry affect ETF pricing? Higher cross-bond PD dispersion may disproportionately incentivize APs to engage in creations, thereby increasing the ETF's share supply and exerting downward pressure on its price. In other words, greater cross-bond PD dispersion should be associated with a lower ETF premium, or equivalently, a deeper ETF discount.

To empirically evaluate this conjecture, we estimate the following model:

$$\textit{ETF Premium}_{f,t} = \gamma \times \textit{Fund PD Dispersion}_{f,t} + \mu_f + \mu_t + \varepsilon_{f,t}, \tag{9}$$

where $\textit{Fund PD Dispersion}_{f,t}$ is measured as the interquartile range of the PD measure across all bonds held by fund f on day t. A higher value of this measure reflects greater dispersion in bond valuations within the fund. We note that this measure is not affected by whether the ETF tends to mark up or down bond prices overall, since we focus on the cross-sectional dispersion (Q75–Q25) rather than absolute pricing levels. The regression includes fund fixed effects (μ_f) and date fixed effects (μ_t) to control for unobservable heterogeneities at the fund and day levels, respectively. Standard errors are double-clustered at both the fund and date levels.

Results in Table XIV support our conjecture. Column (1) shows that the coefficient on *Fund PD Dispersion*$_{f,t}$ is negative and statistically significant, indicating that greater dispersion in bond valuations within an ETF is associated with a lower ETF premium. Our results change little when we further control for liquidity effects by incorporating time-varying fund-level variables, specifically ETF size and ETF bid–ask spreads (column (2)). The findings also hold when we compute *Fund PD Dispersion*$_{f,t}$ using value weights rather than equal weights within each fund (see columns (3) and (4)).

Depending on the ETF's pre-existing demand-supply imbalance and pricing condition, APs' strategic behavior can either enhance or impair price efficiency. When the ETF trades at a premium and experiences strong investor demand, the incentive encourages creations, increasing share supply and helping to realign the market price with NAV. When the ETF trades at NAV, however, APs may still find it profitable to create shares, not because of price-NAV differences but due to strategic bond selection. Moreover, when the ETF is trading at a discount in the presence of selling pressure, a sufficiently strong strategic bond selection incentive may still lead APs to create shares despite the negative arbitrage signal, thereby exacerbating pricing inefficiencies.

We use quantile regressions to examine how the effect of *Fund PD Dispersion*$_{f,t}$ on *ETF Premium*$_{f,t}$ varies across different points of the distribution. This approach allows us to test whether the impact of *Fund PD Dispersion*$_{f,t}$ is stronger in the discount regime (lower quantiles) or in the premium regime (upper quantiles). Figure 2 plots point estimates and 95% confidence intervals from quantile regressions of *ETF Premium*$_{f,t}$ on *Fund PD Dispersion*$_{f,t}$, controlling for fund-level characteristics and fixed effects as in Table XIV. Panel (a) shows that *Fund PD Dispersion*$_{f,t}$ is generally associated with a smaller ETF premium (or a higher discount) across the distribution. Moreover, the negative coefficients become progressively smaller in magnitude across quantiles, indicating that the relationship is more pronounced in the lower part of the distribution. Specifically, the coefficient on *Fund PD Dispersion*$_{f,t}$ is negative and statistically significant at the 10th, 25th, and 50th percentiles, with estimates of -0.23, -0.19, and -0.14, respectively. The economic magnitude is substantial.

For example, at the 10th percentile, where an ETF is on average trading at a discount of 15 bps, an interquartile increase in *Fund PD Dispersion*$_{f,t}$ (17 bps) is associated with an additional 3.8 bps increase in the ETF discount. In contrast, at the 75th and 90th percentiles, where an ETF is on average trading at a premium of 21 bps and 39 bps, respectively, the coefficients on *Fund PD Dispersion*$_{f,t}$ become smaller in magnitude (-0.09 and -0.05, respectively) and statistically insignificant. These findings suggest that the negative effect of *Fund PD Dispersion*$_{f,t}$ is concentrated among ETFs trading at or near a discount and weakens as the ETF premium increases. Panel (b) shows that our results remain largely unchanged when using valued weighted *Fund PD Dispersion*$_{f,t}$. Taken together, these results suggest that bond valuation uncertainty can exacerbate pricing inefficiencies, particularly during periods of market stress.

8 Conclusion

In this paper, we analyze the strategic behavior of APs in response to cross-ETF valuation discrepancies in the primary corporate bond ETF market. Our findings highlight the important role that valuation dispersion plays in shaping APs' decisions about which ETF to select when converting bond inventories into ETF shares. APs strategically exploit these discrepancies by delivering bonds to the ETFs that assign them the highest value, thereby maximizing their potential benefits from the creation process.

We show that the profitability of this strategy depends not only on the relative valuation assigned to a bond by an ETF but also on the ETF's overall tendency to value bonds higher or lower than its peers. Our empirical analysis supports this mechanism, demonstrating that bonds are more likely to be included in the creation basket of an ETF that assigns them the highest value, especially when the ETF exhibits a relatively lower overall valuation tendency compared to others. The robustness of our results across various specifications, including controlling for potential biases from market price deviations or bond-specific shocks, further strengthens our conclusions.

Our findings carry important implications for both ETF flows and pricing. We find that ETFs are more likely to receive creations on a given day when the dispersion in their valuations across bonds is greater. For ETF issuers, this finding highlights the importance of managing valuation discrepancies across holdings, as such discrepancies influence not only creation activity but also the fund's attractiveness to APs. At the same time, ETFs with high valuation dispersion may inadvertently encourage strategic behavior by APs, which can lead to unintended consequences for ETF price efficiency. Indeed, when ETFs trade at a discount, typically during periods of selling pressure, the incentives created by bond price dispersion still promote share creations, causing prices to deviate further from fundamentals and exacerbating pricing inefficiencies. While the corporate bond ETF market plays a critical role in enhancing liquidity in the corporate bond market, the strategic behavior of APs, driven by valuation dispersion across ETFs, can have significant implications for both ETF and market outcomes. Our findings underscore the importance of understanding these dynamics for both ETF issuers and investors, as they navigate the complexities of the bond and ETF markets.

References

Andrews, S., Henderson, B. J., Reed, A. V., 2024. Cross-etf arbitrage. Available at SSRN 4667997.

Ben-David, I., Franzoni, F., Kim, B., Moussawi, R., 2023. Competition for attention in the ETF space. Review of Financial Studies 36, 987–1042.

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

Bhattacharya, A., O'Hara, M., 2018. Can ETFs increase market fragility? Effect of information linkages in ETF markets. Available at SSRN 2740699.

Box, T., Davis, R., Evans, R., Lynch, A., 2021. Intraday arbitrage between ETFs and their underlying portfolios. Journal of Financial Economics 141, 1078–1095.

Boyarchenko, N., Kovner, A., Shachar, O., 2022. It's what you say and what you buy: A holistic evaluation of the corporate credit facilities. Journal of Financial Economics 144, 695–731.

Brogaard, J., Heath, D., Huang, D., 2025. ETF sampling and index arbitrage. Journal of Financial and Quantitative Analysis, Forthcoming.

Choi, J., Kronlund, M., Oh, J., 2025. Silent swing: Do open-end funds tilt holdings valuations in response to flows? Available at SSRN 5048497.

Cici, G., Gibson, S., Merrick Jr, J. J., 2011. Missing the marks? Dispersion in corporate bond valuations across mutual funds. Journal of Financial Economics 101, 206–226.

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

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

Dannhauser, C. D., Hoseinzade, S., 2022. The unintended consequences of corporate bond ETFs: Evidence from the taper tantrum. Review of Financial Studies 35, 51–90.

Dannhauser, C. D., Karmaziene, E., 2023. The dealer warehouse–corporate bond ETFs. Available at SSRN 4676850.

Du, L., 2023. Inventory in liquidity transformation: Evidence from corporate bond ETF creations. Available at SSRN 4653799.

Eisfeldt, A. L., Herskovic, B., Liu, S., 2024. Interdealer price dispersion and intermediary capacity. Available at SSRN 4238166.

Evans, R. B., Karakaş, O., Moussawi, R., Young, M., 2025. Phantom of the opera: Etf shorting and shareholder voting. Management Science Forthcoming.

Evans, R. B., Moussawi, R., Pagano, M. S., Sedunov, J., 2022. Operational shorting and etf liquidity provision. Tech. Rep. 2961954, Darden School of Business, University of Virginia Villanova School of Business, working Paper; last revised August 4, 2022.

Feldhütter, P., 2012. The same bond at different prices: Identifying search frictions and selling pressures. Review of Financial Studies 25, 1155–1206.

Finnerty, J. D., Reisel, N., Zhong, X., 2025. ETFs, creation and redemption processes, and bond liquidity. Journal of Financial and Quantitative Analysis 60, 1891–1924.

Glosten, L., Nallareddy, S., Zou, Y., 2021. ETF activity and informational efficiency of underlying securities. Management Science 67, 22–47.

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

Haddad, V., Moreira, A., Muir, T., 2021. When selling becomes viral: Disruptions in debt markets in the COVID-19 crisis and the Fed's response. Review of Financial Studies 34, 5309–5351.

Holden, C. W., Nam, J., 2024. Market accessibility, corporate bond ETFs, and liquidity. Review of Finance pp. 18–1.

Hotchkiss, E., Jostova, G., 2017. Determinants of corporate bond trading: A comprehensive analysis. Quarterly Journal of Finance 7, 1750003.

Hu, G. X., Pan, J., Wang, J., 2013. Noise as information for illiquidity. Journal of Finance 68, 2341–2382.

Jankowitsch, R., Nashikkar, A., Subrahmanyam, M. G., 2011. Price dispersion in OTC markets: A new measure of liquidity. Journal of Banking & Finance 35, 343–357.

Kargar, M., Lester, B., Lindsay, D., Liu, S., Weill, P.-O., Zúñiga, D., 2021. Corporate bond liquidity during the COVID-19 crisis. Review of Financial Studies 34, 5352–5401.

Koont, N., Ma, Y., Pastor, L., Zeng, Y., 2025. Steering a ship in illiquid waters: Active management of passive funds. Review of Financial Studies Forthcoming.

Li, J. S., O'Hara, M., Rapp, A. C., Zhou, X. A., 2025. Bond market illiquidity: Is portfolio trading the solution? Available at SSRN 4495516.

Ma, Y., Xiao, K., Zeng, Y., 2022. Mutual fund liquidity transformation and reverse flight to liquidity. Review of Financial Studies 35, 4674–4711.

O'Hara, M., Zhou, X. A., 2021. Anatomy of a liquidity crisis: Corporate bonds in the COVID-19 crisis. Journal of Financial Economics 142, 46–68.

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

Reilly, C., 2022. The hidden cost of corporate bond ETFs. Available at SSRN 4147882.

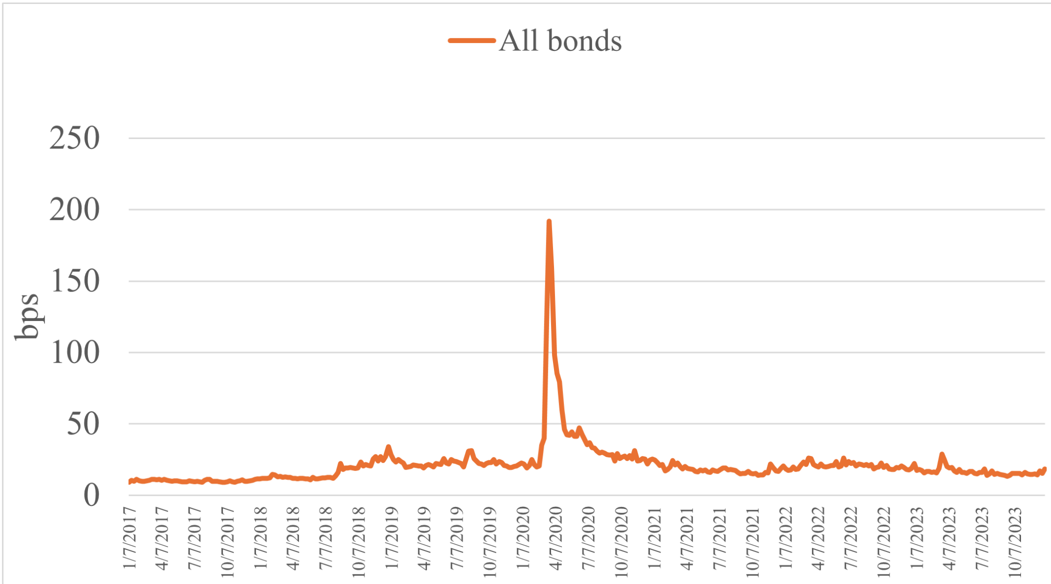
Sağlam, M., Tuzun, T., Wermers, R., 2025. Do ETFs increase liquidity? Available at SSRN 3142081.

Shim, J. J., Todorov, K., 2021. ETFs, illiquid assets, and fire sales. BIS Working Papers, No 975.

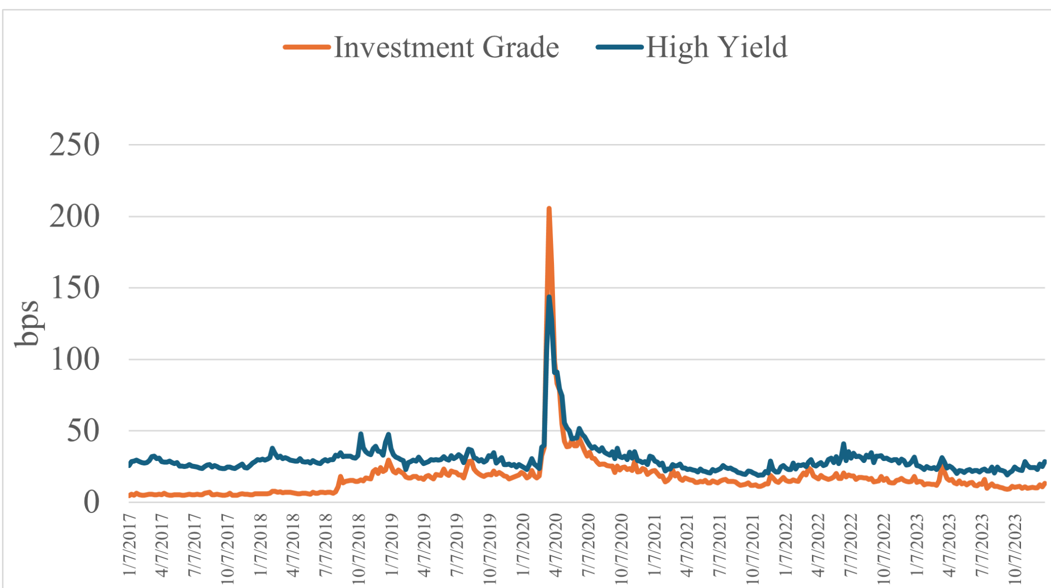
Shim, J. J., Todorov, K., 2023. ETFs, illiquid assets, and fire sales. Available at SSRN 3886881.

Shin, S. S., Zhou, X. A., Zhu, Q., 2025. The growing index effect in the corporate bond market. Available at SSRN 5181327.

Sialm, C., Xu, D. X., 2025. Information acquisition by mutual fund investors: Evidence from stock trading suspensions. Available at SSRN 5047094.

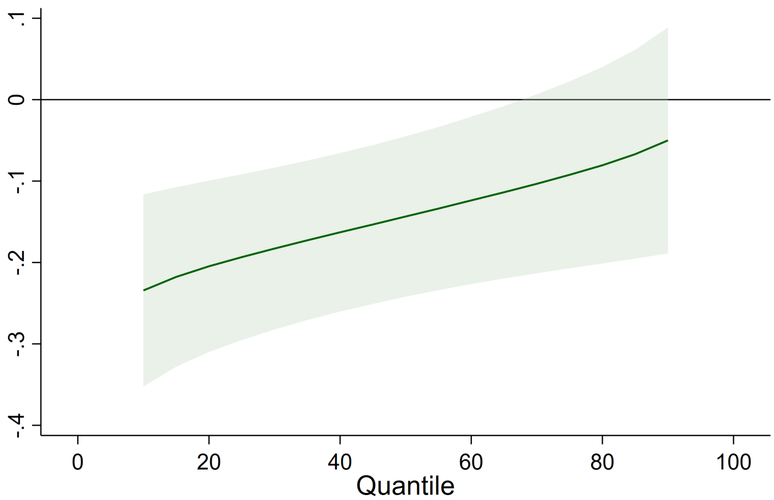


(a) Weekly average of $PD\ Dispersion_{b,t}$

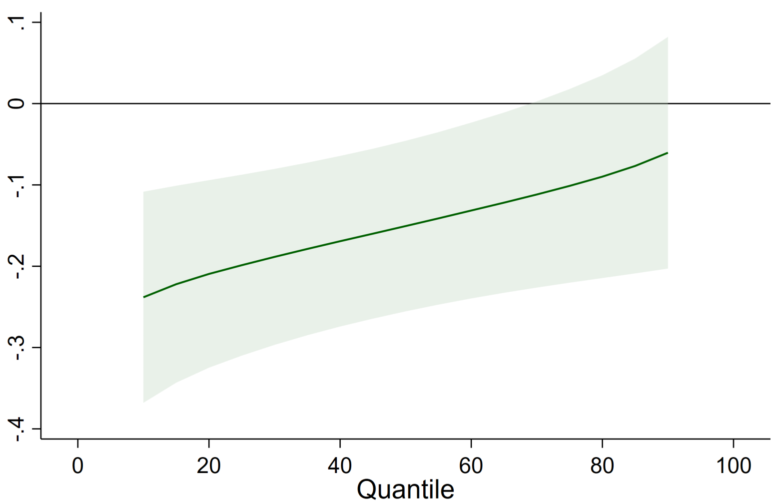


(b) Weekly average of $PD\ Dispersion_{b,t}$ for IG and HY bonds

Figure 1 Price dispersion over time This figure displays the weekly time series of price dispersion over our sample period. We first calculate the interquartile range (IQR) of the price deviations at the bond-day level, using the price deviations $PD_{f,b,t}$ by all ETFs for bond b on day t. We then compute the weekly average of these IQRs and plot the weekly time series in Panel (a). Panel (b) shows the time series for investment-grade bonds and high-yield bonds, respectively.



(a) *Fund PD Dispersion*$_{f,t}$ is equal weighted



(b) *Fund PD Dispersion*$_{f,t}$ is value weighted

Figure 2 Quantile regressions of ETF premium on *Fund PD Dispersion*$_{f,t}$ This figure examines how the relationship between ETF premiums and bond price dispersion varies across the distribution of ETF premiums. It plots the estimated coefficients from quantile regressions of $ETF\,Premium_{f,t}$ on *Fund PD Dispersion*$_{f,t}$. The dependent variable, $ETF\,Premium_{f,t}$, is defined as (Price/NAV) − 1. The independent variable, *Fund PD Dispersion*$_{f,t}$, measures the interquartile range of bond-level price deviations $PD_{f,b,t}$ within each ETF on day t. Panel (a) uses the equal-weighted dispersion measure, while Panel (b) uses the value-weighted measure. Each line traces the estimated coefficient across conditional quantiles of the dependent variable (from 0 to 100), and the shaded area represents the 95% confidence interval. The regressions include controls for ETF size and bid–ask spread, and include fund fixed effects and date fixed effects. Standard errors are clustered by fund. $ETF\,Premium_{f,t}$ and *Fund PD Dispersion*$_{f,t}$ are expressed in basis points.

Table I: An Example of a Bond Held by Multiple ETFs

This table shows the reported dirty prices, as of 09/28/2022, by corporate-bond ETF holders of a bond issued by Oracle Corp. (CUSIP: 68389XBJ3) carrying a Baa2 rating and maturing 07/15/2046. The data is from Morningstar. "Fund inflow" is indicated with a "✓" when the number of ETF shares outstanding increased on 09/28/2022. "Bond in Creation Basket" is marked with a "✓" when the ETF shares outstanding changed and the ETF's par holdings of the bond increased.

ETF Name	Ticker	Index Benchmark	Bond Price (Dirty)	Fund Inflow	Bond in Creation Basket
iShares iBoxx $ Invmnt Grade Corp Bd	LQD	Markit iBoxx Liquid IG TR USD	68.52	✓	✓
iShares 10+ Year Invtmt Grd Corp Bd	IGLB	ICE BofA 10+Y US Corp TR USD	68.33		
iShares Broad USD Invm Grd Corp Bd	USIG	ICE BofA US Corporate TR USD	68.33	✓	
Hartford Total Return Bond	HTRB	Bbg US Agg Bond TR USD	68.32		
Schwab US Aggregate Bond	SCHZ	Bbg US Agg Bond TR USD	68.13		
WisdomTree Yield Enhanced US Aggt Bd	AGGY	Bbg US Agg Enhanced Yield TR USD	68.13	✓	
SPDR Portfolio Long Term Corp Bd	SPLB	Bbg US Long Corporate TR USD	67.91		
SPDR Portfolio Corporate Bond	SPBO	Bbg US Corp Bond TR USD	67.91		
SPDR Portfolio Aggregate Bond	SPAB	Bbg US Agg Bond TR USD	67.91	✓	
iShares Government/Credit Bond	GBF	Bbg US Govt/Credit TR USD	67.90		
iShares ESG U.S. Aggregate Bond	EAGG	Bbg MSCI US Agg ESG Focus TR USD	67.90		
iShares Core US Aggregate Bond	AGG	Bbg US Agg Bond TR USD	67.90	✓	
iShares Core Total USD Bond Market	IUSB	Bbg US Universal TR USD	67.90	✓	
iShares ESG USD Corporate Bond	SUSC	Bbg MSCI US Corp ESG Focus TR USD	67.90		
iShares Core 10+ Year USD Bond	ILTB	Bbg US Universal 10+ Years TR USD	67.90		

Table II: Dispersions in Bond Valuations and Custom Creation Baskets

This table illustrates how dispersions in bond valuations across ETFs influence APs' selection of bonds in custom creation baskets. In this example, four ETFs each hold one unit of Bond 1 and one unit of Bond 2, but assign different valuations to these bonds. ETF NAVs are calculated based on their reported prices. For each ETF, we calculate the cash component required to exchange two units of a bond for one ETF share. This component depends on the bond's reported price and the ETF's NAV, with lower cash components being more attractive to APs. In the case of exchanging Bond 2, negative cash components imply cash paid out by the ETF; more negative values are more favorable to APs. We calculate the price deviation as the ETF-reported bond price minus the cross-ETF median price for that bond-day. This price deviation is then adjusted by the average price deviation for Bonds 1 and 2 within each ETF. We then show that the adjusted price deviation is negatively associated with the cash component required to create one ETF share.

	ETF A	**ETF B**	**ETF C**	**ETF D**
	Reported constituent price			
Bond 1	102	105	105	105
Bond 2	110	110	113	107
ETF NAV	212	215	218	212
	Exchanging Two Units of Bond 1 for one ETF Share			
Bond 1 Value (2 units)	204	210	210	210
Cash component	8	5	8	2
Basket NAV	212	215	218	212
	Exchanging Two Units of Bond 2 for one ETF share			
Bond 2 Value (2 units)	220	220	226	214
Cash component	-8	-5	-8	-2
Basket NAV	212	215	218	212
	Price Deviation (PD): Bond price − Median Bond price			
Bond 1	-3.0	0.0	0.0	0.0
Bond 2	0.0	0.0	3.0	-3.0
	ETF-Specific Adjustment: Average (Bond price − Median bond price)			
ETF Adjustment	-1.5	0.0	1.5	-1.5
	Price Deviation (PD) − ETF-Specific Adjustment			
Bond 1	-1.5	0.0	-1.5	1.5
Bond 2	1.5	0.0	1.5	-1.5

Table III: Summary Statistics of ETFs and Bond Holdings
This table reports the summary statistics of the ETF characteristics and the constituent bonds. Panel A reports the summary statistics at the fund-day level. Panel B reports the summary statistics at the bond-day level. A bond's credit rating at the time of trading has the following numerical value: AAA=1, AA=2, A=3, BBB=4, BB=5, B=6, CCC=7, CC=8, and C and below = 9. See the Internet Appendix for more variable definitions.

	Mean	Median	S.D.	P1	P10	Q1	Q3	P90	P99
Panel A: Fund-day level (N = 121,486)									
ETF Premium (bps)	10	8	24	-69	-15	-2	21	39	91
Fund Size ($1,000,000)	3,239	395	8,521	10	54	119	2,192	7,359	57,830
Bid-Ask Spread (bps)	11	7	12	1	2	4	15	25	69
Panel B: Bond-day level (N = 13,656,655)									
Credit Rating	3.79	4.00	0.98	2	3	3	4	5	7
Remaining maturity (years)	10.37	6.59	9.22	0.9	2.0	3.6	16.6	26.3	37.6
Age (years)	4.23	3.21	3.81	0.1	0.6	1.5	5.9	8.7	19.4
Offering amount ($1,000)	894	700	750	250	350	500	1,000	1,750	3,500

Table IV: Summary Statistics of Bond Price Dispersion at the Bond-Day Level

This table presents the summary statistics of bond price dispersion, $PD\ Dispersion_{b,t}$, at the bond-day level, which is measured as the interquartile range (IQR) or the standard deviation (STD) of price deviation across funds for bond b on day t. Price deviation ($PD_{f,b,t}$) is defined as $P_{f,b,t} - \widetilde{P_{b,t}}$, where $P_{f,b,t}$ is the price for bond b reported by ETF f on day t, and $\widetilde{P_{b,t}}$ is the median price of bond b reported by all ETFs that hold the bond on day t. The number of ETF holders is the number of ETFs that hold the same bond on the same day. The number of TRACE transactions (t) is the total number of transactions for the bond reported in TRACE on that day. There are 13,656,655 observations at the bond-day level.

	Mean	Median	S.D.	P1	P10	Q1	Q3	P90	P99
Number of ETF holders (same bond, same day)	10.46	10.00	4.59	3	4	7	14	17	21
Number of TRACE transactions (t)	5.36	2.00	10.3	0	0	0	7	14	41
Number of TRACE transactions (t) > 0	0.65	1.00	0.50	0	0	0	1	1	1
$PD\ Dispersion_{b,\,t}$ (IQR, bps)	19.16	8.00	36.94	0.02	0.79	1.63	22.68	49.84	139.24
$PD\ Dispersion_{b,\,t}$ (STD, bps)	16.01	10.32	21.93	0.23	1.82	4.78	20.55	35.22	85.12

Table V: Determinants of Bond Price Dispersion

This table presents the relationship between cross-ETF bond price dispersion and bond liquidity. The dependent variable (expressed in basis points) is the interquartile range (IQR) of price deviations for the same bond on the same day across different ETFs in Panel A, and the standard deviation (STD) of these prices in Panel B. Independent variables include liquidity proxies, such as the number of TRACE transactions, and various bond characteristics. A bond's credit rating at the time of trading has the following numerical value: AAA=1, AA=2, A=3, BBB=4, BB=5, B=6, CCC=7, CC=8, and C and below = 9. The t-statistics, shown in parentheses, are adjusted for two-way clustering at the date and bond levels. *p< .1; **p< .05; ***p< .01.

Panel A: Dependent variable $PD\ Dispersion_{b,t}$ **(IQR)**

	(1)	(2)	(3)	(4)
Number of TRACE transactions(t) > 0	-2.7428***	-3.0014***		
	(-15.58)	(-17.23)		
log(1+Number of TRACE transactions(t))			-1.1090***	-1.2230***
			(-10.55)	(-12.56)
Remaining years to maturity	0.6412***	0.6312***	0.6321***	0.6206***
	(37.72)	(37.52)	(38.03)	(37.46)
Years since offering	0.4354***	0.4921***	0.4388***	0.4972***
	(13.11)	(14.01)	(13.21)	(14.11)
log(Offering amount)	-2.4151***	-2.4636***	-2.1803***	-2.2037***
	(-10.54)	(-10.71)	(-9.48)	(-9.51)
Credit Rating	2.9121***	1.4021***	2.9245***	1.4407***
	(12.95)	(3.59)	(13.00)	(3.68)
Observations	13,656,655	13,656,655	13,656,655	13,656,655
Issuer FE	Y		Y	
Date FE	Y		Y	
Issuer × Date FE		Y		Y
Adj R-squared	0.260	0.335	0.260	0.335

Table V: (continued)

Panel B: Dependent variable $PD\ Dispersion_{b,t}$ **(STD)**	(1)	(2)	(3)	(4)
Number of TRACE transactions(t) > 0	-1.8837***	-2.2060***		
	(-17.20)	(-21.11)		
log(1+Number of TRACE transactions(t))			-0.9569***	-1.1511***
			(-14.33)	(-18.69)
Remaining years to maturity	0.6377***	0.6338***	0.6269***	0.6201***
	(54.82)	(55.02)	(55.66)	(55.36)
Years since offering	0.4525***	0.4773***	0.4513***	0.4762***
	(23.42)	(23.36)	(23.56)	(23.47)
log(Offering amount)	-1.1205***	-1.1857***	-0.8451***	-0.8463***
	(-8.74)	(-9.21)	(-6.78)	(-6.73)
Credit Rating	1.4464***	0.6017***	1.4657***	0.6716***
	(11.77)	(2.96)	(11.98)	(3.33)
Observations	13,653,522	13,652,271	13,653,522	13,652,271
Issuer FE	Y		Y	
Date FE	Y		Y	
Issuer × Date FE		Y		Y
Adj R-squared	0.396	0.478	0.396	0.478

Table VI: Summary Statistics for Bond Creations and Redemptions

This table presents the likelihood of a bond being included in the creation or redemption baskets on ETF trading days. ETFs' creation and redemption baskets are identified based on daily changes in ETF holdings in Morningstar and the daily flows from Bloomberg. The variable $BondCreation_{f,b,t}$ ($BondRedemption_{f,b,t}$) is a fund-bond-day level dummy that equals one if the number of shares of bond b held by ETF f increases (decreases) on day t, and zero otherwise. Both positive and negative net flow days may involve creations and redemptions, though a positive (negative) net flow day ensures that the ETF has at least one creation (redemption). We report the mean values of $BondCreation_{f,b,t}$ and $BondRedemption_{f,b,t}$, along with the number of fund-bond-day observations for days with positive, negative and non-zero flows, and all days.

	Positive flow days (Net Creation)	Negative flow days (Net Redemption)	Non-zero flow days (Both)	All Days
	$BondCreation$: Creation = 1; Otherwise=0			
Mean	9.6%	2.9%	7.6%	3.3%
	$BondRedemption$: Redemption = 1; Otherwise=0			
Mean	1.7%	18.4%	6.7%	2.9%
	Number of Observations (Fund-Bond-Day)			
N	42,508,900	19,098,321	62,236,177	157,143,077

Table VII: Price Deviation and ETF Creation Baskets

This table explores the relationship between cross-ETF price dispersion and a bond's likelihood of inclusion in the creation basket. The analysis is conducted at the fund–bond–day level using observations from ETF–days with positive net flows, corresponding to days with net creations. The dependent variable is $BondCreation$, a dummy variable equal to one if the number of bonds held by the ETF increases on the given day, and zero otherwise. Price Deviation ($PD_{f,b,t}$) is defined at the fund-bond-day level as $P_{f,b,t} - \widetilde{P_{b,t}}$, where $P_{f,b,t}$ is the price for bond b reported by ETF f on day t, and $\widetilde{P_{b,t}}$ is the median price of bond b reported by all ETFs that hold the bond on day t. $PD_{f,b,t}$ is measured in dollars per \$1 par to facilitate clearer economic interpretation of the coefficients (see Appendix A). $\overline{PD_{f,t}}$ represents the average $PD_{f,b,t}$ estimated at the fund-day level and captures the ETF adjustment. The replacement of $(PD_{f,b,t} - \overline{PD_{f,t}})$ with $PD_{f,b,t}$ in Columns (2) to (4) is made because $\overline{PD_{f,t}}$ becomes redundant when fund-date fixed effects are added. Columns (5) and (6) utilize rank-based measures derived from the PD measure as independent variables. For bond b on day t, we assign a percentile rank to each ETF based on the price it assigns to bond b. We then create four dummy variables for ETFs falling within the following rank ranges: 0%-20%, 20%-40%, 60%-80%, and above 80%. The t-statistics, shown in parentheses, are adjusted for two-way clustering at the date and fund levels. *p< .1; **p< .05; ***p< .01.

	(1)	(2)	(3)	(4)	(5)	(6)
	$BondCreation$ (Creation =1; Otherwise=0)					
$PD_{f,b,t} - \overline{PD_{f,t}}$	6.437***					
	(5.57)					
$PD_{f,b,t}$		5.942***	7.224***	7.915***		
		(5.63)	(8.98)	(9.61)		
Rank($PD_{f,b,t}$) in (50, 100]					0.0672***	
					(9.79)	
Rank($PD_{f,b,t}$)≤20						-0.0510***
						(-11.24)
Rank($PD_{f,b,t}$) in (20, 40]						-0.0230***
						(-4.49)
Rank($PD_{f,b,t}$) in (60, 80]						0.0418***
						(6.41)
Rank($PD_{f,b,t}$)>80						0.0718***
						(6.11)
Observations	42,508,900	42,508,900	42,508,900	42,508,900	42,508,900	42,508,900
Fund × date FE		Y	Y	Y	Y	Y
Bond × date FE			Y	Y	Y	Y
Fund × bond FE				Y	Y	Y
Adj R-squared	0.002	0.176	0.214	0.233	0.242	0.245

Table VIII: Price Deviation and ETF Creation Baskets: Fund-by-Fund Analysis
This table investigates the relationship between cross-ETF price dispersion and a bond's likelihood of inclusion in the creation basket. For each ETF f, we estimate a separate bond-day level panel regression: $BondCreation_{f,b,t} = \beta_f \times PD_{f,b,t} + \mu_{f,t} + \mu_{f,b} + \epsilon_{f,b,t}$. The analysis is conducted for each fund using bond–day observations from days with positive net inflows, corresponding to periods of net creation activity. The dependent variable, *BondCreation*, is a dummy variable equal to one if the number of bonds held by ETF f increases on day t, and zero otherwise. The independent variable is the price deviation measure, $PD_{f,b,t}$, at the fund-bond-day level. Each regression includes date fixed effects and bond fixed effects. Standard errors are two-way clustering adjusted at the date and bond levels. Panel A reports the distribution of coefficients for price deviation across a sample of 90 ETFs and subsamples conditioning on the sign and statistical significance of β_f. Panel B reports the number of ETFs with positive and statistically significant β_f, disaggregated by fund family.

Panel A: Distribution of β_f for price deviation across a sample of 90 ETFs

		$\beta > 0$			$\beta < 0$
	All ETFs	All	p< 10%	p> 10%	All
N	90	81	63	18	9
Mean	4.74	5.36	6.63	0.91	-0.83
S.D.	6.32	6.37	6.68	0.94	0.65
Percentiles:					
P10	-0.08	0.25	0.65	0.03	-1.86
P25	0.55	1.12	1.85	0.22	-1.50
P50	2.53	3.14	4.55	0.44	-0.69
P75	6.27	6.90	7.99	1.52	-0.28
P90	14.33	14.53	16.63	2.53	-0.19

Table VIII: (continued)

Panel B: Number of ETFs disaggregated by fund family

ETF issuer	All ETFs	$\beta > 0$ All	$\beta > 0$ p< 10%	$\beta > 0$ p< 1%
iShares	27	27	26	25
State Street	8	6	4	4
Invesco	4	4	3	3
Fidelity	4	4	1	
JPMorgan	4	4	1	
American Century	3	3	3	3
Xtrackers	3	3	3	3
Goldman Sachs	3	3	3	2
PIMCO	3	3	2	2
Hartford Funds	3	3	2	1
Nuveen	3	3	2	1
Charles Schwab	3	2	1	1
Dimensional	2	2	2	2
First Trust	2	2	2	2
WisdomTree	2	2	2	1
Flexshares	2	2	1	
ALPS	2	1	1	
PGIM	2	1	1	
Franklin Templeton	2	1		
BlackRock	1	1	1	1
VanEck	1	1	1	1
Columbia	1	1	1	
Regents Park	1	1		
Vident Financial	1	1		
Inspire	1			
LeaderShares	1			
Pacer	1			
Total	90	81	63	52

Table IX: Decomposition of Price Deviation

This table reports the summary statistics of bond price dispersion across ETFs at the bond-day level (N=13,656,655), focusing on ETFs that share the same issuer, the same index provider, or both. For each bond b on each day t, price dispersion is measured as the interquartile range (IQR) or the standard deviation (STD) of the price deviations $PD_{f,b,t}$ across funds within the relevant group. Price deviation ($PD_{f,b,t}$) is measured as $P_{f,b,t} - \widetilde{P_{b,t}}$, where $P_{f,b,t}$ is the price for bond b reported by ETF f on day t, and $\widetilde{P_{b,t}}$ is the median price of bond b reported by all ETFs that hold the bond on day t.

	Mean	Median	S.D.	P1	P10	Q1	Q3	P90	P99
Panel A: Within issuer PD Variation (bps):									
$PD\,Dispersion_{b,t}$ (IQR)	16.40	5.29	36.72	0.00	0.03	0.54	18.24	45.60	138.51
$PD\,Dispersion_{b,t}$ (STD)	11.34	5.07	21.56	0.00	0.21	1.61	13.29	29.94	83.88
Panel B: Within index provider PD Variation (bps):									
$PD\,Dispersion_{b,t}$ (IQR)	11.16	2.07	23.25	0.01	0.56	0.98	11.52	30.70	96.02
$PD\,Dispersion_{b,t}$ (STD)	11.18	6.15	16.50	0.01	0.59	2.09	13.83	28.13	68.21
Panel C: Within index provider and issuer PD Variation (bps):									
$PD\,Dispersion_{b,t}$ (IQR)	0.65	0.03	6.49	0.00	0.01	0.02	0.09	0.28	11.09
$PD\,Dispersion_{b,t}$ (STD)	0.74	0.04	5.77	0.00	0.01	0.01	0.09	0.37	19.44

Table X: Summary Statistics of LQD and LQD's Bond Holdings

This table reports the summary statistics for the characteristics and the constituent bonds of LQD and its tracking index, the iBoxx IG index. Panel A provides the bond-day level distribution of constituent weights in the LQD portfolio and the iBoxx index, as well as the maturities and credit ratings of the bonds. A bond's credit rating at the time of trading has the following numerical value: AAA=1, AA=2, A=3, BBB=4, BB=5, B=6, CCC=7, CC=8, and C and below = 9. Panel B reports the bond-day level distribution of the overall price deviation, $PD_{\text{LQD},b,t} = P_{\text{LQD},b,t} - \widetilde{P_{b,t}}$, where $P_{\text{LQD},b,t}$ is the price of bond b reported by LQD on day t and $\widetilde{P_{b,t}}$ is the median price of the bond b reported by all ETFs on day t. We then decompose the overall price deviation into the active component and the passive component using the index price $P_{\text{iBoxx},b,t}$.Panel C shows the likelihood of a bond being included in the creation or redemption baskets on ETF trading days with non-zero flows.

Panel A: Portfolio weights and bond characteristics at the bond–day level (N=1,292,933)

	Mean	Median	S.D.	P1	P10	P25	P75	P90	P99
$w_{\text{LQD},b,t-1}$ (bps)	4.32	3.57	2.61	0.56	1.94	2.60	5.29	7.85	14.68
$w_{\text{iBoxx},b,t-1}$ (bps)	4.38	3.49	2.53	1.90	2.22	2.67	5.26	7.83	14.80
$w_{\text{iBoxx},b,t-1} - w_{\text{LQD},b,t-1}$ (bps)	0.06	-0.05	0.95	-2.50	-0.93	-0.41	0.46	1.21	3.30
Time to Maturity Dummies:									
$<$ 10 Year	0.58	1	0.49	0	0	0	1	1	1
10–20 Year	0.13	0	0.34	0	0	0	0	1	1
$>$ 20 Year	0.26	0	0.44	0	0	0	1	1	1
Credit Rating	3.40	4	0.70	1	3	3	4	4	4

Table X: (continued)

Panel B: Overall price deviation and decomposition at the bond–day level (N=1,292,933)

	Mean	Median	S.D.	P1	P10	P25	P75	P90	P99
Overall price deviation (PD)									
PD: $P_{\text{LQD},b,t} - \widetilde{P_{b,t}}$(bps)	-9	-4	33	-153	-42	-17	5	19	89
Decomposition using data from iBoxx									
Active Deviation:									
PD: $P_{\text{LQD},b,t} - P_{\text{iBoxx},b,t}$(bps)	1	0	1	0	0	0	0	2	6
Dummy: $P_{\text{LQD},b,t} - P_{\text{iBoxx},b,t} > 2$bps	0.10	0	0.30	0	0	0	0	1	1
Passive Deviation:									
PD: $P_{\text{iBoxx},b,t} - \widetilde{P_{b,t}}$(bps)	-9	-4	32	-145	-42	-18	4	18	89
Dummy: $P_{\text{iBoxx},b,t} > \widetilde{P_{b,t}}$	0.36	0	0.48	0	0	0	1	1	1

Panel C: *BondCreation* and *BondRedemption* Dummies

	Positive-flow days (Net Creation)	Negative-flow days (Net Redemption)	Non-zero Flow days (Both)	All Days
	BondCreation: Creation= 1; Otherwise = 0			
Mean	25.4%	6.3%	18.2%	17.5%
	BondRedemption: Redemption= 1; Otherwise = 0			
Mean	5.5%	40.5%	18.8%	18.0%
	Number of Observations (Fund–Bond–Day)			
N	1,292,933	791,711	2,084,644	2,217,271

Table XI: Price Deviation and LQD Creation Baskets
This table examines the relationship between cross-ETF price dispersion and a bond's likelihood of inclusion in LQD's creation basket. The analysis focuses on bond–day observations from days with positive net inflows for LQD, corresponding to periods of net creation activity. The dependent variable is $BondCreation_{LQD,b,t}$, a dummy variable equal to one if the number of bond b held by the LQD increases on day t, and zero otherwise. The independent variables include the overall price deviation $PD_{\text{LQD},b,t} = P_{\text{LQD},b,t} - \widetilde{P_{b,t}}$, where $P_{\text{LQD},b,t}$ is the price of bond b reported by LQD on day t and $\widetilde{P_{b,t}}$ is the median price of the bond b reported by all ETFs on day t. Additionally, we decompose this price deviation into two components: an active component by the ETF manager, defined as $P_{\text{LQD},b,t} - P_{\text{iBoxx},b,t}$, and a passive component, defined as $P_{\text{iBoxx},b,t} - \widetilde{P_{b,t}}$. We focus on the sample where the LQD flow is positive. PD is measured in dollars per \$1 par $PD_{f,b,t}$ to facilitate clearer economic interpretation of the coefficients (see Appendix A). The t-statistics, shown in parentheses, are adjusted for two-way clustering at the date and bond levels. *p< .1; **p< .05; ***p< .01.

	(1)	(2)	(3)	(4)	(5)
Dep. Var.	$BondCreation_{LQD,b,t}$ (Creation=1; Otherwise=0)				
Price Deviation ($PD_{\text{LQD},b,t}$)	8.7958***				
	(18.17)				
Active Deviation ($P_{\text{LQD},b,t} - P_{\text{iBoxx},b,t}$)		2,049.3***		2,042.6***	
		(36.22)		(36.26)	
Passive Deviation ($P_{\text{iBoxx},b,t} - \widetilde{P_{b,t}}$)			8.5680***	5.5031***	
			(17.99)	(11.50)	
Active Dummy ($P_{\text{LQD},b,t} - P_{\text{iBoxx},b,t} > 2bps$)					0.7013***
					(60.05)
Passive Dummy ($P_{\text{iBoxx},b,t} > \widetilde{P_{b,t}}$)					0.0398***
					(17.91)
Observations	1,292,933	1,292,933	1,292,933	1,292,933	1,292,933
Bond FE	Y	Y	Y	Y	Y
Date FE	Y	Y	Y	Y	Y
Rating FE	Y	Y	Y	Y	Y
Maturity group FE	Y	Y	Y	Y	Y
Adj R-squared	0.142	0.389	0.141	0.390	0.329

Table XII: Active Price Deviation and Portfolio Weights in LQD
This table investigates whether bonds that are underweighted in the LQD portfolio are associated with active price deviation reported by the LQD manager relative to the iBoxx index price. The dependent variables are active price deviation: $P_{\text{LQD},b,t} - P_{\text{iBoxx},b,t}$, and a dummy variable of whether this variable is greater than 2 bps. The independent variables include the difference between the bond's portfolio weight in LQD and its weight in the iBoxx IG Index, and a dummy variable (*Underweight Dummy*) equal to one if the bond is underweighted by LQD relative to the iBoxx IG Index on a given day and zero otherwise. We focus on the sample where the LQD flow is non-zero and at least three ETFs hold the bond on the given day. To facilitate clearer economic interpretation of the coefficients, $P_{\text{LQD},b,t} - P_{\text{iBoxx},b,t}$ is measured in dollars per \$1 par (see Appendix A), and $w_{\text{iBoxx},b,t-1} - w_{\text{LQD},b,t-1}$ is measured in decimal units, where 0.001 corresponds to 0.1%. The t-statistics, shown in parentheses, are adjusted for two-way clustering at the date and bond levels. *p< .1; **p< .05; ***p< .01

	(1)	(2)	(3)	(4)	(5)	(6)
Sample time to maturity (years)	All	<3.5	>3.5	All	<3.5	>3.5
	Active Deviation: $P_{\text{LQD},b,t} - P_{\text{iBoxx},b,t}$			Active Dummy: $P_{\text{LQD},b,t} - P_{\text{iBoxx},b,t} > 2$bps		
$w_{\text{iBoxx},b,t-1} - w_{\text{LQD},b,t-1}$	0.0286***	0.0095	0.0483***			
	(8.26)	(1.27)	(12.89)			
Underweight Dummy				0.0211***	0.0033	0.0254***
				(13.98)	(1.32)	(15.73)
Observations	2,135,349	122,351	2,012,989	2,151,851	123,220	2,028,622
Bond FE	Y	Y	Y	Y	Y	Y
Date FE	Y	Y	Y	Y	Y	Y
Rating FE	Y	Y	Y	Y	Y	Y
Maturity group FE	Y		Y	Y		Y
Adj R-squared	0.246	0.244	0.252	0.203	0.173	0.209

Table XIII: AP Profits from Actual Holdings

This table summarizes the profits (P&L) realized by APs from ETF creation baskets, based on changes in Morningstar holdings and cross-sectional differences in bond valuations across ETFs. For each bond included in a creation basket, we first compute the dollar profit as the difference between the ETF adjusted bond price and the cross-ETF median adjusted price. These fund-bond-level P&Ls are then aggregated at the fund-day level and expressed as a percentage of the creation basket's total value. Panel A reports value-weighted (VW) averages, where each fund-day observation is weighted by the dollar value of the creation basket. Panel B presents equal-weighted (EW) averages across fund-days.

	(1)	(2)	(3)	(4)
	All ETFs	$\beta > 0$	$\beta > 0$ & p< .1	$\beta > 7$ & p< .01
Number of ETFs	90	81	63	20
Average TNA ($1,000,000)	3,545	3,776	4,553	4,618
Median TNA ($1,000,000)	409	471	779	1,006
Panel A: VW average fund-day				
	APs' profit from creation baskets (bps)			
Mean	1.6	1.6	1.7	3.2
P10	-4.6	-4.7	-5.0	-3.4
P25	-1.2	-1.3	-1.3	0.5
Median	1.2	1.2	1.3	3.2
P75	4.3	4.4	4.5	6.1
P90	8.6	8.7	8.9	10.5
Panel B: EW average fund-day				
	APs' profit from creation baskets (bps)			
Mean	2.0	2.1	2.1	2.4
P10	-5.0	-5.1	-5.3	-3.7
P25	-1.0	-0.9	-1.0	0.0
Median	1.0	1.1	1.2	2.1
P75	4.2	4.4	4.6	5.6
P90	9.7	9.9	10.2	10.2

Table XIV: Bond Price Dispersion and ETF Pricing
This table examines the relationship between within-ETF bond price dispersion and ETF pricing. The dependent variable is $ETF\,Premium_{f,t}$ (Price/NAV $-$ 1). *Fund PD Dispersion*$_{f,t}$ is the interquartile range of $PD_{f,b,t}$ across bonds in ETF f on day t, computed either equal-weighted or value-weighted within fund. ETF premium, *Fund PD Dispersion*$_{f,t}$, and ETF bid–ask spread are in basis points. t-statistics (in parentheses) are two-way clustered by date and fund. *p< .1; **p< .05; ***p< .01.

	(1)	(2)	(3)	(4)
Dep. Var.:	ETF premium			
Fund PD Dispersion$_{f,t}$ (equal weighted)	-0.15**	-0.14***		
	(-2.63)	(-2.70)		
Fund PD Dispersion$_{f,t}$ (value weighted)			-0.16***	-0.15***
			(-2.65)	(-2.68)
log(Fund Size)		-2.92***		-2.90***
		(-4.26)		(-4.24)
ETF bid–ask spread		-0.03		-0.03
		(-0.41)		(-0.40)
Observations	121,486	121,486	121,486	121,486
Fund FE	Y	Y	Y	Y
Date FE	Y	Y	Y	Y
Adj R-squared	0.423	0.430	0.423	0.430

Internet Appendix for "Underlying Valuation Uncertainty, Strategic ETF Creation and Pricing Efficiency"

A Variable Definitions

Variable	Definition
Price Deviation ($PD_{f,b,t}$)	The ETF's reported price of the bond minus the median reported price of all ETFs ($PD_{f,b,t} = P_{f,b,t} - \widetilde{P_{b,t}}$), where $P_{f,b,t}$ is the price for bond b reported by ETF f on day t, and $\widetilde{P_{b,t}}$ is the median price of bond b reported by all ETFs that hold the bond on day t. For clearer economic interpretation, $PD_{f,b,t}$ is reported in basis points per \$1 par for summary statistics and in dollars per \$1 par for regressions. For example, if an ETF reports a bond price of 99.84 and the median price is 99.75, the price deviation equals 9 basis points when expressed in bps, and 0.0009 when expressed in dollars per \$1 par for regressions.
$\textit{PD Dispersion}_{b,t}$	A bond-day level variable measured by the interquartile range (IQR) or standard deviation (STD) for the price deviations, $PD_{f,b,t}$, for bond b reported by all ETFs on day t.
$\textit{Fund PD Dispersion}_{f,t}$	A fund-day level variable measured by the interquartile range (IQR) for the price deviations, $PD_{f,b,t}$, for all bonds in ETF f on day t.
$BondCreation_{f,b,t}$	A fund-bond-day level dummy that equals one if the number of shares of bond b held by ETF f increases on day t, and zero otherwise.
$BondRedemption_{f,b,t}$	A fund-bond-day level dummy that equals one if the number of shares of bond b held by ETF f decreases on day t, and zero otherwise.
ETF Premium	The ratio of the ETF price to the ETF's NAV from CRSP, minus one (Price/NAV − 1).
Fund Size	The total market cap of the ETF from CRSP.
Credit Rating	A bond's credit rating at the time of trading: AAA=1, AA=2, A=3, BBB=4, BB=5, B=6, CCC=7, CC=8, and C and below = 9.
Time to Maturity (years)	A bond's remaining maturity (in years) at the time of trading.
Age (years)	The number of years that elapsed between a bond's issuance date and the time of trade.
Issue Size	The principal amount of the bond issuance.
Number of ETF Holders	The number of ETFs that hold the same bond on the same day.
Number of TRACE transactions (t)	The total number of transactions for the bond reported in TRACE on day-t.
Number of TRACE transactions (t) > 0	A dummy variable equal to one if the Number of TRACE transactions (t) is greater than zero; otherwise, it is equal to zero.
$w_{\text{LQD},b,t-1}$	Weight of bond b in LQD's portfolio on day $t-1$.
$w_{\text{iBoxx},b,t-1}$	Weight of bond b in the iBoxx IG Index on day $t-1$.
$P_{\text{LQD},b,t} - \widetilde{P_{b,t}}$	The bond-day level overall price deviation by LQD, where $P_{\text{LQD},b,t}$ is the price of bond b reported by LQD on day t, and $\widetilde{P_{b,t}}$ is the median price of bond b reported by all ETFs on day t.
$P_{\text{LQD},b,t} - P_{\text{iBoxx},b,t}$	The difference in reported price between LQD and iBoxx IG Index.
$P_{\text{iBoxx},b,t} - \widetilde{P_{b,t}}$	iBoxx IG Index's reported price of the bond minus the median reported price of all ETFs.
$BondCreation_{\text{LQD},b,t}$	A dummy variable equal to one if the number of bond b held by the LQD increases on day t, and zero otherwise.

B ETF Sample Selection

We begin with a sample of 344 fixed income ETFs from Morningstar, spanning seven categories: U.S. Fund Corporate Bond, U.S. Fund High Yield Bond, U.S. Fund Long-Term Bond, U.S. Fund Intermediate Core Bond, U.S. Fund Intermediate Core-Plus Bond, U.S. Fund Multisector Bond, and U.S. Fund Short-Term Bond. We exclude 74 ETFs due to missing ticker symbols, non-USD base currency, or affiliation with Vanguard. Vanguard ETFs are removed because they typically do not report daily portfolio holdings. This yields a preliminary sample of 270 ETFs.

We then exclude 21 target date ETFs, resulting in a sample of 249. Next, we remove four ETFs lacking Morningstar holdings data and two ETFs (HYGH and BYLD) whose reported bond prices appear erroneous (e.g., being inflated by a factor of 10), reducing the sample to 243. To calculate fund flows, we use Bloomberg's daily shares outstanding data and exclude eight ETFs for which this information is unavailable, leaving 235 ETFs. We further exclude ETFs for which we cannot reliably infer creation baskets based on the correlation between flow and holdings data, as detailed in Appendix C. Specifically, we drop 64 ETFs with consistently low same-day and lead-lag correlations, resulting in a sample of 171 ETFs.

Finally, we exclude ETFs that are small, have limited corporate bond exposure, or exhibit minimal primary market activity. Specifically, we remove ETFs that (i) hold fewer than 50 unique corporate bonds over the full sample period, (ii) have fewer than 30 days with positive net inflows between 2017 and 2023 (approximately 1,764 trading days), or (iii) do not report daily holdings, even if not affiliated with Vanguard. Our final sample consists of 90 ETFs.

C Morningstar Holdings Timestamp: Before or After End-of-Day Creations/Redemptions?

To ensure consistency across data sources, we compare fund flow data from Bloomberg with holdings data from Morningstar, retaining only ETFs with clearly aligned reporting schedules. We adjust for observable lead-lag patterns when necessary, following the methodology of Koont et al. (2025), who document systematic reporting lags in Morningstar's holdings data relative to Bloomberg's shares outstanding data. Some ETFs report end-of-day holdings that incorporate the day's creation and redemption activity (see Panel (a) of Figure A1). For these ETFs, the inferred creation basket from day $t-1$ to t corresponds to the change in Bloomberg shares outstanding over the same interval. In contrast, other ETFs report holdings prior to incorporating that day's activity (Panel (b) of Figure A1). For these funds, the holdings-based basket from $t-1$ to t aligns with Bloomberg flows from $t-2$ to $t-1$.

To determine each ETF's reporting convention, we calculate daily flow estimates from Morningstar holdings data and correlate them with Bloomberg flows under both timing assumptions. For each ETF-year, we compute two correlation coefficients: (1) the same-day correlation between Morningstar and Bloomberg flows on day t, and (2) the lead-lag correlation between Morningstar flow on day t and Bloomberg flow on day $t-1$. Typically, one of these correlations is high (above 0.6), while the other is near zero. We manually inspect these patterns to classify each ETF's reporting convention and identify any changes in convention over time. For example, as shown in Table A1, the same-day correlations for LQD range from 0.6 to 0.8 between 2017 and 2023, while lead-lag correlations remain between 0.1 and 0.3, indicating that LQD includes same-day creations/redemptions in its reported holdings. In contrast, SPIB exhibits near-zero or negative same-day correlations (–0.1 to 0.1) and higher lead-lag correlations (0.4 to 0.8), suggesting it reports holdings prior to incorporating daily basket activity. For ETFs with stronger lead-lag correlations (e.g., SPIB), we shift the Morningstar date associated with creation and redemption baskets by one day to align it with Bloomberg flows.

This adjustment is applied only to the quantity, that is, changes in the par amount of holdings. The end-of-day valuation of bond holdings is not affected by the reporting convention. We verify this by examining the return correlations of identical bonds held by ETFs with differing reporting conventions. Aside from minor discrepancies attributable to price deviations discussed in the paper, the same-day correlations of bond returns across these ETFs are close to one, while the lead-lag correlations are close to zero. This indicates that valuation data is consistently reported as of the end of the Morningstar date stamp, regardless of the reporting convention.

For instance, Panel (a) of Figure A2 presents a scatter plot of the same-day daily price changes of a bond held by both LQD and SPIB in 2019. Bond price changes by LQD from $t-1$ to t are on the x-axis and bond price changes by SPIB over the same interval are on the y-axis. The points cluster tightly around the 45-degree line, reflecting a high degree of correlation.[1] In contrast, Panel (b) displays a lead-lag comparison, plotting price change by LQD from $t-1$ to t on the x-axis against price changes by SPIB from $t-2$ to $t-1$ on the y-axis. The absence of a clear pattern and the near-zero correlation confirm that daily price changes are aligned with a consistent Morningstar valuation timestamp, regardless of whether the ETF's Morningstar reporting convention reflects holdings before or after the primary market creation/redemption process, as illustrated in Figure A1. Therefore, ETF bond holding valuations from Morningstar do not require any adjustment to align with the Bloomberg flow timestamp.

[1] Two noticeable outliers occur when LQD reports price changes of less than –1.5%, while SPIB's corresponding price changes are near zero. These deviations arise from differences in dirty versus clean price reporting conventions.

D Pricing Reporting Convention (Dirty vs. Clean) in Morningstar

Bond pricing conventions vary across ETFs and fund families: some report clean prices, while others report dirty prices. Regardless of the reporting convention, ETF primary market bond transactions are based on dirty prices. To ensure comparability, we identify ETFs that report clean prices and adjust them by adding accrued interest.

Although prices for the same bond on the same day may differ across ETFs for various reasons, discrepancies arising from pricing conventions are readily identifiable. In particular, clean versus dirty pricing leads to a predictable cyclical pattern: prices drop on the coupon payment date and gradually rise thereafter as interest accrues. For example, Panel (c) of Figure A2 presents the time series of prices reported by LQD and SPIB in 2019 for a bond issued by UnitedHealth Group Incorporated. The bond carries a 3.45% semiannual coupon and matures on January 15, 2027. Throughout the sample period, LQD consistently reports higher prices than SPIB on the same trading day, indicating that LQD uses dirty pricing (which includes accrued interest), while SPIB uses clean pricing (which excludes it). The price differential ranges from approximately 0.015 to 0.017 in early January but drops to zero on January 15, 2019, the bond's coupon payment date. The gap then gradually widens again, resetting on July 15, 2019, consistent with the bond's semiannual coupon schedule. This recurring pattern supports the conclusion that LQD and SPIB use different pricing conventions.

To systematically identify the pricing convention used by each ETF, we begin by verifying that iShares' LQD reports dirty prices, based on a comparison with both clean and dirty prices from its index provider, Markit/S&P, as described in Section 6. For other ETFs, we compute the average absolute price difference relative to LQD, using both the raw reported price and the price adjusted for accrued interest. An ETF is classified as using dirty prices if the unadjusted price yields a smaller average difference; otherwise, it is classified as using clean prices. Our analysis indicates that only iShares and BlackRock ETFs report dirty prices, while all other ETFs in our sample report clean prices. This conclusion is consistent with information obtained through conversations with industry practitioners.

E Are Market Prices Inferred from ETF Daily Holdings Used in ETF Creations?

We use LQD as a case study to examine whether the market prices inferred from daily holdings are used in NAV calculations and in ETF creations.

Step 1: Validating Holdings Data Consistency

We begin by comparing LQD's daily bond holdings from Morningstar Direct, its shares outstanding from Bloomberg, and corresponding fund-level information published on the iShares website. Per regulatory requirements, iShares publishes current fund-level data and bond holdings at the start of each trading day (e.g., March 21, 2025), which reflect end-of-day information from the prior trading day (e.g., March 20, 2025). See the timeline in Figure A3 for clarification.

We verify that both Morningstar and Bloomberg reflect post-creation/redemption (C/R) share counts for LQD, whereas the iShares website reports pre-C/R data. After lagging the iShares data by one day, we find an exact match between Morningstar/Bloomberg and the iShares website, confirming the consistency and reliability of the Morningstar and Bloomberg datasets for our analysis. Notably, Morningstar reports bond market values based on dirty prices, while the iShares website also discloses clean prices.

Step 2: Reconstructing Total Net Assets (TNA)

We then compute LQD's total net assets (TNA) on a given day by applying the following procedure. For each bond, we infer its dirty price on day t (e.g., March 21, 2025) by dividing the reported market value by its par amount. We multiply this inferred dirty price by the par amount held as of day $t-1$ (e.g., March 20, 2025), as NAV on day t is used for C/R activity on that same day and thus should be calculated before the C/R activity. We sum across all bonds to calculate TNA. For March 21, 2025, this procedure yields a reconstructed TNA of $29,794,528,379.

Since dirty prices decline when coupon payments are made, we adjust for coupon income that increases the fund's cash holdings. However, changes in reported cash balances from day $t-1$ to day t are not informative, as they may reflect cash used or received in creation/redemption activity. Instead, we identify bonds with significant discrepancies between changes in clean and dirty prices, which signal coupon payments.

For instance, a CAMPBELLS CO bond maturing on March 21, 2034 with a 5.4% semiannual coupon shows a clean price drop of -0.2% but a dirty price drop of -2.9%. The 2.7% difference matches the semiannual coupon rate, implying a coupon payment of $13,310,000 $\times$ 2.7% = $359,370.

In total, 8 out of 2,897 bonds held by LQD show this pattern on March 21, 2025, leading to total coupon income of $2,201,909. Adjusting for these coupons, we revise the TNA on March 21 to $29,794,528,379 + $2,201,909 = $29,796,730,287.

Step 3: Validating NAV Calculation

NAV is computed before any creation/redemption occurs on a given day. We use the reconstructed TNA from Step 2 and divide it by the pre-C/R share count (274.1 million shares), reported by Bloomberg on March 20: $29,796,730,287 $\div$ 274,100,000 = $108.7075. This closely matches the official NAV of $108.71 (which is provided only up to two decimal places) reported on the iShares website and other sources such as Bloomberg and CRSP.

Step 4: Validating Creation/Redemption Flow

We infer the value of creation/redemption activity from Morningstar holdings data by:

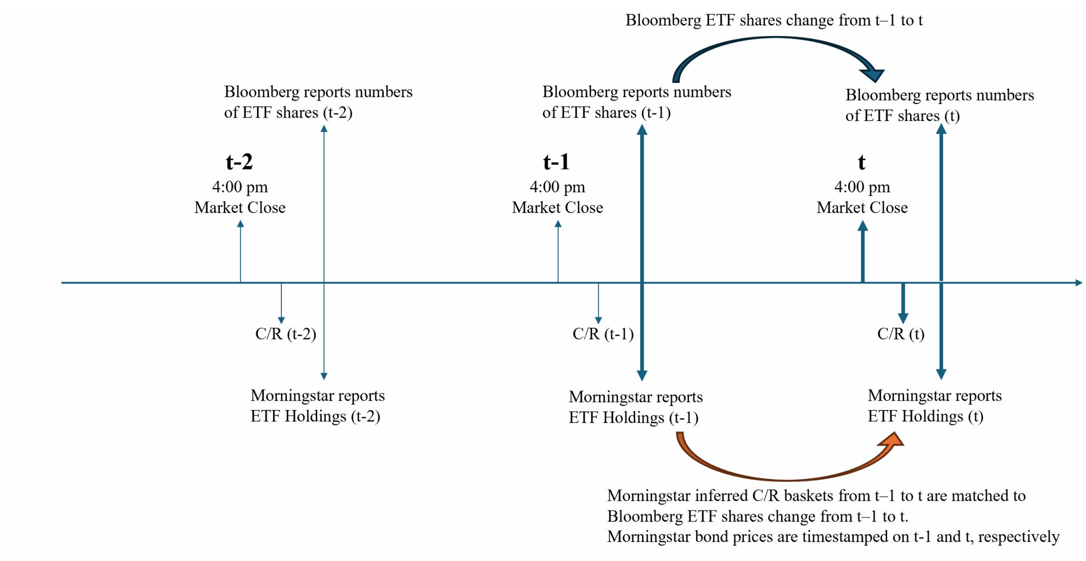
1. computing the change in par amount for each bond (and cash) between March 20 and March 21;

2. multiplying this change by the March 21 bond price; and
3. summing across all bonds and cash.

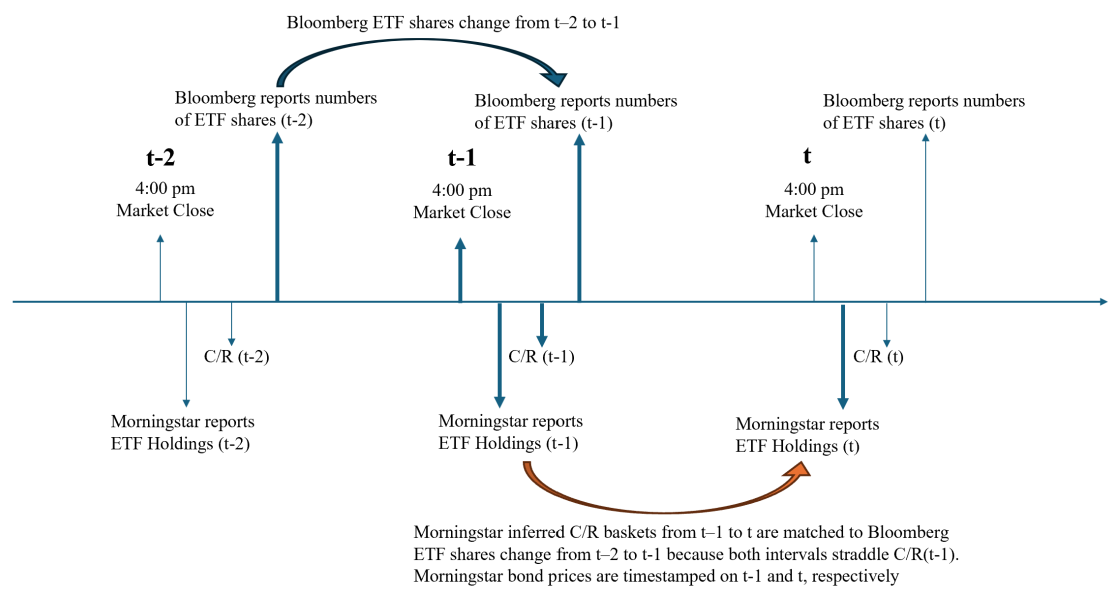
This yields a net redemption value of $-\$454,444,831$. After adjusting for the coupon income, as reflected in the change in the fund's cash position from Step 2, we obtain a final estimated net outflow of $456,646,740.

From Bloomberg, we observe a reduction in shares outstanding from 274.1 million to 269.9 million, a drop of 4.2 million shares. At an NAV of $108.71, this corresponds to a net outflow of: 4.2 million $\times$ 108.71 = $456,585,629. The difference between the two estimates is only $61,111, a negligible amount relative to the fund's size (over $29 billion in TNA) and the scale of the redemption. For context, LQD's daily expenses (based on an annual expense ratio of 0.14%) are approximately: $29.9 billion $\times$ 0.14% $\div$ 365 = $114,932. Other small adjustments—such as securities lending revenue or fund operating expenses—could account for the remaining gap.

This analysis confirms that the market prices inferred from daily ETF holdings data are indeed the same prices used in NAV calculation and by extension, in ETF share creation.

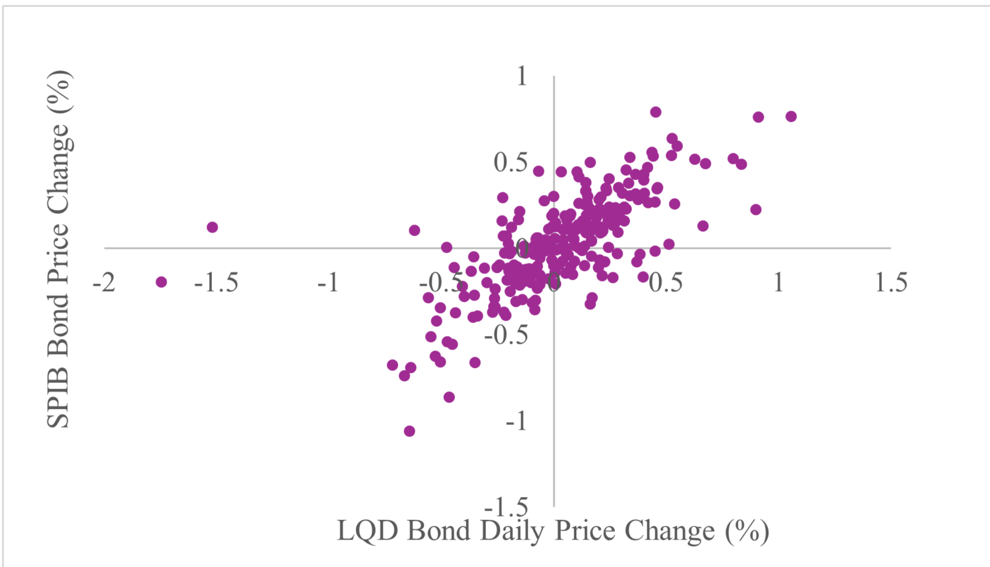


(a) ETF reports end-of-day holdings that incorporate the day's creation and redemption



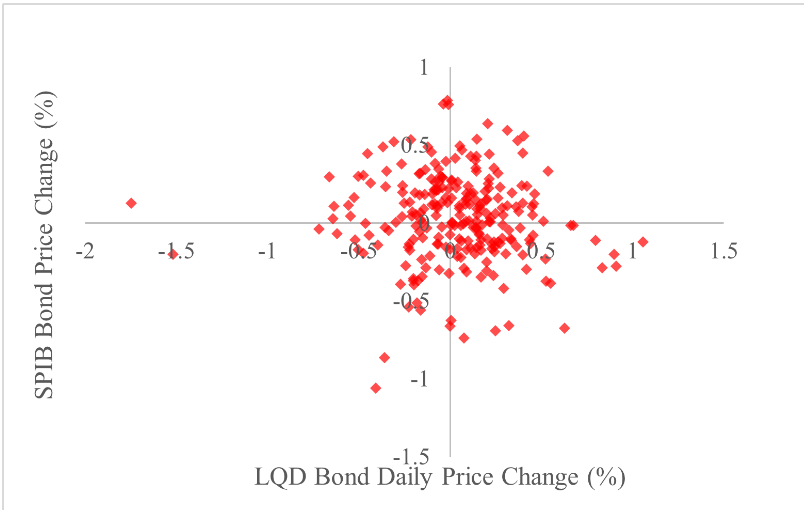
(b) ETF reports end-of-day holdings prior to incorporating that day's activity

Figure A1 Morningstar Holdings Timestamp This figure displays the timing that Morningstar reports ETF holdings with respect to the flows reported by Bloomberg. In Panel (a), ETFs report end-of-day holdings that incorporate the day's creation and redemption activity. For these ETFs, the inferred creation basket from day t–1 to t corresponds to the change in Bloomberg shares outstanding over the same interval. In Panel (b), ETFs report holdings prior to incorporating that day's activity. For these funds, the holdings-based basket from t–1 to t aligns with Bloomberg flows from t–2 to t–1.

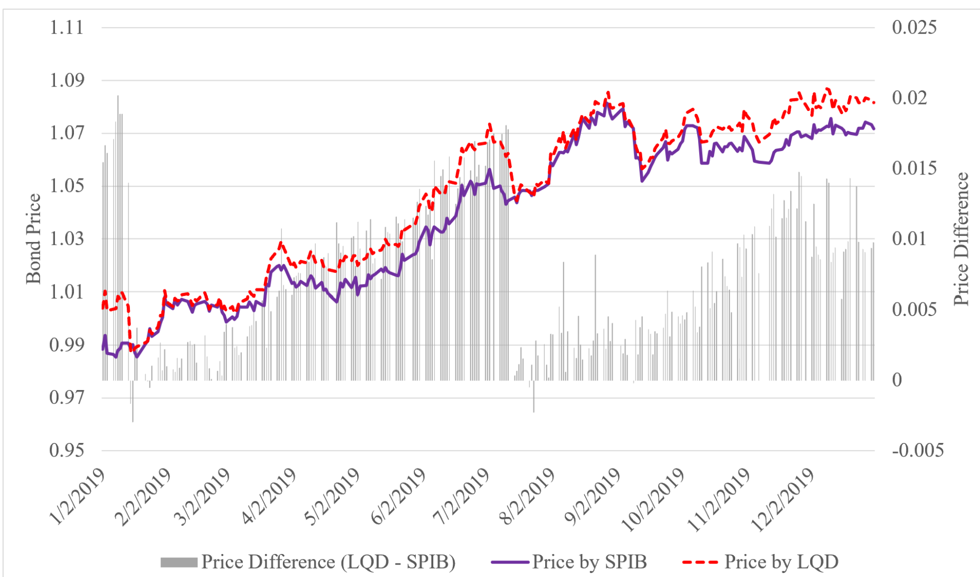


(a) Same day bond price changes by LQD and SPIB

Figure A2 Pricing reporting convention in Morningstar This figure illustrates Morningstar's price reporting convention, focusing on timing and the inclusion of accrued interest. In Panels (a) and (b), we plot same-day and lead-lag price changes for a bond (CUSIP: 91324PCW0, issued by UnitedHealth Group Incorporated) held by LQD and SPIB, knowing that LQD reports end-of-day holdings that incorporate the day's creation and redemption, while SPIB reports end-of-day holdings before incorporating that day's activity. The strong price correlations in Panel (a) and the lack of correlation in Panel (b) indicate that valuation data is consistently reported as of the end of the Morningstar timestamp, regardless of the reporting convention. In Panel (c), we plot the price of the same bond held by LQD and SPIB in 2019. LQD consistently reports higher prices than SPIB for the same trading day, indicating that LQD uses dirty pricing (which includes accrued interest), while SPIB uses clean pricing.



(b) Lead-lag bond price changes by LQD and SPIB



(c) Morningstar Bond Valuation by LQD and SPIB

Figure A2 (continued)

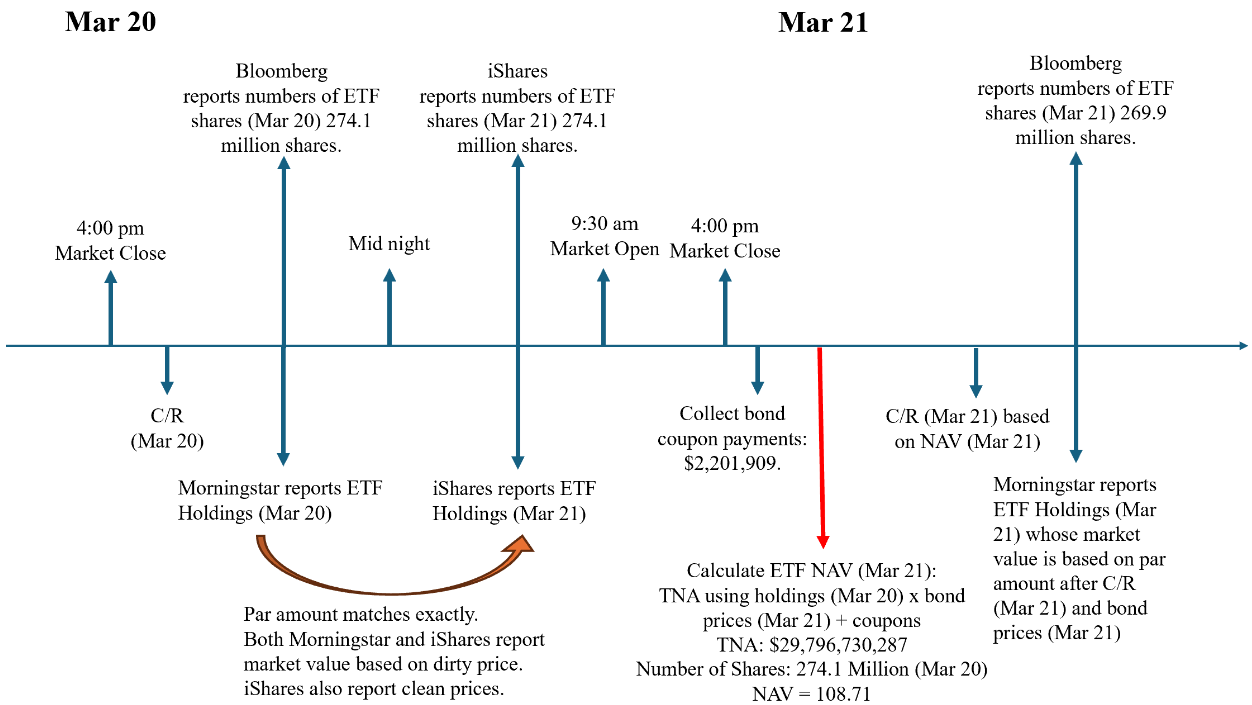


Figure A3 Timing in NAV calculation This figure displays timing for the holdings and flows data from Morningstar, Bloomberg, and the iShares website to calculate the NAV of the fund on Mar 21 and the value of the redemption basket on Mar 21. We note that the NAV on day t is used for C/R activity on that same day, and, thus, should be calculated before the C/R activity. Hence, the par amount of holdings on $t-1$ should be used to calculate the TNA on day t. Actual coupon payments from the holdings on $t-1$ should be adjusted.

Table A1: Determine ETF's reporting convention

This table shows how we identify each ETF's reporting convention using LQD and SPIB as examples. Some ETFs report end-of-day holdings after creation/redemption activity, while others report before. We estimate daily flows from Morningstar and correlate them with Bloomberg flows under two assumptions: (1) same-day correlation, and (2) lead-lag correlation. LQD aligns with the same-day convention, while SPIB reports holdings before C/R activity.

	LQD		**SPIB**	
	Same Day	Lead/Lag	Same Day	Lead/Lag
2017	0.7	0.1	-0.1	0.4
2018	0.7	0.1	0.1	0.6
2019	0.7	0.2	0.0	0.7
2020	0.6	0.3	0.0	0.6
2021	0.8	0.1	-0.1	0.8
2022	0.7	0.2	0.1	0.7
2023	0.7	0.2	-0.1	0.8

Table A2: Price Deviation and ETF Creation Baskets: Extended Samples
This table explores the relationship between cross-ETF price dispersion and a bond's likelihood of inclusion in the creation basket. Panel A replicates the analyses from Table VII using a broader sample that includes all fund-days with non-zero flows (both inflows and outflows). Panel B repeats the same analyses using the full sample that includes all fund-days, whether or not there are net flows. The dependent variable is $BondCreation$, a dummy variable equal to one if the number of bonds held by the ETF increases on the given day, and zero otherwise. Price Deviation ($PD_{f,b,t}$) is the fund-bond-day level measure: $PD_{f,b,t} = P_{f,b,t} - \widetilde{P_{b,t}}$, where $P_{f,b,t}$ is the price for bond b reported by ETF f on day t, and $\widetilde{P_{b,t}}$ is the median price of bond b reported by all ETFs that hold the bond on day t. $PD_{f,b,t}$ is measured in dollars per \$1 par to facilitate clearer economic interpretation of the coefficients (see Appendix A). $\overline{PD_{f,t}}$ represents the average $PD_{f,b,t}$ at the fund-day level and captures the ETF adjustment. The replacement of ($PD_{f,b,t} - \overline{PD_{f,t}}$) with $PD_{f,b,t}$ in columns (2)–(4) is made because $\overline{PD_{f,t}}$ becomes redundant when fund-date fixed effects are added. Columns (5) and (6) utilize rank-based measures derived from the PD measure as independent variables. For bond b on day t, we assign a percentile rank to each ETF based on the price it assigns to bond b. We then create four dummy variables for ETFs falling within the following rank ranges: 0%–20%, 20%–40%, 60%–80%, and above 80%. The t-statistics, shown in parentheses, are adjusted for two-way clustering at the date and fund levels. *p< .1; **p< .05; ***p< .01.

Panel A: Fund-bond-day sample with non-zero flows days (both inflows and outflows)

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var.	$BondCreation$ (Creation=1; Otherwise=0)					
$PD_{f,b,t} - \overline{PD_{f,t}}$	4.918***					
	(5.72)					
$PD_{f,b,t}$		4.525***	5.455***	6.027***		
		(5.76)	(8.57)	(9.06)		
Rank($PD_{f,b,t}$) in (50, 100]					0.0515***	
					(8.83)	
Rank($PD_{f,b,t}$)≤20						-0.0404***
						(-12.46)
Rank($PD_{f,b,t}$) in (20, 40]						-0.0179***
						(-4.05)
Rank($PD_{f,b,t}$) in (60, 80]						0.0311***
						(5.81)
Rank($PD_{f,b,t}$)>80						0.0544***
						(5.65)
Observations	62,236,177	62,236,177	62,236,177	62,236,177	62,236,177	62,236,177
Fund × date FE		Y	Y	Y	Y	Y
Bond × date FE			Y	Y	Y	Y
Fund × bond FE				Y	Y	Y
Adj R-squared	0.001	0.179	0.208	0.221	0.227	0.229

Table A2: (continued)

Panel B: Fund-bond-day sample with all fund-days

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var.	$BondCreation$ (Creation =1; Otherwise=0)					
$PD_{f,b,t} - \overline{PD_{f,t}}$	2.136***					
	(4.05)					
$PD_{f,b,t}$		1.959***	2.353***	2.594***		
		(4.11)	(5.12)	(5.16)		
Rank($PD_{f,b,t}$) in (50, 100]					0.0228***	
					(6.62)	
Rank($PD_{f,b,t}$)≤20						-0.0175***
						(-9.19)
Rank($PD_{f,b,t}$) in (20, 40]						-0.00667***
						(-3.45)
Rank($PD_{f,b,t}$) in (60, 80]						0.0154***
						(5.44)
Rank($PD_{f,b,t}$)>80						0.0243***
						(4.53)
Observations	157,143,077	157,143,077	157,143,077	157,143,077	157,143,077	157,143,077
Fund × date FE		Y	Y	Y	Y	Y
Bond × date FE			Y	Y	Y	Y
Fund × bond FE				Y	Y	Y
Adj R-squared	0.000	0.215	0.231	0.242	0.245	0.246

Table A3: Price Deviation and ETF Creation Baskets under CP+ Bid and Mid Adjustments

This table explores the relationship between cross-ETF price dispersion and a bond's likelihood of inclusion in the creation basket under three pricing treatments: (i) bond prices adjusted to the bid using the CP+ spread; (ii) bond prices adjusted to the mid using the CP+ spread; and (iii) no adjustment but restricted to the CP+ sample. The CP+ adjustment combines consolidated intraday secondary-market bond quotes with accrued interest data to produce "dirty" bid and mid prices for each bond. We use the quotes at the market close to align the bond-level prices with ETF-reported end-of-day valuations. For each ETF, we determine whether it reports bid- or mid-based valuations by selecting the smaller of the two average absolute deviations between its reported dirty prices and the CP+ bid and mid dirty prices, respectively. For (i) bond prices adjusted to the bid using the CP+ spread, we adjust the reported prices of ETFs identified as using mid prices by subtracting half of the CP+ bid-ask spread. For (ii) bond prices adjusted to the mid using the CP+ spread, we adjust the reported prices of ETFs identified as using bid prices by adding half of the CP+ spread. We use (iii), the unadjusted prices restricted to the CP+ sample, as the benchmark for comparison. The dependent variable is *BondCreation*, a dummy equal to one if the number of bonds an ETF holds increases on a given day. Price Deviation ($PD_{f,b,t}$) is defined as $P_{f,b,t} - \widetilde{P}_{b,t}$, measured in dollars per \$1 par to facilitate clearer economic interpretation of the coefficients (see Appendix A). Rank-based variables use the percentile rank of $PD_{f,b,t}$ across ETFs holding bond b on day t. We report dummies for 0–20%, 20–40%, 60–80%, and >80%. Panel A limits the sample to fund-days with net inflows, Panel B includes both inflow and outflow days, and Panel C covers the full sample of all fund-days. All regressions include fund–date, bond–date, and fund–bond fixed effects. t-statistics (in parentheses) are adjusted for two-way clustering by date and fund. *p< .1; **p< .05; ***p< .01.

Panel A: Fund-bond-day sample with positive flows days

	Bond prices adjusted to bid based on CP+ spread			Bond prices adjusted to mid based on CP+ spread			No adjustment but use bond-day in CP+ sample		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dep. Var.	*BondCreation* (Creation =1; Otherwise=0)								
$PD_{f,b,t}$	8.770***			8.750***			8.755***		
	(9.44)			(9.43)			(9.75)		
Rank($PD_{f,b,t}$) in (50, 100]		0.0624***			0.0623***			0.0702***	
		(10.09)			(10.07)			(9.20)	
Rank($PD_{f,b,t}$)< 20			-0.0426***			-0.0426***			-0.0544***
			(-10.47)			(-10.51)			(-10.72)
Rank($PD_{f,b,t}$) in (20, 40]			-0.0165***			-0.0166***			-0.0248***
			(-3.47)			(-3.49)			(-4.28)
Rank($PD_{f,b,t}$) in (60, 80]			0.0357***			0.0356***			0.0432***
			(5.61)			(5.60)			(5.85)
Rank($PD_{f,b,t}$)> 80			0.0726***			0.0723***			0.0724***
			(6.19)			(6.20)			(6.22)
Observations	30,626,734	30,626,734	30,626,734	30,626,734	30,626,734	30,626,734	30,626,734	30,626,734	30,626,734
Fund × date FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Bond × date FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Fund × bond FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Adj R-squared	0.242	0.249	0.253	0.242	0.249	0.253	0.242	0.251	0.254

Table A3: (continued)

Panel B: Fund-bond-day sample with non-zero flows days (both inflows and outflows)

	Bond prices adjusted to bid based on CP+ spread			Bond prices adjusted to mid based on CP+ spread			No adjustment but use bond-day in CP+ sample		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dep. Var.	$BondCreation$ (Creation =1; Otherwise=0)								
$PD_{f,b,t}$	6.639***			6.625***			6.641***		
	(8.34)			(8.34)			(8.69)		
Rank($PD_{f,b,t}$) in (50, 100]		0.0479***			0.0477***			0.0533***	
		(9.44)			(9.42)			(8.41)	
Rank($PD_{f,b,t}$)$<$ 20			-0.0336***			-0.0336***			-0.0428***
			(-11.15)			(-11.19)			(-11.77)
Rank($PD_{f,b,t}$) in (20, 40]			-0.0129***			-0.0129***			-0.0190***
			(-3.12)			(-3.13)			(-3.83)
Rank($PD_{f,b,t}$) in (60, 80]			0.0263***			0.0262***			0.0316***
			(5.11)			(5.11)			(5.39)
Rank($PD_{f,b,t}$)$>$ 80			0.0540***			0.0538***			0.0541***
			(5.74)			(5.75)			(5.70)
Observations	45,007,982	45,007,982	45,007,982	45,007,982	45,007,982	45,007,982	45,007,982	45,007,982	45,007,982
Fund × date FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Bond × date FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Fund × bond FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Adj R-squared	0.230	0.235	0.238	0.230	0.235	0.237	0.230	0.236	0.238

Table A3: (continued)

Panel C: Fund-bond-day sample with all fund-days

	Bond prices adjusted to bid based on CP+ spread			Bond prices adjusted to mid based on CP+ spread			No adjustment but use bond-day in CP+ sample		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dep. Var.	*BondCreation* (Creation =1; Otherwise=0)								
$PD_{f,b,t}$	2.938***			2.932***			2.970***		
	(4.86)			(4.85)			(4.97)		
Rank($PD_{f,b,t}$) in (50, 100]		0.0210***			0.0209***			0.0249***	
		(6.59)			(6.58)			(6.64)	
Rank($PD_{f,b,t}$)< 20			-0.0142***			-0.0142***			-0.0200***
			(-6.61)			(-6.62)			(-9.16)
Rank($PD_{f,b,t}$) in (20, 40]			-0.00441**			-0.00442**			-0.00745***
			(-2.56)			(-2.57)			(-3.26)
Rank($PD_{f,b,t}$) in (60, 80]			0.0124***			0.0123***			0.0165***
			(4.82)			(4.81)			(5.29)
Rank($PD_{f,b,t}$)> 80			0.0261***			0.0259***			0.0253***
			(4.81)			(4.81)			(4.65)
Observations	105,090,958	105,090,958	105,090,958	105,090,958	105,090,958	105,090,958	105,090,958	105,090,958	105,090,958
Fund × date FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Bond × date FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Fund × bond FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Adj R-squared	0.250	0.252	0.253	0.250	0.252	0.253	0.250	0.252	0.254

Table A4: Price Deviation and ETF Creation Baskets: iShares ETFs

This table explores the relationship between cross-ETF price dispersion and a bond's likelihood of inclusion in the creation basket. We replicate the analyses from column (4) in Table VII and Table A2 using sample of only iShares ETFs. The dependent variable is $BondCreation$, a dummy variable equal to one if the number of bonds held by the ETF increases on the given day, and zero otherwise. Price Deviation ($PD_{f,b,t}$) is the fund-bond-day level measure: $PD_{f,b,t} = P_{f,b,t} - \widetilde{P_{b,t}}$, where $P_{f,b,t}$ is the price for bond b reported by ETF f on day t, and $\widetilde{P_{b,t}}$ is the median price of bond b reported by all ETFs that hold the bond on day t. $PD_{f,b,t}$ is measured in dollars per $1 par to facilitate clearer economic interpretation (see Appendix A). The t-statistics, shown in parentheses, are adjusted for two-way clustering at the date and fund levels. *p< .1; **p< .05; ***p< .01.

	(1)	(2)	(3)
	Positive flow days	Non-zero flow days	All days
Dep. Var.	$BondCreation$ (Creation =1; Otherwise=0)		
$PD_{f,b,t}$	11.10***	8.631***	4.599***
	(10.01)	(10.52)	(6.16)
Observations	24,128,064	34,818,239	90,888,600
Fund × date FE	Y	Y	Y
Bond × date FE	Y	Y	Y
Fund × bond FE	Y	Y	Y
Adj R-squared	0.217	0.206	0.228

Table A5: Price Deviation and ETF Redemption Baskets

This table explores the relationship between cross-ETF price dispersion and a bond's likelihood of inclusion in the redemption basket. We replicate the analyses from Columns (4) and (6) in Table VII and Table A2 with a new dependent variable, $BondRedemption_{f,b,t}$, which equals one if the number of shares of bond b held by ETF f decreases on day t, and zero otherwise. Price Deviation ($PD_{f,b,t}$) is the fund-bond-day level measure: $PD_{f,b,t} = P_{f,b,t} - \widetilde{P_{b,t}}$, where $P_{f,b,t}$ is the price for bond b reported by ETF f on day t, and $\widetilde{P_{b,t}}$ is the median price of bond b reported by all ETFs that hold the bond on day t. $PD_{f,b,t}$ is measured in dollars per \$1 par to facilitate clearer economic interpretation of the coefficients (see Appendix A). Columns (2), (4) and (6) utilize rank-based measures derived from the PD measure as independent variables. For bond b on day t, we assign a percentile rank to each ETF based on the price it assigns to bond b. We then create four dummy variables for ETFs falling within the following rank ranges: 0%-20%, 20%-40%, 60%-80%, and above 80%. The t-statistics, shown in parentheses, are adjusted for two-way clustering at the date and fund levels. *p< .1; **p< .05; ***p< .01.

	(1)	(2)	(3)	(4)	(5)	(6)
	Negative flow days		Non-zero flow days		All days	
Dep. Var.	*BondRedemption*: Redemption =1; Otherwise=0					
$PD_{f,b,t}$	0.890***		0.193*		0.0839	
	(3.68)		(1.68)		(1.06)	
Rank($PD_{f,b,t}$)≤20		-0.00178***		0.00000620		-0.000203
		(-3.11)		(0.02)		(-1.13)
Rank($PD_{f,b,t}$) in (20, 40]		-0.000551		0.0000222		-0.000493
		(-1.37)		(0.12)		(-0.57)
Rank($PD_{f,b,t}$) in (60, 80]		0.000634*		-0.000465***		-0.000187
		(1.76)		(-2.90)		(-1.49)
Rank($PD_{f,b,t}$)>80		-0.000135		-0.00115***		-0.000508***
		(-0.17)		(-4.14)		(-2.63)
Observations	13,599,677	13,599,677	62,236,177	62,236,177	157,143,077	157,143,077
Fund × date FE	Y	Y	Y	Y	Y	Y
Bond × date FE	Y	Y	Y	Y	Y	Y
Fund × bond FE	Y	Y	Y	Y	Y	Y
Adj R-squared	0.484	0.484	0.501	0.498	0.504	0.500

Table A6: Price Deviation and ETF Creation Baskets: Alternative Approach

This table explores the relationship between cross-ETF price dispersion and a bond's likelihood of inclusion in the creation basket. The dependent variable is *BondCreation*, a dummy variable equal to one if the number of bonds held by the ETF increases on the given day, and zero otherwise. We replicate the analyses from Columns (4) and (6) in Table VII and Table A2 using raw bond prices $P_{f,b,t}$ as the independent variable, instead of $PD_{f,b,t} - \overline{PD_{f,t}}$. $P_{f,b,t}$ is measured in dollars per \$1 par to facilitate clearer economic interpretation of the coefficients. Our main results remain robust, as fund × date fixed effects absorb any fund-day-level variation in $P_{f,b,t}$ and bond × date fixed effects absorb bond-day-level variation. Since directly winsorizing $P_{f,b,t}$ is undesirable, we first compute $PD_{f,b,t} = P_{f,b,t} - \widetilde{P_{b,t}}$, winsorize $PD_{f,b,t}$ at 1% and 99%, and reconstruct $P_{f,b,t}(winsorized)$ as $\widetilde{P_{b,t}} + PD_{f,b,t}(winsorized)$. The t-statistics, shown in parentheses, are adjusted for two-way clustering at the date and fund levels. *p< .1; **p< .05; ***p< .01.

	(1)	(2)	(3)	(4)	(5)	(6)
	Positive flow days		Non-zero flow days		All days	
Dep. Var.	*BondCreation* (Creation =1; Otherwise=0)					
$P_{f,b,t}$	2.407***		1.464***		0.667***	
	(6.61)		(5.65)		(4.40)	
$P_{f,b,t}(winsorized)$		7.915***		6.027***		2.594***
		(9.61)		(9.06)		(5.16)
Observations	42,508,900	42,508,900	62,236,177	62,236,177	157,143,077	157,143,077
Fund × date FE	Y	Y	Y	Y	Y	Y
Bond × date FE	Y	Y	Y	Y	Y	Y
Fund × bond FE	Y	Y	Y	Y	Y	Y
Adj R-squared	0.232	0.233	0.220	0.221	0.242	0.242

Table A7: Impact of Bond Valuation on ETF Creation Over Time

This table explores the relationship between cross-ETF price dispersion and a bond's likelihood of inclusion in the creation basket. We replicate the analyses from Column (4) in Table VII and Table A2 with yearly sample from 2017 to 2023. Panel A uses a sample that includes only net inflows days, while Panels B and C include, respectively, non-zero flows days (both inflows and outflows) and all fund days. The dependent variable is *BondCreation*, a dummy variable equal to one if the number of bonds held by the ETF increases on the given day, and zero otherwise. Price Deviation ($PD_{f,b,t}$) is the fund-bond-day level measure: $PD_{f,b,t} = P_{f,b,t} - \widetilde{P_{b,t}}$, where $P_{f,b,t}$ is the price for bond b reported by ETF f on day t, and $\widetilde{P_{b,t}}$ is the median price of bond b reported by all ETFs that hold the bond on day t. $PD_{f,b,t}$ is measured in dollars per \$1 par to facilitate clearer economic interpretation of the coefficients (see Appendix A). The t-statistics, shown in parentheses, are adjusted for two-way clustering at the date and fund levels. *p< .1; **p< .05; ***p< .01.

Panel A: Fund-bond-day sample with net inflows days

	(1) 2017	(2) 2018	(3) 2019	(4) 2020	(5) 2021	(6) 2022	(7) 2023
Dep. Var.	*BondCreation* (Creation =1; Otherwise=0)						
$PD_{f,b,t}$	7.659***	10.06***	9.121***	7.069***	12.96***	6.998***	5.417***
	(6.22)	(7.15)	(7.01)	(7.80)	(11.22)	(6.90)	(5.83)
Observations	1,697,889	2,911,248	5,191,757	7,783,688	7,941,857	7,307,275	9,657,388
Fund × date FE	Y	Y	Y	Y	Y	Y	Y
Bond × date FE	Y	Y	Y	Y	Y	Y	Y
Fund × bond FE	Y	Y	Y	Y	Y	Y	Y
Adj R-squared	0.183	0.266	0.201	0.219	0.233	0.271	0.272

Table A7: (continued)

Panel B: Fund-bond-day sample with non-zero flows days (both inflows and outflows)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	2017	2018	2019	2020	2021	2022	2023
Dep. Var.	*BondCreation* (Creation =1; Otherwise=0)						
$PD_{f,b,t}$	6.067***	7.613***	7.392***	5.204***	10.09***	4.802***	4.386***
	(6.66)	(8.77)	(6.77)	(7.27)	(9.71)	(5.98)	(5.39)
Observations	2,319,101	4,549,534	7,160,317	10,440,484	11,465,615	11,819,851	14,470,058
Fund × date FE	Y	Y	Y	Y	Y	Y	Y
Bond × date FE	Y	Y	Y	Y	Y	Y	Y
Fund × bond FE	Y	Y	Y	Y	Y	Y	Y
Adj R-squared	0.168	0.241	0.189	0.214	0.218	0.262	0.247

Panel C: Fund-bond-day sample with all days

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	2017	2018	2019	2020	2021	2022	2023
Dep. Var.	*BondCreation* (Creation =1; Otherwise=0)						
$PD_{f,b,t}$	2.017***	2.456***	2.762***	2.382***	4.738***	1.893***	2.111***
	(2.75)	(4.00)	(4.58)	(4.96)	(5.31)	(4.14)	(3.91)
Observations	11,286,321	14,969,175	18,503,407	23,339,362	27,230,026	30,356,403	31,457,059
Fund × date FE	Y	Y	Y	Y	Y	Y	Y
Bond × date FE	Y	Y	Y	Y	Y	Y	Y
Fund × bond FE	Y	Y	Y	Y	Y	Y	Y
Adj R-squared	0.200	0.263	0.203	0.224	0.248	0.287	0.262

Table A8: Dispersions in Bond Valuations within an ETF, Creations, and Redemptions
This table examines the relationship between within-ETF relative bond price dispersion and a fund's likelihood of receiving creations or redemptions. The dependent variables are $FundCreation_{f,t}$ and $BondRedemption_{f,t}$, which equal one if the number of ETF shares outstanding increases or decreases on day t, respectively, and zero otherwise. The key fund-day level independent variable, $Fund\,PD\,Dispersion_{f,t}$, is the interquartile range of price deviation ($PD_{f,b,t}$) for all bonds in ETF f on day t. $Fund\,PD\,Dispersion_{f,t}$, ETF premiums, and ETF bid–ask spreads are expressed in basis points. To facilitate clearer economic interpretation of coefficients, both dependent dummies are multiplied by 100. t-statistics (in parentheses) are adjusted for two-way clustering at the date and fund levels. *p< .1; **p< .05; ***p< .01.

	(1)	(2)	(3)	(4)
Dep. Var.	$FundCreation_{f,t}(\times 100)$		$BondRedemption_{f,t}(\times 100)$	
$Fund\ PD\ Dispersion_{f,t}$ (equal weighted)	0.1017**		-0.0002	
	(2.24)		(-0.48)	
$Fund\ PD\ Dispersion_{f,t}$ (value weighted)		0.0914*		-0.0001
		(1.85)		(-0.27)
ETF Premium	0.1871***	0.1868***	0.0103***	0.0103***
	(7.26)	(7.24)	(2.88)	(2.89)
log(Fund Size)	0.0511***	0.0510***	-0.0013***	-0.0013***
	(7.95)	(7.91)	(-5.73)	(-5.73)
ETF bid–ask spread	-0.0881***	-0.0883***	0.0002	0.0002
	(-3.78)	(-3.78)	(1.05)	(1.05)
Observations	121,486	121,486	121,486	121,486
Fund FE	Y	Y	Y	Y
Date FE	Y	Y	Y	Y
Adj R-squared	0.213	0.213	0.187	0.187