Bond Valuation Dispersion and ETF Creation*

Alan Guoming Huang
University of Waterloo
aghuang@uwaterloo.ca

Russ Wermers
University of Maryland
wermers@umd.edu

Jinming Xue
Southern Methodist University
jinmingx@smu.edu

Xing (Alex) Zhou
Southern Methodist University
and Office of Financial Research at the Department of the Treasury
axzhou@smu.edu

November 2024

Abstract

This paper presents a novel mechanism through which the lack of consensus on corporate bond valuation affects corporate bond ETF creation. Although multiple ETF issuers can accept the same bond during the "in-kind" creation process, they often value it differently. Valuation dispersion is driven by external pricing sources but can also reflect ETFs' efforts to minimize tracking errors. In their role as Authorized Participants (APs), dealers can benefit by delivering a bond to the ETF that values it higher while maintaining a lower overall valuation tendency. This strategic behavior benefits APs but dilutes the holdings of existing ETF investors to an extent similar to that of the funds' expense ratio.

JEL Classification: G12, G23

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

* The authors gratefully acknowledge helpful comments from Chotibhak (Pab) Jotikasthira, Jimmy Karam, Ananth Madhavan, Ke Shen, John Shim, Karamfil Todorov, David Xiaoyu Xu, and Haibei Zhao. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Office of Financial Research or the Department of the Treasury.

Bond Valuation Dispersion and ETF Creation

Abstract

This paper presents a novel mechanism through which the lack of consensus on corporate bond valuation affects corporate bond ETF creation. Although multiple ETF issuers can accept the same bond during the "in-kind" creation process, they often value it differently. Valuation dispersion is driven by external pricing sources but can also reflect ETFs' efforts to minimize tracking errors. In their role as Authorized Participants (APs), dealers can benefit by delivering a bond to the ETF that values it higher while maintaining a lower overall valuation tendency. This strategic behavior benefits APs but dilutes the holdings of existing ETF investors to an extent similar to that of the funds' expense ratio.

JEL Classification: G12, G23

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

1. Introduction

Trading in corporate bond market is largely intermediated by a small number of large bond dealers, many of whom also serve as authorized participants (APs) for exchange-traded funds (ETFs). Over the past decade, corporate bond ETFs have become increasingly crucial in supporting dealers' liquidity provision in the corporate bond market.[1] To avoid accumulating large positions in illiquid bonds, dealers can turn to the primary ETF market, where they deliver bonds to the ETF issuer in exchange for liquid ETF shares. This process is facilitated by the use of representative sampling in the creation of new shares, which allows APs to exchange a small, negotiated basket of bonds—rather than replicating the entire index tracked by the ETF—when creating new shares.[2]

A key characteristic of the corporate bond ETF primary market is its concentration: a small number of APs are responsible for creating shares in nearly all bond ETFs.[3] Although many of these ETFs are capable of accepting the same bonds, a single bond may not necessarily be valued at the same price across different ETFs. Due to the lack of continuous trading in corporate bonds, ETF issuers rely on a variety of data sources and valuation models, which can result in significant discrepancies in bond valuations across ETFs. This raises an important question: how does valuation dispersion across ETFs influence APs' decisions on which ETF to choose when converting their bond inventories into new ETF shares?

In this paper, we analyze the strategic behavior of APs in response to discrepancies in bond valuations across ETFs and the subsequent impact on ETF investors. We show that an AP can maximize the potential benefits from the creation process by delivering a bond to the ETF that values it the most. To illustrate, Table 1 presents the prices reported by eleven ETFs that held a

[1] See for example, Dannhauser and Karmaziene (2023), Du (2023), Koont, Ma, Pastor, and Zeng (2024), Pan and Zeng (2017), Shim and Todorov (2023), Reilly (2022), among others.

[2] Our paper focuses on the strategic behavior of APs during the creation process, as they have the flexibility to negotiate customized creation baskets with an ETF sponsor. This flexibility, however, does not necessarily extend to the redemption process, where redemption baskets could potentially be non-negotiable. For further details, see Koont, Ma, Pastor, and Zeng (2024) and Shim and Todorov (2023), as well as the commentary released by BlackRock, the largest provider of corporate bond ETFs. See https://www.blackrock.com/corporate/literature/publication/sec-proposed-rule-exchange-traded-funds-092618.pdf.

[3] See 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.

bond issued by AT&T Inc. (CUSIP: 00206RJY9) on July 6, 2020, along with information on whether the bond was included in the creation basket for each ETF. As shown in Table 1, the eleven ETFs assigned different values to the bond, with prices ranging from 105.24% to 105.83% of its par value—a spread of 59 basis points (bps). Among the eleven ETFs holding the bond, six experienced net inflows. However, this bond was included only in LQD's creation basket, which assigned it the highest value among the ETFs.

The case of AT&T 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 eight ETFs at any given time. The prices reported by different ETFs vary considerably, with the standard deviation and interquartile range of cross-ETF price deviations from the median reaching 76.8 bps and 41.5 bps, respectively. These large price discrepancies across ETFs appear to be driven by the illiquidity of corporate bonds. ETFs are more likely to disagree on bond prices when a bond trades infrequently (or not at all). Valuation discrepancies tend to be larger for older bonds, lower-rated bonds, those with smaller issuance sizes, and bonds with longer maturities. 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] While valuation dispersion for investment-grade (IG) bonds is generally lower than that for high-yield (HY) bonds, it skyrocketed, surpassing the dispersion observed for HY bonds at the peak of the COVID liquidity crisis. This aligns with the greater deterioration of liquidity in the IG sector during that period.[5]

Valuation dispersion across ETFs creates an opportunity for APs to exploit during the creation process. For a given underlying bond, an AP can exchange it for more ETF shares if the ETF issuer assigns a higher value to the bond. It is important to note that the profitability of exchanging an underlying bond for ETF shares also depends on the overall valuation tendencies of the ETF. The benefit of using bonds to create shares with an ETF that assigns the highest value to those bonds diminishes—or may even disappear—if the ETF generally values all bonds highly.

[4] See for example, Boyarchenko, Kovner, and Shachar (2022), Kargar, Lester, Lindsay, Liu, Weill, and Zúñiga (2021), and O'Hara and Zhou (2021).
[5] See for example, Haddad, Moreira, and Muir (2021), and Ma, Xiao, and Zeng (2022).

Using a simple model, we illustrate that APs gain the most from delivering a bond to an ETF that assigns it the highest value while maintaining a lower overall valuation tendency.

Empirically, we find strong evidence supporting this mechanism. Using a panel constructed at the fund-bond-day level, we show that, on any given day, a bond is more likely to be included in the creation basket of an ETF that assigns it the highest value, after controlling for the ETF's overall valuation tendency. Specifically, when an ETF ranks in the top 10% for valuing a bond relative to all other ETFs holding that bond, the probability of an AP delivering the bond to that ETF increases by 16%. This effect is substantial, especially considering the relatively low unconditional probability of only 7.1% for a bond to be included in a creation basket.

Our results remain robust even after controlling for several stringent fixed effects. One potential concern is that the likelihood of a bond's inclusion in an ETF's creation basket may be influenced by deviations between the ETF's market price and its net asset value (NAV). When an ETF trades at a premium, it creates an arbitrage opportunity for APs, increasing the probability that the ETF will receive bond deliveries for creation on a given day. To address this concern and account for other time-varying ETF-level characteristics that might affect creation probabilities, we include fund-day fixed effects in our analysis, and our results hold robustly. Another potential concern is that shocks to a specific bond—such as fire sales triggered by credit events or liquidity shocks—may cause dealers to hold more inventory of that bond, thereby increasing the likelihood of the bond's inclusion in the creation basket. To control for potential time-varying bond-level shocks, we include bond-day fixed effects, and again, our results remain robust. Furthermore, by incorporating fund-bond fixed effects to account for potential ETF preferences for specific bonds, we find that our results continue to hold, reinforcing the robustness of our findings.

What drives the dispersion in valuations across ETFs? A key factor is the role of tracking index providers, who typically supply indicative or end-of-day pricing 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, when comparing the prices reported by the iShares iBoxx $ Investment Grade Corporate Bond ETF (LQD) with those provided by Markit Indices GmbH for the iBoxx Liquid Investment Grade (USD) Index—the benchmark tracked by LQD—LQD's prices generally align closely with Markit's. Yet, when discrepancies occur, even

small deviations can have a substantial impact on the bonds included in LQD's creation basket. Our analysis suggests that these price adjustments are likely aimed at minimizing tracking error by selectively attracting certain bonds into the ETF's portfolio.

We explore the implications of APs' strategic behavior in the context of cross-ETF valuation dispersion for both ETFs and their investors. Our findings show that strategic behavior by APs during the creation process can significantly influence the ETF industry. Whether an ETF issuer receives a bond in its creation basket depends not only on the value it assigns to the bond but also on its overall valuation strategy. An AP would generally prefer to create new ETF shares when an ETF exhibits considerable dispersion in bond valuations across their holdings—assigning high values to some bonds while valuing others less—rather than when it assigns consistently high or low prices to all bond holdings. Indeed, our analysis shows that ETFs are more likely to receive creations on a given day when the dispersion in their valuations—relative to peers— is greater across bonds.

While APs benefit from this strategic behavior, it comes at the expense of existing ETF investors. When the ETF sponsor accepts bonds at valuations higher than those of its peers, it results in the dilution of existing shareholders' holdings. Our analysis shows that the magnitude of this dilution is comparable to the funds' expense ratios. Specifically, for ETFs where valuation deviations significantly attract bonds during the creation process, the annual dilution rate mirrors the average expense ratio, leading to total losses exceeding $114 million per annum.

Our paper contributes to the growing literature on corporate bond ETFs, particularly in understanding the creation and redemption mechanism and its implications for both ETFs and the underlying corporate bond markets. Several studies have examined these processes in detail. For example, Shim and Todorov (2021) document that in the primary market for corporate bond ETFs, APs often use fractional baskets—baskets that represent only a small fraction of ETF holdings—particularly in the creation process. Koont, Ma, Pastor, and Zeng (2024) show that corporate bond ETFs actively adjust their creation and redemption baskets to correct tracking errors while simultaneously facilitating liquidity transformation. While such active portfolio management enhances liquidity in the underlying bonds during normal market conditions, it can exacerbate liquidity issues during periods of stress. Shim and Todorov (2023) propose a model illustrating

how liquidity shocks in ETFs holding illiquid corporate bonds are transmitted to the underlying bond markets through the creation and redemption process. Their findings highlight how APs act as a buffer between the ETF and the bond markets, with liquidity-demanding ETF investors absorbing transaction costs through ETF premiums and discounts. Reilly (2022) further demonstrates that APs may strategically include underperforming bonds in creation baskets, which can lower the future performance of ETFs.

Our paper is particularly aligned with studies that investigate the role played by ETFs in facilitating dealers' inventory management in corporate bond markets. Pan and Zeng (2017) highlight how balance sheet constraints can create conflicts between APs' dual roles as ETF arbitrageurs and bond dealers. Even when an ETF is trading at a discount, APs may opt to create more shares if they hold large positions in the underlying bonds, potentially exacerbating ETF mispricing. Du (2023) finds that corporate bond dealers often include bonds already in their inventory when creating ETF baskets, while Dannhauser and Karmaziene (2023) show that dealers use ETFs to manage inventory risks, including the delivery of downgraded bonds in the creation process. Building on this literature, our paper explores how APs leverage ETFs to manage their bond inventories. Unlike previous studies, which generally treat ETFs as a homogeneous group, we investigate the significant variations across ETFs and demonstrate how these differences influence the allocation of specific bonds to different ETFs.

Our study contributes to this literature by shedding light on the previously underexplored role of cross-ETF valuation differences in shaping the creation process. We add to the body of research examining the liquidity transformation role of ETFs and the behavior of APs in managing bond inventories. In doing so, we highlight how seemingly small valuation discrepancies can have a significant impact on ETF flows and, ultimately, on investor returns. Moreover, by showing how valuation dispersion affects the allocation of bonds to different ETFs, our paper provides valuable insights into how APs' strategic behavior can influence the broader functioning of the corporate bond ETF market.

Our paper is also related to research that documents the dispersion in month-end valuations of identical bonds across different mutual funds and investigates the factors driving this variation. Cici, Gibson, and Merrick (2011) link valuation dispersion, when aggregated to fund NAV, to

return smoothing practices by fund managers, while Choi, Kronlund, and Oh (2024) argue that differences in bond valuations between mutual funds are influenced by same-day fund flows. They suggest that funds adjust the prices of their holdings, and consequently fund NAV, in response to investor flows, often without explicit disclosure. In a similar vein, we find that ETF issuers adjust bond valuations to minimize their tracking error. However, our study differs in that we focus on how valuation discrepancies at the bond level affect the behavior of APs, a feature unique to the ETF market. Additionally, we examine how the strategic behavior of APs impacts ETF flows and investor wealth, an aspect not addressed in previous research.

The paper is organized as follows: Section 2 provides the institutional background for our study. Section 3 introduces a toy model to illustrate 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 uses LQD as a case study to explore the drivers behind bond valuation dispersions across ETFs. Section 7 examines the impact of APs' strategic behavior on ETF flows and investor wealth. Finally, Section 8 concludes.

2. Institutional Background

The illiquid nature of the corporate bond market poses significant challenges for dealers' liquidity provision. Dealers often find themselves holding substantial inventories of bonds that may be difficult to trade, particularly in times of market stress or when liquidity is low. Over the past decade, corporate bond ETFs have emerged as an important tool for dealers to address these challenges by facilitating liquidity transformation and inventory management. To avoid accumulating large positions in illiquid bonds, dealers can turn to the corporate bond ETF market, where they can deliver bonds to an ETF issuer in exchange for liquid ETF shares. This process allows dealers to offload illiquid bonds while gaining exposure to a diversified basket of bonds via ETF shares. These shares, being highly liquid and traded on exchanges, offer dealers an easy exit from their bond positions without incurring significant transaction costs or price discounts associated with selling individual bonds in the secondary market. In this way, corporate bond ETFs

serve as an important vehicle for dealers to manage their bond inventories and reduce liquidity risk.

A distinctive feature of corporate bond ETFs is the in-kind creation process, which allows APs — typically bond dealers or large institutional investors — to exchange a basket of bonds for ETF shares. In the primary market, when new shares of an ETF are created, an AP delivers a specific basket of bonds 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 does not receive cash but rather ETF shares, which represent a pro-rata stake in the ETF's underlying bond portfolio. By allowing APs to exchange illiquid, hard-to-trade bonds for liquid ETF shares, this creation mechanism supports efficient inventory management and reduces the risk of holding large amounts of less liquid securities.

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.[6] This 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 managers an efficient means for portfolio rebalancing, ultimately benefiting investors. For APs who also serve as bond dealers, this system presents an opportunity to overweight bonds that they wish to offload. For example, an AP can include more of the bonds it needs to get rid 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 sample.

A key aspect of the corporate bond ETF market is that a single dealer can serve as an AP for multiple ETFs. Since many corporate bond ETFs track similar indices or focus on similar

[6] 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 differenc between the value of the bonds baskets and the NAV of a creation unit.

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. This allows the dealer to offload bonds across a wide range of ETFs, increasing the likelihood that they can effectively manage their inventory. Moreover, given the relatively small number of APs active in the corporate bond ETF market, dealers often serve as APs for several competing ETFs. This market concentration means that the same bond could potentially 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, leading to valuation dispersion across ETFs.

These 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 to ETFs that assign higher values to those bonds, as this will maximize the number of ETF shares they receive in exchange. However, this strategy is also dependent on the overall valuation tendencies of the ETF. If an ETF tends to value all bonds highly, the benefit of delivering a bond to that ETF may be diminished, making it more advantageous for APs to seek out ETFs that assign a lower overall value to their portfolios while valuing specific bonds highly.

3. A Toy Model

We present a toy model in Table 2 to demonstrate how dispersions in bond valuations across ETFs influence the preferred custom creation baskets by APs. Suppose there are two bonds, Bond 1 and Bond 2, held by four ETFs with identical portfolios, each consisting of one unit of Bond 1 and one unit of Bond 2. The reported bond prices differ across the four ETFs A, B, C, and D: the prices for Bond 1 are 102, 105, 105, and 105, respectively, while the prices for Bond 2 are 100, 100, 103, and 97, respectively. Fixed income ETFs typically allow for custom creation baskets

that do not exactly replicate the ETFs' portfolio holdings. Due to the differences in bond valuations, if an AP decides to create ETF shares using only Bond 1, the AP will receive a different number of ETF shares for each ETF in exchange for one unit of Bond 1.

Comparing ETF A and ETF B, both assign a value of 100 to Bond 2, but ETF A values Bond 1 at 102 while ETF B values Bond 1 higher at 105. Consequently, the NAVs differ, with ETF A's NAV at 202 and ETF B's NAV at 205. Suppose an AP uses a custom creation basket containing only Bond 1. By surrendering one unit of Bond 1, the AP can create 0.505 shares of ETF A, whereas with ETF B, the AP would receive 0.512 shares.[7] Given that the underlying composition of the two ETFs is identical, if the market is efficient enough, the secondary market prices for the ETFs would be the same.[8] Thus, by receiving more shares, the AP would prefer to create ETF B over ETF A using Bond 1.

The above comparison between ETFs A and B suggests that, all else being equal, it is optimal to include a bond in the creation basket for the ETF that assigns the bond a higher value. Extending this analysis to ETFs C and D, we demonstrate that the valuation dispersions of both Bond 1 and Bond 2 influence the decision to include Bond 1 in the custom creation basket, as Bond 2's valuation affects the ETF's NAV and thereby affects the number of ETF shares received in exchange. For example, ETF C values both bonds higher (Bond 1 at 105 and Bond 2 at 103), yet an AP would receive only 0.505 shares of ETF C for one unit of Bond 1, the same as with ETF A, which values both bonds lower by 2. Therefore, creating shares with ETF C is also less desirable than creating shares with ETF B, which values Bond 1 at 105 and Bond 2 at 100. This example highlights what ultimately matters is the relative price dispersion—how an ETF values a bond in comparison to the other bonds in its portfolio. This concept is reinforced by the sample of ETF D, which values Bond 1 at 105 and assigns an even lower value of 97 to Bond 2. With an NAV of

[7] Given the NAV of ETF A is 202 and Bond 1 is valued at 102, the AP could get 102/202=0.505 shares of ETF for each unit of Bond 1. Similarly, AP would receive 105/205=0.512 share.

[8] Without the assumption on the secondary market ETF prices, APs can exit the position by redeeming the ETF shares in the primary market before selling to customers. APs are often bond dealers, who at the same time often utilize the primary corporate bond ETF markets to manage their bond inventories. Since redemption baskets are typically more diversified, APs are less likely to receive a redemption basket concentrated in bonds with relatively higher price deviation.

202, ETF D allows an AP to receive the highest amount (0.520 shares) using one unit of Bond 1 in the creation basket, compared to other ETFs.

We note that this mechanism hinges on the within-fund relative valuation dispersion across bonds—specifically, the deviation in a bond's valuation by an ETF, adjusted for the ETF's overall tendency to value its portfolio higher or lower. To illustrate this, we first adjust each ETF's valuation of Bond 1 by the ETF's overall valuation across both bonds. We then link this adjusted valuation to the amount of ETF shares that can be exchanged for one unit of Bond 1. 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 Bond 1 and Bond 2 within the given fund. For example, ETF B reports prices of 105 for Bond 1 and 100 for Bond 2, and the median prices of the two bonds are also 105 and 100, respectively, as well. Consequently, Bond 1 in ETF B has an adjusted PD of 0, as both the PD and the ETF adjustment are 0. In contrast, while the PD for Bond 1 is also 0 in ETFs C and D that both value Bond 1 at 105, their ETF adjustments are 1.5 and -1.5, respectively, because ETF C values Bond 2 at 103 and ETF D at 97. 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. Next, we connect adjusted PD to the shares of ETFs that can be created using one unit of Bond 1. Bond 1 in ETF D has the highest adjusted PD of 1.5, with the greatest number of ETF shares to be created, followed by ETF B with the second-highest adjusted PD and the second largest number of share creation. In contrast, ETFs A and C have the lowest adjusted PD values, which are associated with the smallest number of ETF shares that can be created with one unit of Bond 1.[9] Taken together, the adjusted PD for a bond held by an ETF reflects the attractiveness of including this bond in the creation basket for that ETF.

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

[9] Although our discussion focuses on Bond 1, a similar analysis can be conducted on Bond 2, revealing that it is optimal to use Bond 2 to create ETFs A and C.

Income" funds in Morningstar with substantial investments in corporate bonds.[10] Target date ETFs are excluded, as these funds tend to increase cash holdings as they approach their closure dates (Koont et al. (2024)). 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.[11] For each corporate bond ETF, we gather historical daily portfolio holdings from Morningstar covering the period from 2017 to 2021. Along with par amount of the holding, the data also include each ETF's estimated valuation for every bond it holds. Our sample comprises 60 corporate bond ETFs holding a total of 24,256 unique bonds, resulting in 86,156,441 fund-bond-day observations.[12]

Since the actual creation baskets made by APs 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. (2024)). We identify these flow days using changes in the daily number of shares outstanding from 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. Net inflow days ensure that the ETF experiences at least one creation, while net outflow days can result from both creations and redemptions. By restricting our sample to days with non-zero ETF-level flows, we derive a primary dataset comprising 32,946,739 fund-bond-day observations. 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 issue 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 construct a daily composite rating for each bond that integrates the ratings from these agencies.[13]

---

[10] 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" and "US Fund Intermediate Core-Plus Bond," "US Fund Multisector Bond," and "US Fund Short-Term Bond."

[11] 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.

[12] Our sample starts in 2017, as there are few corporate bond ETFs prior to 2017. Following Cici, Gibson, and Merrick (2011), we require that a bond be held by at least three ETFs on a given day.

[13] 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

Table 3 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 $220 million, indicating the presence of a few large ETFs. During the sample period, ETFs trade at a slight premium, with both the mean and median price-to-NAV ratio at 0.08%. Notably, price deviations from NAV do occur, with the top and bottom deciles reflecting a 0.20% discount and a 0.41% premium, respectively. Panel B shows that over three-quarters of the bonds are investment-grade, with an average (median) remaining time to maturity of 10.22 (6.55) years and an average (median) age of 4.16 (3.20) years.

5. ETF Adjusted Bond Valuation Dispersion and AP Creation Basket Composition

In this section, we demonstrate that the mechanism by which adjusted bond valuation dispersion across ETFs influences APs' creation basket decisions, as illustrated in our toy model in Section 3, is broadly evident in our sample. We start with estimating the bond valuation dispersions across ETFs and analyze the factors driving 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.

5.1 Dispersion of bond valuations across ETFs and Bond Liquidity

To generalize the example, we assess how an ETF's valuation of a specific bond differs from that of other ETFs holding the same bond on the same day. We achieve this by estimating a fund-bond-date level price deviation measure ($PD_{f,b,t}$):

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

where $P_{f,b,t}$ is the price for bond b reported by ETF f on day t, and $\widetilde{P_{g,b,t}}$ is the median price of bond b reported by any ETF g that holds the bond on day t. At the fund-bond-day level, we report both the standard deviation and the interquartile range of the price deviation (PD) measure as proxies for the dispersion of bond valuations across ETFs. Table 4 shows the standard deviation and interquartile range of price deviation are 76.8 bps and 41.5 bps, respectively. At both ends 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.

the distribution, the 1 and 99th percentiles of the reported price deviation exceed 200 bps. We also analyze price dispersion separately for IG and HY bonds, finding that absolute price dispersions are greater for HY bonds. Specifically, the standard deviation and interquartile range of the PD measure are 72.2 bps and 38.3 bps for IG bonds, compared to 94.2 bps and 58.9 bps for HY bonds.

Bond illiquidity appears to significantly contribute to price dispersion across ETFs. Panel B shows that, on average, a bond is held by eight ETFs on any given day. For each bond on each day, we compute a price dispersion measure, $PD\ Dispersion_{b,t}$, which is the standard deviation of the price deviations, $PD_{f,b,t}$, for bond b reported by all ETFs on day t. $PD\ Dispersion_{b,t}$ varies across bonds and days, with a sample mean (median) of 59 (51) bps, and the 10th and 90th percentiles are 12 and 117 bps, respectively. Figure 1 presents the weekly time series of the median of the $PD\ Dispersion_{b,t}$ throughout our sample period. Panel A of Figure 1 indicates that median price dispersion across ETFs typically fluctuates around 50 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.[14] During this period, price dispersion spiked to over 200 basis points. In addition, in line with the greater deterioration of liquidity in the IG sector,[15] Panel B illustrates that while price dispersion for IG bonds is generally about 30% lower than that for HY bonds during normal market conditions, it experienced a sudden surge, exceeding the price dispersion for HY bonds by approximately 43% 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 obtained 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 construct a bond-day level sample and estimate the following empirical model:

[14] See for example, O'Hara and Zhou (2021), Kargar, Lester, Lindsay, Liu, Weill, and Zúñiga (2021), Boyarchenko, Kovner, and Shachar (2022).
[15] See for example, Haddad, Moreira, and Muir (2021), and Ma, Xiao, and Zeng (2022).

$$PD\ Dispersion_{b,t} = \beta_1 \times Traded\ Dummy_{b,t} + \beta_2 \times Age_{b,t} + \beta_3 \times Issue\ Size_{b,t} + \beta_4 \times Credit\ Rating_{b,t} + \beta_5 \times Time\ to\ Maturity_{b,t} + \beta_6 \times Log(Number\ of\ ETF\ holders)_{b,t} + \mu_i + \mu_t + \varepsilon_{b,t}, \quad (2)$$

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.[16] We also control for a bond's credit rating, time to maturity, and the number of ETFs holding the bond. 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 greater liquidity leading to smaller price dispersions, Column (1) of Panel A, Table 5 shows that price dispersion among ETFs declines when a bond gets traded, and it tends to be smaller for younger bonds and for larger issues. In addition, lower rated, longer term bonds and those held by more ETFs 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 $Traded\ Dummy$ with the logarithm of 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 interquartile range of prices across ETFs to capture price dispersion (Panel B, Table 5). [17]

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 an ETF's creation basket based on changes in its holdings. Specifically, on days with non-zero ETF flows, a positional increase in a bond's holdings 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

[16] See for example, Hotchkiss and Jostova (2017).
[17] Trading frequency absorbs part of the liquidity effect associated with issue size, making the coefficient for issue size insignificant.

creation basket for a given ETF on a specific day. $Bond\ Creation_{f,b,t}$ equals one if the number of shares of bond b held by ETF f increases on day t, and zero otherwise. As shown in Table 6, 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 8.7%. In contrast, on days with net outflows, the probability of a bond being included in the redemption basket rises to 20.2%, suggesting that redemption baskets typically contain more than twice the number of bonds compared to creation baskets. [18]

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

$$Bond\ Creation_{f,b,t} = \beta \times \left(PD_{f,b,t} - \overline{PD_{f,t}}\right) + \varepsilon_{f,b,t}, \qquad (3)$$

where $PD_{f,b,t}$ is defined as in model (1), and $\overline{PD_{f,t}}$ represents the average $PD_{f,b,t}$ estimated at the fund-day level, and captures the ETF adjustment as illustrated in the model. The difference between $PD_{f,b,t}$ and $\overline{PD_{f,t}}$ captures how a fund values a bond on a given day relative to its peers, after controlling for its overall valuation tendencies. These tendencies could, for instance, arise from fund-specific differences in pricing practices, such as some funds consistently using bid prices to report the value of their holdings, while others relying on mid prices.[19] Standard errors are double-clustered at the date and fund levels.

If a bond is more likely to be included in the creation basket of an ETF that values it higher—after controlling for the ETF's overall valuation tendency—we would expect a positive β. The results in Table 7 strongly support this hypothesis. Specifically, Column (1) shows that the coefficient for the PD measure is both positive and highly significant, indicating that a higher

[18] 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). As shown in Table 6, there is a 1.7% (3.0%) probability that a bond being in 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.

[19] See for example, Cici, Gibson, and Merrick (2011).

adjusted valuation of a bond by an ETF increases the probability of its inclusion in the ETF's creation basket.

Next, we 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 sizable AP creation orders, which may significantly increase the probability of creation basket inclusion for all bonds in the ETF portfolio. Conversely, on days when the ETF does not receive any AP creation orders, the probability of creation basket inclusion for any bond would be zero. By incorporating fund-date fixed effects, we can account for these potential biases. More overall, adding fund-date fixed effects naturally 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 $\left(PD_{f,b,t} - \overline{PD_{f,t}}\right)$ 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 timing 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. Controlling for time-varying bond characteristics does not materially alter our conclusion (Column (3)). This finding holds even when we further include fund-bond fixed effects to account for potential ETF preferences for different bonds (Column (4)).

To assess the economic magnitude of the effect of ETFs' adjusted bond valuation on APs' creation basket decisions, we introduce a rank-based measure derived from the PD measure. Specifically, for bond b on day t, we first assign a percentile rank to each ETF based on the price it assigns to bond b. We then create three dummy variables for ETFs falling within the following rank ranges: 50%–70%, 70%–90%, and above 90%. ETFs in any of these ranges assign a higher valuation to the bond than the median ETF, with those in the top 10% (above 90%) attaching highest valuations.

We then re-estimate model (3) by replacing the PD measure with these three dummy variables. Column (5) of Table 7 shows a convex relationship between price deviation and the likelihood of a bond's inclusion in a creation basket. The coefficient of the dummy variable for the 50% – 70% range is 0.0006 and statistically insignificant, while the coefficients of the dummies for the 70% – 90% and above 90% ranges are 0.036 and 0.165, respectively, and both are statistically significant. This result suggests that, compared to the baseline group, where the assigned price for bond b does not exceed the median price assigned by all ETFs, there is no significant difference in the probability of inclusion when the rank of PD falls between 50^{th} and 70^{th} percentiles. However, APs tend to focus on ETFs that assign higher relative prices to a bond, often selecting those ETFs that rank among the very top in terms of valuation. Specifically, when a bond's price rank is between the 70^{th} and 90^{th} percentiles, its probability of inclusion in a creation basket increases by 3.6%. Additionally, if the bond's price deviation ranks in the top decile, the probability of inclusion is 16.5% higher compared to the baseline group. These probabilities are considerable given that unconditionally, a bond's probability of being included in a creation basket is only 7.1% (Table 6). In sum, these results show that when the bond's price deviation is below or close to the 70^{th} percentile, its likelihood of being included in the creation basket is smaller; and this likelihood increases significantly with the degree of positivity in the price deviation.

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. Table 8 reports the distribution of coefficients for price deviation across a sample of 60 ETFs. Of these, 42 ETFs exhibit a positive coefficient, indicating a positive association between the ETF price deviation and the likelihood of creation basket inclusion. Among these 42 ETFs, 31 (21) show statistical significance at the 10% (1%) level, and 11 are statistically insignificant. Conversely, 18 ETFs have a negative coefficient, with 3 (1) showing statistical significance at the 10% (1%) level and the remaining 15 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 0.67, while the mean for the negative-coefficient ETFs is only -0.26. These results suggest that the relationship between ETFs' adjusted bond valuation differences and the inclusion in creation baskets, as documented earlier, is present for the majority of ETFs in the industry.

6. Decomposition of Price Deviation: A Case Study of LQD

What drives valuation dispersions across ETFs? Each corporate bond ETF in our sample tracks a certain index, whose provider typically supplies indicative or end-of-day pricing for the bonds that make up the index. These prices can be used by ETF sponsors to calculate the NAV of the ETF. However, if certain bonds in the portfolio do not have reliable or recent prices (due to illiquidity or infrequent trading), the ETF may use pricing from other sources, such as pricing services or dealer quotes, or rely on their own proprietary valuation models to determine their value.

To understand how index providers and ETF sponsors each contribute to the valuation dispersion across ETFs, we decompose an ETF's reported bond price deviation from the median into two components:

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

where $P_{p,b,t}$ refers to the price reported by the tracking index provider for bond b on day t. The first term reflects how an ETF sponsor 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 other ETFs (i.e., a passive component). If the ETF sponsor 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—specifically, the index provider's reported price bias relative to the market.

We obtain daily constituent and pricing information for the iBoxx Liquid Investment Grade (USD) Index (hereafter, iBoxx IG Index) from Markit Indices GmbH, a subsidiary of S&P Dow Jones Indices LLC. This index is tracked by iShares iBoxx $ Investment Grade Corporate Bond ETF (ticker symbol: LQD), the most liquid corporate bond ETF as measured by trading volume, and provides a broad coverage of liquid USD-denominated investment-grade corporate bonds.[20]

[20] Daily constituent-level data for indexes tracked by fixed-income ETFs is generally not accessible to academic institutions. This data limitation restricts our ability to construct a more comprehensive sample for decomposing the price deviation measure, as daily constituent-level data for other indexes must be obtained directly from their respective providers.

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 bonds and closely tracks the iBoxx index, with a mean tracking difference of just -0.07 bps and a standard deviation of 1.06 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 9).

While LQD tends to value bonds higher compared to its peers, with the mean (median) price deviation being 28 (13) bps, there are considerable variations in its variable deviation (Panel B of Table 9). The standard deviation of price deviation is 62 bps, and more than one quarter of the reported price deviations are negative. The magnitude of the dispersion of LQD's price deviation is in the same order as the full-sample price dispersion (see Table 4).

Panel B of Table 9 shows that most of LQD's valuation deviation from its peers is primarily driven by the passive component.[21] This component has a mean of 28 bps and a standard deviation of 62 bps, which align closely with the overall price deviation. In 67% of cases, iBoxx-reported bond prices are higher than the median reported prices across ETFs. In contrast, the active deviation component—the difference in reported prices between LQD and iBoxx—is generally close to zero, with a mean of just 0.58 bps and falling within a narrow range from the 1st to the 75th percentile. Interestingly, when LQD does adjust the prices reported by iBoxx, it consistently adjusts them upward. There is an 11% (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.

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 9 shows that, on net-creation days, an average of 22% of the bonds are

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

included in the creation baskets, while on net-redemption days, an average of 42% are in the redemption baskets. On average, the probability of a bond being included in a creation basket is 16%. These figures are notably higher than the overall sample averages, reflecting LQD's greater liquidity due to its frequent creation and redemption activities.

We begin by re-estimating model (3) 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 credit ratings (categorized into broad rating groups: AAA, AA, A, and BBB) and for time to maturity, which is divided into three groups: short-term (1–10 years), medium-term (10–20 years), and long-term (20–30 years). Column (1) of Table 10 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 $PD_{f,b,t}$ 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 16.4% in Column (3) to 50.1% in Column (4).

We also create dummy variables for each of the two components to better assess their individual impacts on *BondCreation*. Specifically, the *Passive Dummy* 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* 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 2.3% 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 66.4% increase in *BondCreation*. This finding suggests that while LQD does not frequently adjust its price relative to the iBoxx Index by a large amount, when it does—even by a small margin—it has a substantial impact on the likelihood of a bond being included in LQD's creation basket.

An interesting question is what motivates ETF sponsors 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 empirical model:

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

where $w_{LQD,b,t-1}$ ($w_{iBoxx,b,t-1}$) 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 11 support our hypothesis. Column (1) shows that the coefficient for the weighted difference between LQD and iBoxx from the previous day is negatively related to the active component of the price deviation (PD) measure. 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 concentrated in these underweighted bonds. iBoxx IG Index exclusively includes bonds with at least three years expected remaining life.[22] 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

[22] 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.

unlikely to occur. We re-estimate Model (5) for bond subsamples based on a maturity cutoff of 3.5 years. Columns (2) and (3) show that the negative association between bond underweighting and the next day active component of the price deviation is not significant for those bonds with a remaining time of maturity less than 3.5 years. In contrast, the negative relationship is even stronger for bonds with more than 3.5 years to mature.

To assess the economic significance, we create a dummy variable, *Underweight Dummy*, which takes the value of one if $w_{LQD,b,t-1}$ is less than $w_{iBoxx,b,t-1}$, and zero otherwise. We then re-estimate model X, replacing $P_{LQD,b,t} - P_{iBoxx,b,t}$ with the Active Price Deviation Dummy, and $\left(w_{LQD,b,t-1} - w_{iBoxx,b,t-1}\right)$ 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 3.28%. This represents 30% 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 remaining life, Column (6) shows that the probability of creation basket inclusion increases by 3.98%. 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. The Ultimate Impact of Valuation Dispersions on Corporate 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 receive more ETF shares 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 explore these questions in more detail.

7.1 Effect on Fund Flows

As discussed earlier, for an ETF to receive a bond through the creation process, it must assign a higher value to the bond than its peers, while maintaining a relatively conservative overall bond valuation. This requires the ETF to assign lower prices to some bonds in its portfolio, resulting in a significant dispersion in its bond price differential (PD) measure. Based on this, we

hypothesize that the greater the cross-bond dispersion in an ETF's PD measures, the more likely that the ETF will attract creation from APs.

To test this hypothesis, we estimate the following empirical model:

$$Fund\ Creation_{f,t} = \beta \times PD\ Dispersion_{f,t} + \mu_f + \mu_t + \varepsilon_{f,t}\ , \tag{6}$$

where $Fund\ Creation_{f,t}$ is a dummy that takes the value of one if fund f experiences creation on day t, and zero otherwise. $PD\ Dispersion_{f,t}$ is the standard deviation of the PD measure across bonds for fund f on day t. A higher value of this measure indicates a greater dispersion in bond valuations within the fund, relative to its peers. We include fund fixed effects (μ_f) and day 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.

Table 12 provides strong support for our hypothesis. The coefficient of $PD\ Dispersion_{f,t}$ is positive and highly significant, suggesting that an ETF is more likely to experience creation when its PD measure shows greater dispersion across bonds. Moreover, the effect is economically meaningful. The point estimate of the coefficient indicates that one standard deviation increase in $PD\ Dispersion_{f,t}$ results in a 1.8% higher likelihood of creations (0.035×0.5087). This economic significance is sizable especially considering that, on average, funds experience creation on only 17% of days.

To address potential concerns that some time-varying fund characteristics may also affect $Fund\ Creation_{f,t}$, we include some fund-level variables that vary over time. Specifically, we control for liquidity effects by including the log of ETF size, and for potential arbitrage driven creations by including the ETF premium. Column (2) shows that larger ETFs and those trading at a premium are indeed more likely to experience creation. More importantly, the coefficient on $PD\ Dispersion_{f,t}$ remains positive and highly significant, with only a small change in magnitude. Since different bonds within an ETF's portfolio may carry different weights, we re-estimate $PD\ Dispersion_{f,t}$ using a value-weighted standard deviation of the PD measure across bonds, with each bond's par amount held by the ETF serving as its weight. Columns (3) and (4) confirm that our results hold. Taken together, these findings provide further support for our hypothesis that

ETFs receive creation when assigning high value to some bonds while assigning low value to others.

7.2 Wealth effect

If an ETF receives a creation basket that is more concentrated in bonds it values higher than its peers, the value of existing ETF holders could potentially be diluted; in other words, existing holders would receive overpriced securities relative to their other holdings. To gauge the potential dilution effect, we estimate a dilution measure for ETF *f* on a given day *t* as follows:

$$Dilution_{f,t} = \sum_b \left(\Delta Shares^{+}_{f,b,t} \times \left(PD_{f,b,t} - \overline{PD_{f,t}} \right) \right), \quad (7)$$

where $\Delta Shares^{+}_{f,b,t}$ is the number of bond b in the creation basket for fund f on day t, and $PD_{f,b,t} - \overline{PD_{f,t}}$ represents the estimated dollar amount of value dilution per bond. Specifically, assuming that an ETF's overall valuation strategy is fair, $PD_{f,b,t}$ captures the potential loss to existing ETF investors (or gains to APs) when one additional unit of bond b is included in the creation basket. The term $\overline{PD_{f,t}}$ adjusts for the ETF's overall valuation tendency.

For each day t, we calculate $Dilution_{f,t}$ for each fund f and then sum across all funds to obtain the daily dollar amount of dilution for the corporate bond ETF market. We then compute the daily dilution percentage by dividing this dollar amount by the total market value of corporate bonds held by all ETFs on that day. Finally, we aggregate daily dilution percentages to generate annual figures from 2017 to 2021.

Table 13 reports the average annual dilution percentage and compares it to the ETF expense ratios. For the full sample of 60 ETFs (Column (1)), the average annual dilution is approximately 6.5 bps, or about 40% of the average expense ratio of 16.8 bps. Columns (2)–(4) focus on subsamples of ETFs that show a positive association between ETF price deviation and the likelihood of creation basket inclusion. For the 42 ETFs exhibiting this positive association, the average annual dilution increases to about 10.9 bps. The stronger the association between price deviation and creation basket inclusion, the more pronounced the dilution effects. In the subsample of 21 ETFs that exhibit the strongest association (p-value < 1%), the average annual dilution rises

to 11.3 bps, which is equivalent to approximately 80% of the average expense ratio of these funds. These ETFs are large, collectively holding $97 billion in corporate bonds. After accounting for the size of these funds, the average annualized dilution amounts to over $114 million. Together, our results indicate that while APs benefit from selectively delivering bonds to ETFs that value them the highest, this practice results in significant losses for existing ETF investors. The magnitude of dilution through this mechanism is comparable to the expense fees that ETF investors pay.

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. This behavior, while advantageous for APs, comes at the expense of existing ETF investors, leading to a dilution of their holdings that mirrors the magnitude of the funds' expense ratios.

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.

The implications of these findings are significant for both ETF issuers and investors. For ETF sponsors, the need to balance valuation discrepancies across their holdings is crucial, as these discrepancies affect not only their creation processes but also the fund's attractiveness to APs. At the same time, ETFs that exhibit high valuation dispersion may inadvertently encourage strategic behavior by APs, which can lead to unintended consequences for fund performance and investor wealth. For investors, understanding the dynamics of valuation dispersion and its impact on ETF

creation can help them better assess the risks associated with investing in corporate bond ETFs, particularly in periods of market stress, such as during the COVID-19 pandemic.

In conclusion, 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 investor 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

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*(3), 695-731.

Choi, J., Kronlund, M., & Oh, J. Y. J. (2024). Silent swing: Do open-end funds tilt holdings valuations in response to flows? *Northern Finance Association Annual Meeting*

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*(1), 206-226.

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*.

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*(11), 5309-5351.

Hotchkiss, E., & Jostova, G. (2017). Determinants of corporate bond trading: A comprehensive analysis. *Quarterly Journal of Finance*, *7*(02), 1750003.

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*(11), 5352-5401.

Koont, N., Ma, Y., Pastor, L., & Zeng, Y. (2024). Steering a ship in illiquid waters: Active management of passive funds. *Review of Financial Studies. Forthcoming*

Ma, Y., Xiao, K., & Zeng, Y. (2022). Mutual fund liquidity transformation and reverse flight to liquidity. *Review of Financial Studies*, *35*(10), 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*(1), 46-68.

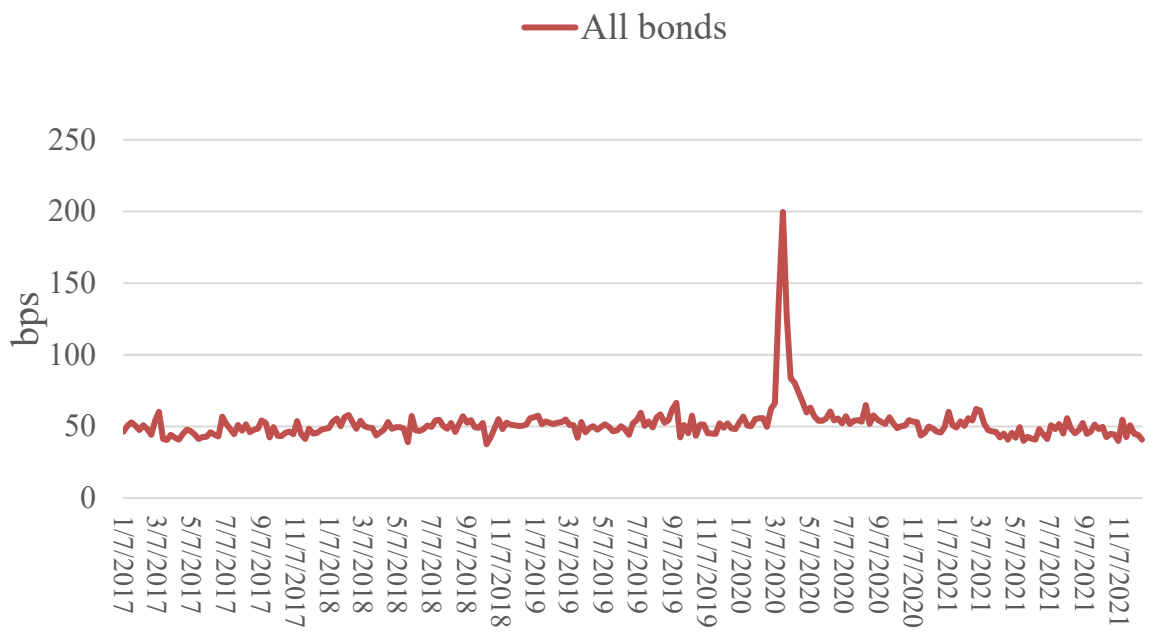
Pan, K., & Zeng, Y. (2017). ETF arbitrage under liquidity mismatch. *Available at SSRN 3723406*.

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

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*.

Panel A: Weekly Median of PD Dispersion$_{b,t}$



Panel B: Weekly Median of PD Dispersion$_{b,t}$ for IG and HY Bonds

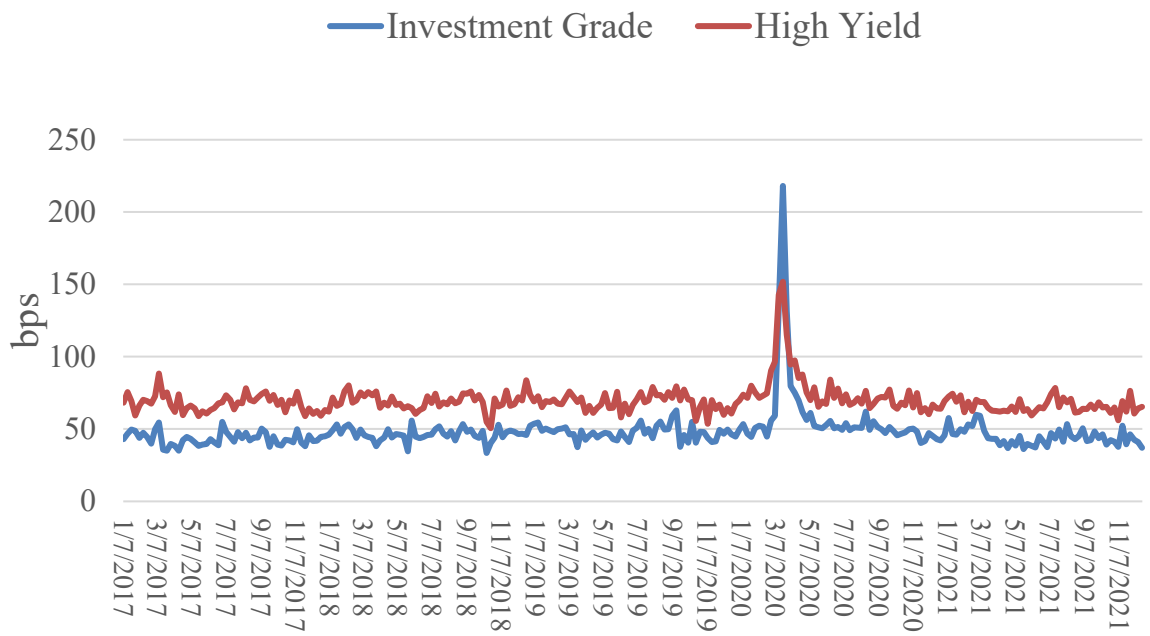


Figure 1 Price dispersion over time. This figure displays the weekly time series of price dispersion over our sample period. First, the standard deviation (S.D.) of the price deviations is calculated at the bond-day level, using the price deviations $PD_{f,b,t}$ by all ETFs for bond b on day t. Then, we compute the weekly median of these S.D. and plot the weekly time series in Panel A. Panel B shows time series for investment grade bonds and high yield bonds, respectively.

Table 1: An example of a Bond Held by Multiple ETFs

This table presents holding information from Morningstar for a bond issued by AT&T Inc (CUSIP: 00206RJY9), with a maturity date of 01/03/2031. As of 07/06/2020, the bond holds a Baa2 rating from Moody's and is included in the portfolios of the following eleven ETFs. These ETFs report different valuations for the bond. ETF daily flows are calculated using the daily number of shares outstanding from Bloomberg.

ETF Name	Ticker	Par Value ($)	Par Value Change	Bond Price	ETF Daily Flow ($Million)
SPDR® Portfolio Aggregate Bond ETF	SPAB	500,000	0	105.24	15.4
SPDR® Portfolio Corporate Bond ETF	SPBO	100,000	0	105.24	3.5
Schwab US Aggregate Bond ETF™	SCHZ	1,200,000	0	105.39	0.0
FlexShares Credit-Scored US Corp Bd ETF	SKOR	140,000	0	105.53	0.0
American Century Dvrs Corp Bd ETF	KORP	409,000	0	105.55	0.0
iShares Broad USD Invm Grd Corp Bd ETF	USIG	1,050,000	0	105.57	3.0
iShares Core Total USD Bond Market ETF	IUSB	315,000	0	105.68	10.8
iShares Core US Aggregate Bond ETF	AGG	10,551,000	0	105.68	141.6
iShares Government/Credit Bond ETF	GBF	200,000	0	105.68	0.0
Goldman Sachs Acss Invmt Grd Corp Bd ETF	GIGB	470,000	0	105.72	0.0
iShares iBoxx $ Invmt Grade Corp Bd ETF	LQD	41,241,000	2,110,000	105.83	404.1

Table 2: A Toy Model—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, the number of shares created using one unit of Bond 1 is determined by the ratio of the reported price of Bond 1 to the ETF NAV, with a higher number of shares preferred by the AP. We calculate the adjusted price deviation for Bond 1 by first subtracting the report prices of Bond 1 by each ETF from the median price of Bond 1 across all ETFs. This price deviation is then adjusted by the average price deviation for Bond 1 and Bond 2 within each ETF. We show that the adjusted price deviation for Bond 1 is positively associated with the number of ETFs created using one unit of Bond 1.

	ETF A	**ETF B**	**ETF C**	**ETF D**
	Reported constituent prices:			
Bond 1	102	105	105	105
Bond 2	100	100	103	97
ETF NAV	202	205	208	202
Shares of ETFs created using one unit of Bond 1	0.505	0.512	0.505	0.520
	Adjusted Price Deviation for Bond 1:			
Price Deviation (PD): Bond 1 price - Median Bond 1 price	-3.0	0.0	0.0	0.0
ETF Adjustment: Average(Bond price - Median bond price)	-1.5	0.0	1.5	-1.5
Price Deviation (PD) - ETF Adjustment	-1.5	0	-1.5	1.5

Table 3: 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 of the bond-day level. A bond's credit rating at the time of trading is a 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 Online Appendix for more variable definitions.

Panel A: Fund-day level (N=75,354)

	Mean	Median	S.D.	P1	P10	Q1	Q3	P90	P99
ETF Premium (Price/NAV) (%)	0.08	0.08	0.32	-1.53	-0.20	-0.04	0.23	0.41	0.97
Fund Size ($1,000,000)	3,018	220	8,838	5	33	87	1,427	6,043	56,931

Panel B: Bond-day level (N=9,178,366)

	Mean	Median	S.D.	P1	P10	Q1	Q3	P90	P99
Credit Rating	3.93	4.00	1.10	2	3	3	4	6	7
Time to maturity (years)	10.22	6.55	8.98	1.0	2.1	3.8	15.5	26.0	36.4
Age (years)	4.16	3.20	3.78	0.1	0.5	1.4	5.8	8.6	19.2
Offering amount ($1000)	872	650	739	250	350	450	1,000	1,700	3,500

Table 4: Summary Statistics of Bond Price Dispersion

This table presents the summary statistics of price deviation ($PD_{f,b,t}$): $PD_{f,b,t} = P_{f,b,t} - \widetilde{P_{g,b,t}}$, where $P_{f,b,t}$ is the price for bond b reported by ETF f on day t, and $\widetilde{P_{g,b,t}}$ is the median price of bond b reported by any ETF g that holds the bond on day t. Panel A reports the summary statistics of $PD_{f,b,t}$ at the fund-bond-day level. Panel B reports price dispersion, $PD\ Dispersion_{b,t}$, at the bond-day level, which is measured as the standard deviation (and the interquartile range) of the $PD_{f,b,t}$ across funds for bond *b* on day *t*. See the Online Appendix for more variable definitions.

Panel A: Fund-bond-day level

	N	Mean	Median	S.D.	Q3-Q1	P1	P10	Q1	Q3	P90	P99
Price Deviation (bps): $PD_{f,b,t} = P_{f,b,t} - \widetilde{P_{g,b,t}}$											
All	32,946,739	-6.37	0.00	76.8	41.5	-263.2	-97.8	-26.0	15.5	74.3	229.6
Investment Grade	24,911,030	-8.11	0.00	72.2	38.3	-263.2	-96.5	-24.5	13.9	64.8	192.8
High Yield	6,146,309	-1.06	0.00	94.2	58.9	-263.2	-109.2	-33.9	25.0	125.0	230.4

Panel B: Bond-day level

	N	Mean	Median	S.D.	Q3-Q1	P1	P10	Q1	Q3	P90	P99
Number of ETF holders	9,178,366	8.34	8.00	3.80	6.0	3	4	5	11	14	18
$PD\ Dispersion_{b,t}$	9,178,366	58.71	50.91	40.90	57.2	1.13	12.42	26.20	83.42	116.64	192.95
Number of TRACE transactions (t)	9,178,366	4.78	2	9.8	6	0	0	0	6	13	37
Number of TRACE transactions (t) > 0	9,178,366	0.64	1	0.5	1	0	0	0	1	1	1

Table 5: Determinants of Bond Price Dispersion

This table presents the relationship between cross-ETF bond price dispersion and bond liquidity. The dependent variable in Panel A is the standard deviation of prices for the same bond on the same day across different ETFs, while in Panel B, it is the interquartile range (Q3-Q1) of these prices. Independent variables include liquidity proxies, such as the number of TRACE transactions, and various bond characteristics. 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: Dependent variable is PD $Dispersion_{b,t}$

	(1)	(2)	(3)	(4)
Trade Dummy: Number of TRACE transaction (t) > 0	-2.7937***	-3.2227***		
	(-18.90)	(-20.50)		
log(1+Number of TRACE transaction (t))			-1.9733***	-2.2377***
			(-22.20)	(-23.26)
Time to Maturity (years)	0.9427***	0.9508***	0.9183***	0.9228***
	(41.29)	(42.38)	(40.51)	(41.46)
Age (years)	2.4472***	2.4350***	2.4372***	2.4232***
	(59.89)	(59.93)	(60.13)	(60.11)
log(Issue Size)	-2.1859***	-2.3278***	-1.4815***	-1.5205***
	(-6.92)	(-7.42)	(-4.71)	(-4.88)
Credit Rating	2.7674***	1.9103***	2.8566***	2.0774***
	(11.02)	(4.83)	(11.42)	(5.28)
log(Number of ETF Holders)	5.9278***	6.2160***	6.0053***	6.2947***
	(18.26)	(17.93)	(18.51)	(18.17)
Observations	9,178,366	8,253,064	9,178,366	8,253,064
Issuer FE	Y		Y	
Date FE	Y		Y	
Issuer × Date FE		Y		Y
Adj R-squared	0.285	0.369	0.285	0.370

Panel B: Dependent variable is the interquartile range (Q3-Q1) of $PD_{f,b,t}$ across funds for bond b on day t.

	(1)	(2)	(3)	(4)
Trade Dummy: Number of TRACE transaction (t) > 0	-3.5017***	-3.8585***		
	(-11.17)	(-11.11)		
log(1+Number of TRACE transaction (t))			-2.4802***	-2.6165***
			(-13.37)	(-12.71)
Time to Maturity (years)	1.0248***	1.0381***	0.9941***	1.0060***
	(23.56)	(24.21)	(22.89)	(23.47)
Age (years)	3.4919***	3.4478***	3.4792***	3.4349***
	(35.91)	(34.86)	(35.84)	(34.76)
log(Issue Size)	-1.5203**	-1.7528**	-0.6337	-0.8210
	(-2.26)	(-2.58)	(-0.94)	(-1.20)
Credit Rating	5.2368***	4.3795***	5.3490***	4.5701***
	(9.26)	(4.79)	(9.48)	(5.01)
log(Number of ETF Holders)	19.0534***	19.8580***	19.1516***	19.9429***
	(28.37)	(26.85)	(28.47)	(26.92)
Observations	9,178,366	8,253,064	9,178,366	8,253,064
Issuer FE	Y		Y	
Date FE	Y		Y	
Issuer × Date FE		Y		Y
Adj R-squared	0.214	0.280	0.214	0.281

Table 6: 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 with non-zero flows. ETFs' creation and redemption baskets are identified based on daily changes in ETF holdings in Morningstar. The variable $Bond\ Creation_{f,b,t}$ ($Bond\ Redemption_{f,b,t}$, respectively) is a fund-bond-day level dummy that equals one if the number of shares of bond b held by ETF f increases (decreases, respectively) 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 $Bond\ Creation_{f,b,t}$ and $Bond\ Redemption_{f,b,t}$ along with the number of fund-bond-day observations for days with positive, negative, and non-zero ETF net flows.

Days when ETF flow is:	Positive (Net Creation)	Negative (Net Redemption)	Non-zero Flow (Both)
	$Bond\ Creation_{f,b,t}$: Creation =1; Otherwise=0		
Mean	8.7%	3.0%	7.1%
	$Bond\ Redemption_{f,b,t}$: Redemption =1; Otherwise=0		
Mean	1.7%	20.2%	6.7%
	Number of Observations (Fund-Bond-Day)		
N	24,095,515	8,851,224	32,946,739

Table 7: 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 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* is the fund-bond-date level price deviation measure ($PD_{f,b,t}$): $PD_{f,b,t} = P_{f,b,t} - \widetilde{P_{g,b,t}}$, where $P_{f,b,t}$ is the price for bond b reported by ETF f on day t, and $\widetilde{P_{g,b,t}}$ is the median price of bond b reported by any ETF g that holds the bond on day t. $\overline{PD_{f,t}}$ is the fund-date level price deviation measure, estimated by averaging $PD_{f,b,t}$ across bonds. Column (4) utilizes 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 three dummy variables for ETFs falling within the following rank ranges: 50%–70%, 70%–90%, and above 90%. 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)
Dep. Var.	*BondCreation* (Creation =1; Otherwise=0)				
Price Deviation - $\overline{PD}$	0.6273***				
	(3.07)				
Price Deviation		0.6353***	0.8160***	0.9284***	
		(3.17)	(3.60)	(3.96)	
Rank(Price Deviation) in (50, 70]					0.0006
					(0.22)
Rank(Price Deviation) in (70, 90]					0.0364***
					(10.21)
Rank(Price Deviation)>90					0.1650***
					(8.64)
Observations	32,969,730	32,969,719	31,729,376	31,723,837	31,723,837
Fund × date FE		Y	Y	Y	Y
Bond × date FE			Y	Y	Y
Fund × bond FE				Y	Y
Adj R-squared	0.000	0.173	0.193	0.206	0.222

Table 8: 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: $Bond\ Creation_{f,b,t} = \beta_f \times PD_{f,b,t} + \mu_{f,t} + \mu_{f,b} + \varepsilon_{f,b,t}$, where 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-date level. Each regression includes date fixed effects and bond fixed effects. Standard errors are two-way clustering adjusted at the date and bond levels. Column (1) reports the distribution of the estimated coefficients β_f across our sample of 60 ETFs. Column (2) focuses on ETFs with a positive β_f. Column (3) provides results for ETFs where β_f is positive and statistically significant, with 3 ETFs significant at the 10% level, 7 at the 5% level, and 21 at the 1% level. Column (4) shows ETFs with a positive but statistically insignificant β_f. Column (5) reports the results for the 18 ETFs where the β_f estimate is negative, among which 2 ETFs show significance at the 10% level and 1 at the 1% level.

	(1)	(2)	(3)	(4)	(5)
		$\beta > 0$			$\beta < 0$
Summary statistics of β_f	All ETFs	All	p-value<10%	p-value>10%	All
N	60	42	31	11	18
Mean	0.39	0.67	0.83	0.19	-0.26
S.D.	0.74	0.71	0.76	0.15	0.25
Percentiles:					
P10	-0.27	0.08	0.11	0.06	-0.51
P25	-0.09	0.15	0.26	0.08	-0.37
P50	0.16	0.43	0.66	0.15	-0.17
P75	0.70	0.81	0.99	0.27	-0.11
P90	1.37	2.12	2.23	0.42	-0.06

Table 9: 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. Panel B reports the bond-day level distribution of the overall price deviation, $PD_{f,b,t} = P_{LQD,b,t} - \widetilde{P_{g,b,t}}$, where $P_{LQD,b,t}$ is the price of bond b reported by LQD on day t and $\widetilde{P_{g,b,t}}$ is the median price of the bond b reported by all other ETFs on day t. We then decompose the overall price deviation into the active component and passive component using the index price $P_{iBoxx,b,t}$. 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=1,465,726)

	Mean	**Median**	**S.D.**	**P1**	**P10**	**P25**	**P75**	**P90**	**P99**
$w_{LQD,b,t-1}$ (bps)	4.50	3.74	2.76	0.49	1.95	2.67	5.60	8.26	15.25
$w_{iBoxx,b,t-1}$ (bps)	4.58	3.64	2.63	1.96	2.33	2.79	5.52	8.17	15.59
$w_{LQD,b,t-1} - w_{iBoxx,b,t-1}$ (bps)	-0.07	0.08	1.06	-3.61	-1.43	-0.50	0.44	1.03	2.74
Time to Maturity:									
<10 Year	0.59	1	0.49	0	0	0	1	1	1
10-20 Year	0.12	0	0.33	0	0	0	0	1	1
>20 Year	0.26	0	0.44	0	0	0	1	1	1
Credit Rating	3.40	4	0.71	1	3	3	4	4	4

Panel B: Overall price deviation and decomposition at the Bond-day level (N=1,465,726)

	Mean	**Median**	**S.D.**	**P1**	**P10**	**P25**	**P75**	**P90**	**P99**
Overall price deviation (PD)									
PD: $P_{LQD,b,t} - \widetilde{P_{g,b,t}}(\%)$	0.28	0.13	0.62	-1.47	-0.30	-0.05	0.57	1.14	2.29
Decomposition using data from iBoxx									
Active Deviation:									
PD: $P_{LQD,b,t} - P_{iBoxx,b,t}(\%)$	0.0058	0.0000	0.0129	0.000	0.000	0.000	0.000	0.022	0.067
Dummy: $P_{LQD,b,t} - P_{iBoxx,b,t} > 2bps$	0.11	0	0.31	0	0	0	0	1	1
Passive Deviation:									
PD: $P_{iBoxx,b,t} - \widetilde{P_{g,b,t}}(\%)$	0.28	0.13	0.62	-1.42	-0.30	-0.06	0.57	1.13	2.28
Dummy: $P_{iBoxx,b,t} > \widetilde{P_{g,b,t}}$	0.67	1	0.47	0	0	0	1	1	1

Panel C: BondCreation and BondRedemption Dummies

Days when ETF flow is:	Positive (Net Creation)	Negative (Net Redemption)	Non-zero Flow (Both)
	$Bond\ Creation_{LQD,b,t}$: Creation =1; Otherwise=0		
Mean	22.2%	5.7%	15.7%
	$Bond\ Redemption_{LQD,b,t}$: Redemption =1; Otherwise=0		
Mean	5.5%	42.1%	19.9%
	Number of Observations (Fund-Bond-Day)		
N	890,636	575,166	1,465,802

Table 10: 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 dependent variable is $Bond\ Creation_{LQD,b,t}$, 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 $PD_{f,b,t} = P_{LQD,b,t} - \widetilde{P_{g,b,t}}$, where $P_{LQD,b,t}$ is the price of bond b reported by LQD on day t and $\widetilde{P_{g,b,t}}$ is the median price of the bond b reported by all other ETFs on day t. Additionally, we decompose this price deviation into two components: an active component by the ETF manager, defined as $P_{LQD,b,t} - P_{iBoxx,b,t}$, and a passive component, defined as $P_{iBoxx,b,t} - \widetilde{P_{g,b,t}}$. We focus on the sample where the LQD flow is non-zero and at least three ETFs hold the bond on the given day. 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)
Dep. Var.	*BondCreation* (Creation =1; Otherwise=0)				
Price Deviation ($PD_{LQD,b,t}$)	2.4506***				
	(15.76)				
Active Deviation: $P_{LQD,b,t} - P_{iBoxx,b,t}$		1,864.4***		1,862.9***	
		(37.49)		(37.50)	
Passive Deviation: $P_{iBoxx,b,t} - \widetilde{P_{g,b,t}}$			2.3311***	1.4689***	
			(15.68)	(10.99)	
Active Dummy: $P_{LQD,b,t} - P_{iBoxx,b,t} > 2bps$					0.6640***
					(56.32)
Passive Dummy: $P_{iBoxx,b,t} > \widetilde{P_{g,b,t}}$					0.0233***
					(15.59)
Observations	1,465,726	1,465,726	1,465,726	1,465,726	1,465,726
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	0.164	0.501	0.164	0.501	0.431

Table 11: 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 the *Active Deviation:* $P_{LQD,b,t} - P_{iBoxx,b,t}$ and associated dummy variables. 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. 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)
Sample time to maturity (years)	All	<3.5	>3.5	All	<3.5	>3.5
Dep. Var.	Active Deviation: $P_{LQD,b,t} - P_{iBoxx,b,t}$			Active Dummy: $P_{LQD,b,t} - P_{iBoxx,b,t} > 2bps$		
$w_{LQD,b,t-1} - w_{iBoxx,b,t-1}$	-0.0352***	0.0019	-0.0608***			
	(-8.16)	(0.23)	(-13.27)			
Underweight Dummy				0.0328***	0.0018	0.0398***
				(15.73)	(0.61)	(18.40)
Observations	1,465,726	83,422	1,382,299	1,465,726	83,422	1,382,299
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	0.232	0.254	0.236	0.193	0.180	0.198

Table 12: Bond Price Dispersion and Likelihood of Fund Creations

This table examines the relationship between with-ETF relative bond price dispersion and a fund's likelihood of receiving creations. The dependent variable is $Fund\ Creation_{f,t}$, a dummy variable equal to one if the number of shares outstanding of the ETF increases on day t, and zero otherwise. The key fund-day level independent variable, $PD\ Dispersion_{f,t}$, is the standard deviation of price deviation $PD_{f,b,t}$ for all bonds in ETF f on day t. 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)
Dep. Var.	$Fund\ Creation_{f,t}$(Creation =1; Otherwise=0)			
$PD\ Dispersion_{f,t}$ − Equal Weighted	0.5087***	0.4777***		
	(3.56)	(3.86)		
$PD\ Dispersion_{f,t}$ − Value Weighted			0.5497***	0.4985***
			(3.57)	(3.77)
log(Fund Size)		0.0533***		0.0532***
		(5.03)		(5.01)
ETF Premium (Price/NAV)		0.1093***		0.1092***
		(3.77)		(3.76)
Observations	75,354	75,354	75,342	75,342
Fund FE	Y	Y	Y	Y
Date FE	Y	Y	Y	Y
Adj R-squared	0.219	0.231	0.219	0.231

Table 13: Wealth Effect of Bond Price Deviation Induced Creations

This table presents the average value dilution experienced by investors holding corporate bond ETFs due to bond price deviation-induced creations. We first calculate the dollar value dilution at the bond-fund-day level, as detailed in Section 6.2. We then aggregate this dollar value dilution across bonds and funds to determine the total dilution for the corporate bond ETF market. This total is divided by the same-day total corporate bond holdings of these ETFs and annualized to obtain a percentage dilution for each year from 2017 to 2021. The reported figures are the average of the annualized dilution. Additionally, we present the TNA-weighted average expense ratio, the average ETF market capitalization, and the annualized dollar amount of dilution. Column (1) reports the results for all 60 ETFs in our sample. Columns (2) – (4) reports the subsample results based on funds with estimated β_f and the associated *p*-values from the regression $Bond\ Creation_{f,b,t} = \beta_f \times PD_{f,b,t} + \mu_{f,t} + \mu_{f,b} + \varepsilon_{f,b,t}$, as in Table 8.

	(1)	(2)	(3)	(4)
			$\beta_f > 0$	
	All ETFs	All	*p*-value<10%	*p*-value<1%
Average Annualized Dilution (bps)	6.5	10.9	11.2	11.3
Average Expense Ratio (bps)	16.8	14.6	14.0	14.2
Average Market Capitalization ($Billion)	127.9	109.5	107.9	105.0
Average Annualized Dilution ($Million)	82.2	114.4	115.6	113.9
# Funds	60	42	31	21

Appendix

A Variable Definitions

Price Deviation: The ETF's reported price of the bond minus the median reported price of all ETFs (measured in bps relative to a $1 par). $PD_{f,b,t} = P_{f,b,t} - \widetilde{P_{g,b,t}}$, where $P_{f,b,t}$ is the price for bond b reported by ETF f on day t, and $\widetilde{P_{g,b,t}}$ is the median price of bond b reported by any ETF g that holds the bond on day t.

$PD\ Dispersion_{b,t}$: A bond-day level variable measured by the standard deviation (S.D.) for the price deviations, $PD_{f,b,t}$, for bond b reported by all ETFs on day t.

$PD\ Dispersion_{f,t}$: A fund-day level variable measured by the standard deviation (S.D.) for the price deviations, $PD_{f,b,t}$, for all bonds in ETF f on day t.

$Bond\ Creation_{f,b,t}$: 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.

$Bond\ Redemption_{f,b,t}$: 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 (Price/NAV): The ratio of the ETF price to the ETF's NAV from CRSP

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 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 transaction (t): The total number of transactions for the bond reported in TRACE on that day.

Number of TRACE transaction (t) > 0: A dummy is equal to one if the Number of TRACE transaction (t) is greater than zero; otherwise, it is equal to zero.

$w_{LQD,b,t-1}$: Weight of the bond b in LQD's portfolio on day $t-1$

$w_{iBoxx,b,t-1}$: Weight of the bond b in the iBoxx IG Index on day $t-1$

$P_{LQD,b,t} - \widetilde{P_{g,b,t}}$: The bond-day level overall price deviation by LQD, where $P_{LQD,b,t}$ is the price of bond b reported by LQD on day t and $\widetilde{P_{g,b,t}}$ is the median price of the bond b reported by all other ETFs on day t.

$P_{LQD,b,t} - P_{iBoxx,b,t}$: The difference in reported price between LQD and iBoxx IG Index.

$P_{iBoxx,b,t} - \widetilde{P_{g,b,t}}$: iBoxx IG Index's reported price of the bond minus the median reported price of all other ETFs (measured in bps relative to a \$1 par).

$Bond\ Creation_{LQD,b,t}$: A dummy variable equal to one if the number of bond b held by the LQD increases on day t, and zero otherwise.