BANK FOR INTERNATIONAL SETTLEMENTS

BIS Working Papers

No 975

ETFs, illiquid assets, and fire sales

by John J Shim and Karamfil Todorov

Monetary and Economic Department

November 2021

JEL classification: G01, G11, G12, G23.

Keywords: bonds, ETFs, fire sales, liquidity.

BIS Working Papers are written by members of the Monetary and Economic Department of the Bank for International Settlements, and from time to time by other economists, and are published by the Bank. The papers are on subjects of topical interest and are technical in character. The views expressed in them are those of their authors and not necessarily the views of the BIS.

This publication is available on the BIS website (www.bis.org).

ISSN 1020-0959 (print)
ISSN 1682-7678 (online)

ETFs, Illiquid Assets, and Fire Sales*

John J. Shim[†] and Karamfil Todorov[‡]

July 14, 2021

Abstract

We document several novel facts about exchange-traded funds (ETFs) holding corporate bonds. First, the portfolio of bonds that are exchanged for new or existing ETF shares (called creation or redemption baskets) often represents a small fraction of ETF holdings – a fact that we call "fractional baskets." Second, creation and redemption baskets exhibit high turnover. Third, creation (redemption) baskets tend to have longer (shorter) durations and smaller (larger) bid-ask spreads relative to holdings. Lastly, ETFs with fractional baskets exhibit persistent premiums and discounts, which is related to the slow adjustment of NAV returns to ETF returns. We develop a simple model to show that an ETF's authorized participants (APs) can act as a buffer between the ETF market and the underlying illiquid assets, and help mitigate fire sales. Our findings suggest that ETFs may be more effective in managing illiquid assets than mutual funds.

Keywords: bonds, ETFs, fire sales, liquidity

JEL classification: G01, G11, G12, G23

*The views expressed herein are those of the authors and do not necessarily reflect the views of the Bank for International Settlements. We thank Zhi Da, Andreas Schrimpf, and Sophie Shive for helpful comments, and seminar participants at the BIS and Notre Dame.

[†]University of Notre Dame Mendoza College of Business, jshim2@nd.edu

[‡]Bank for International Settlements, karamfil.todorov@bis.org

1 Introduction

Bond exchange-traded funds (ETFs) have grown steadily over the past decade and as of early 2021 manage more than $1.2 trillion compared to less than $220 billion in 2011. This growth has drawn attention in light of the mismatch between the very liquid ETF market and the less liquid bond market. This mismatch may be important because when ETF prices deviate from the net asset value (NAV) of underlying holdings, ETF's authorized participants (APs) engage in arbitrage to correct the discrepancy. This arbitrage takes place through a process called creation/redemption, in which the AP exchanges a representative portfolio of assets (called the creation or redemption basket) for shares of the ETF. The fact that bond ETFs often trade at a premium or discount to NAV has led to speculation that ETF arbitrage is ineffective when the underlying asset market is illiquid.

In this paper, we study the creation and redemption activity of bond ETFs. We implement a novel method of inferring creation and redemption baskets using data on daily ETF holdings and document several facts that distinguish bond ETFs from the more commonly studied equity ETFs. Our main empirical finding is that bond ETF baskets contain a small fraction of holdings (a fact we refer to as "fractional baskets"), which contributes to persistent discrepancies between ETF price and NAV. We argue that these discrepancies may be a feature of ETFs holding illiquid assets. In this spirit, we present a model with a bond ETF facing redemptions and an AP holding bond inventory. The key takeaway from the model is that an ETF discount arises because the AP acts as a buffer between the ETF and the bond market, allowing the ETF price to fall while avoiding selling bonds in quantities that would trigger a fire sale.

As outlined above, our main empirical finding is that bond ETFs employ fractional baskets. For ETFs holding corporate bonds (i.e., illiquid assets), we find that roughly 10% of holdings are in creation baskets and 20% are in redemption baskets, on average. This is in stark contrast to ETFs holding equities and Treasuries (i.e., liquid assets), where ETF baskets are typically equivalent to holdings. Fractional baskets challenge the common assumption that baskets are representative of holdings, and this finding has important implications for the ETF arbitrage process.

We document several additional facts that follow from fractional baskets. First, corporate bond ETFs' creation baskets exhibit high turnover: a bond in the creation basket today has on average a 25% chance to be included in the next day's creation basket. This pattern is also true for redemption baskets, though to a lesser degree, with an average of roughly 50%. Second, we find that creation baskets generally contain longer-duration bonds relative to holdings, whereas redemption baskets contain shorter-duration bonds. This fact is consistent with the idea that ETF managers use creation/redemption to maintain a relatively constant duration of their holdings. Third, we find that creation baskets tend to include bonds with wider bid-ask spreads relative to holdings, whereas redemption baskets include bonds with narrower spreads. One explanation for this pattern is that ETF managers could be less concerned about bid-ask spreads compared to APs, who are typically bond dealers. Since ETFs utilize in-kind transactions through creation/redemption, they do not incur bid-ask spread costs in contrast to APs, who transact on the bond market.

A direct implication of fractional baskets is that the ETF arbitrage mechanism may be less effective in eliminating ETF premiums and discounts. APs earn arbitrage profits by eliminating discrepancies between the ETF price and the NAV of the basket. When baskets are identical to holdings, this APs' activity also eliminates premiums and discounts. However with fractional baskets, discrepancies between the ETF price and NAV can persist since not all bonds in holdings are involved in the arbitrage. We find evidence that NAV returns imperfectly follow ETF returns for corporate bond ETFs but not for equity or Treasury ETFs, suggesting that for corporate bond ETFs, NAV responds to ETF price movements with a lag. In some respects, this lagged NAV adjustment is purely mechanical – if not all bonds in holdings are included in the basket, then APs simply do not trade these bonds when performing arbitrage. This finding suggests that volatility (Ben-David et al., 2018) and comovement (Shim, 2020; Da and Shive, 2018) effects associated with arbitrage trading in equity ETFs may be muted in bond ETFs.

Persistent premiums and discounts can arise not only through lagged NAV adjustment, but also via ETF price overreaction. To investigate this channel, we next study episodes with large premiums or discounts. We find that large premiums arise mostly due to lagged changes in NAV. That is, NAV lags behind ETF price, which allows the premium to persist. The premium is eventually eliminated solely from NAV "catching up" over the course of several days. However, discount episodes paint a different picture. Similar to premiums, large discounts are in part related to slow NAV adjustment. In contrast however, large discounts are also created by initial ETF price overreaction. That is, significant ETF price declines contribute to initial large discounts, which are eliminated in part by subsequent ETF price reversals.

In some respects, delayed NAV response to ETF price movements, which is related to fractional baskets, may seem like a "flaw." Indeed, equity ETFs often tout the lack of significant premiums and discounts as a prominent feature. However, as mentioned above, we argue that the delayed response of NAV is a potential benefit of ETFs managing illiquid assets. To understand why, consider the case of rapid selloff in the ETF market. If the ETF were to immediately transmit this selling pressure to the underlying illiquid assets, that could severely depress prices and lead to fire sales. The delayed response of NAV allows the ETF to act as a buffer, absorbing panic selling in the more liquid ETF market, while mitigating the impact on the more fragile and less liquid market for underlying assets.

We build a simple model to illustrate that ETFs can be effective in preventing the rapid liquidation of illiquid assets in a fire sale. The model features an ETF that holds illiquid bonds and an AP that is a dealer in those bonds (as is the case in reality – see, e.g., Pan and Zeng (2021)). As a dealer, the AP holds inventory in the bonds. When ETF investors sell shares, the AP sets a price to buy those shares based on her cost of handling (combination of holding and selling) bonds in the redemption basket. We also allow for the possibility of a fire sale by assuming that trading large quantities of a bond incurs large transaction costs and depresses prices. This definition of a fire sale is in the spirit of Shleifer and Vishny (1992) in the sense that APs are better off selling illiquid assets to specialist buyers, but rapid selling of large quantities may require trading with non-specialist buyers at significantly depressed prices.

The model shows that when the AP holds inventory, she endogenously avoids fire sales and puts a greater fraction of the redeemed bonds on her balance sheet than she otherwise would. The reason behind this stems from AP's role as a dealer. Due to her initial inventory, the AP optimally avoids a fire sale since selling bonds at fire sale prices would lead to large mark-to-market losses. In other words, the AP internalizes the cost of fire sales because she has "skin in the game."[1]

The model has several implications. First, the ETF creation/redemption process allows the AP to serve as a buffer between ETF investors and illiquid asset markets. The AP effec-

[1] Avoiding fire sales is not only beneficial for the AP but also for bond markets. Preventing a rise in yields could help to maintain a well-functioning primary corporate bond market despite economic downturns, as shown in Becker and Benmelech (2021).

tively acts as an agent who manages the inventory of illiquid assets on behalf of redeeming ETF investors and ensures orderly liquidation based on the cost of selling versus the cost of holding additional inventory. Second, redeeming ETF investors benefit from the presence of an AP. In our model, the AP passes along the costs of managing the assets in the redemption basket to the ETF investors by buying ETF shares at a larger discount. However, this discount is strictly smaller than the discount if there were no AP. For example, in the case of a mutual fund, redeeming investors would trigger forced selling of bonds to meet redemptions, even if that leads to a fire sale. This forced selling would result in lower price for mutual fund investors and in depressed bond prices.[2]

Third, non-redeeming ETF investors remain insulated from large ETF redemptions. Since the AP acts as a buffer between the ETF market and the illiquid bond market, redeeming ETF investors pay most of the costs associated with large redemptions, which are reflected in the ETF discount. The AP internalizes the cost of holding illiquid assets and thus, her incentives are aligned with those of non-redeeming ETF investors, who also effectively hold the underlying assets. That is, by minimizing the costs of selling in times of fire sales, the AP acts to protect non-redeeming ETF investors.

Related Literature Our paper is related to three distinct strands of literature: ETFs, bond market illiquidity and fire sales.

First, this paper is related to the growing literature on ETFs. Ben-David et al. (2018) show that ETF arbitrage transmits noise from the ETF market to the underlying securities and increases volatility. Malamud (2015) demonstrates that ETFs can create a transmission mechanism for non-fundamental shocks to underlying securities. Saglam et al. (2019) show that ETFs increase stock liquidity, whereas Da and Shive (2018) find that they increase return comovement. Brogaard et al. (2019) show that the impact of ETFs on liquidity depends on the fund's index replication strategy. Shim (2020) argues that ETF arbitrage mistranslates systematic information from ETFs to constituent securities. Todorov (2019)

[2]It is common for mutual funds to hold cash or liquid assets to avoid forced selling of illiquid assets (Jin et al. (2021)). However, doing so also incurs a cost in the form of tracking error or cash drag. Mutual funds can also use the mechanism of swing pricing, which allows them to adjust the price at which investors redeem by a swing factor. The purpose of swing pricing is to effectively charge redeeming investors the costs associated with selling bonds. The swing factor is usually much smaller than typical ETF discounts during times of market stress and is generally hard to observe (Jin et al. (2021), Lewrick and Schanz (2017)). In addition, the maximum swing factor is typically set by mutual funds in advance and is not adjusted in response to real-time market factors.

shows that ETFs in VIX and commodity markets put pressure on prices and amplify price changes. Sushko and Turner (2018) document the increase in the share of ETFs in several markets and study the impact for liquidity and volatility.

Several papers analyze bond ETFs in particular. Dannhauser (2018) finds that ETF ownership increases prices but reduces bond liquidity. Holden and Nam (2019) find that ETFs improve liquidity in less accessible corporate bonds. Pan and Zeng (2021) show that APs' dual role as major bond dealers and ETF arbitrageurs can create conflict of interest. Haddad et al. (2020), D'Amico et al. (2020), Aramonte and Avalos (2020), Laipply and Madhavan (2020), and Boyarchenko et al. (2020) study bond ETF discounts during the COVID crisis and the subsequent Fed intervention. Todorov (2021) describes the specifics of the bond ETF arbitrage mechanism.

Our contribution relative to existing studies on ETFs is threefold. First, we introduce a new methodology to infer realized baskets. Second, we show that these baskets are only a fraction of holdings, which differentiates the arbitrage mechanism of bond ETFs from that of more traditionally studied equity ETFs. We illustrate that fractional baskets can create persistent ETF premiums and discounts. Third, we argue that fractional baskets for ETFs holding illiquid bonds can be beneficial at times of stress by allowing APs to act as a buffer and prevent fire sales.

Our research is also related to studies of the bond market. Bessembinder et al. (2020) provide a summary of the main differences between equity and fixed income markets. Bessembinder et al. (2018) show that the median trade size in bond markets is more than \$1 million, which is much larger than that in equity markets. Bessembinder et al. (2009) find that round-lot trades account for approximately 90% of corporate bond dollar trading volume. Goldstein and Hotchkiss (2020) show that dealers endogenously adjust their behavior to mitigate inventory risk from trading in illiquid and higher risk bonds. Bao et al. (2011) show that the illiquidity of corporate bonds impacts their prices. We illustrate that the specifics of the bond market lead to different arbitrage mechanics of bond ETFs relative to equity ETFs.

The research presented here also contributes to the extensive literature on fire sales. This literature stems from the classic paper of Shleifer and Vishny (1992), which shows that forced sales can lead to depressed liquidation values and can have significant implications for firms' cost of capital. Coval and Stafford (2007) illustrate that mutual fund outflows can lead to fire sales and financial distress in equity markets. Our paper adapts the logic of fire sales to the setting of an ETF that manages illiquid assets.

The remainder of this paper is organized as follows. Section 2 describes the data and institutional details. Section 3 documents new facts about bond ETFs. Section 4 studies the implications for the arbitrage mechanism. Section 5 presents the model and Section 6 concludes.

2 Institutional Details and Data

In this section, we explain how ETF arbitrage works conceptually and outline the main differences for bond ETFs. We then discuss the data.

2.1 Bond ETF Arbitrage in Theory and Practice

There are two markets for ETFs: a primary market, where new ETF shares are created and destroyed, and a secondary market, where investors trade existing ETF shares.[3] In the primary market, Authorized Participants (APs) create new or redeem existing ETF shares in exchange for a portfolio of assets (bonds or stocks); this portfolio is referred to as the creation/redemption basket. The secondary market is where market participants trade ETF shares, just like they trade stocks. APs are usually large broker-dealers or liquidity providers in the underlying ETF assets and may operate both in the primary and secondary market.

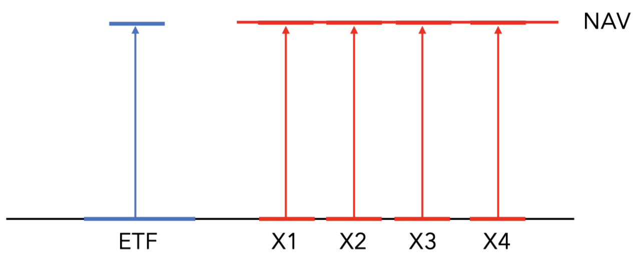
The process of creation/redemption works conceptually because APs are incentivized to eliminate ETF premiums/discounts by arbitrage profits.[4] The textbook version of this arbitrage is as follows. During the course of a typical trading day, price pressure may cause the ETF price to deviate from NAV. APs can then buy the asset with the lower price and sell the asset with the higher price until the two converge. APs do so because they are able to use the creation/redemption mechanism to convert the portfolio of ETF holdings into ETF shares and vice-versa. This conversion nets out the AP's position and locks in arbitrage profits from buying low and selling high.

A critical assumption in this textbook ETF arbitrage is that ETF holdings and creation/redemption baskets are identical. This ensures that when APs eliminate discrepancies between the ETF price and the NAV of the basket, they also eliminate discrepancies be-

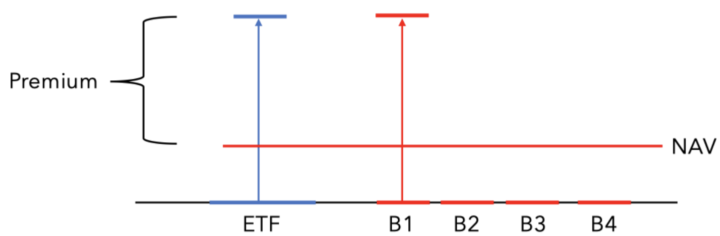
[3]See Lettau and Madhavan (2018) for more details on ETF mechanics.
[4]APs are not legally obligated to correct ETF premiums or discounts.

Figure 2.1: ETF Arbitrage Illustration

(a) Textbook Case



(b) Fractional Baskets Case



Notes. The figure illustrates the ETF arbitrage mechanism when the creation/redemption basket is identical to holdings and there are no trading frictions. In Panel (a), the ETF holds four assets (X1, X2, X3 and X4). When there is an increase in the price of the ETF, the ETF temporarily trades at a premium to the NAV of its holdings. Authorized participants (APs) sell the ETF at the new higher price and buy all four assets, pushing up their prices and eliminating the premium. In Panel (b), we present the same case but when the ETF utilizes baskets that are a subset of holdings, i.e., the case of "fractional baskets" as seen in non-Treasury bond ETFs. Panel (b) shows that the ETF also holds four assets (B1, B2, B3, and B4 to denote these assets as bonds), but only one is in the creation basket and thus only the NAV of the basket containing the single bond is traded by the APs.

tween the ETF price and the NAV of the holdings.[5] Figure 2.1 presents an example of the described textbook arbitrage for an ETF that holds an equal-weighted portfolio of four assets. In Panel (a), the increase in the ETF price creates a premium which is eliminated by arbitrageurs who purchase the assets and sell the ETF to capture arbitrage profits and eliminate the premium.

The assumption that baskets are equal to holdings is generally true for equity ETFs but not for non-Treasury bond ETFs (we show this fact in Section 3.1). The fact that bond ETF baskets are a fraction of holdings is likely due to bonds being more illiquid than equities, making it costly to trade the entire portfolio of bond holdings. As a result, it is relatively common for bond ETF sponsors to create so-called custom baskets. These baskets are typically different from holdings and the pre-announced basket that ETF sponsors publicly declare, and may vary from one AP to another. The baskets can be also different for creation and redemption. In 2019, the Securities and Exchange Commission (SEC) effectively allowed for more widespread use of custom baskets (see, e.g., Kaminska (2020)). Custom baskets are an important element of ETFs holding relatively illiquid assets. They allow for more flexibility in the creation/redemption process by taking trading frictions and inventory constraints into account. Importantly, in Section 3.1 we show that the actual baskets used for creation and redemption in bond ETFs contain only a small fraction of holdings.

When creation/redemption baskets are a fraction of holdings, the ETF arbitrage mechanism may be less effective in eliminating premiums and discounts compared to the textbook arbitrage case. Panel (b) of Figure 2.1 presents a scenario of ETF arbitrage with fractional baskets: the creation basket contains only one of the four assets held by the ETF. In this case, APs eliminate the discrepancy between the NAV of the *basket* and the ETF price, just like in the textbook case above. However, since the basket is only a quarter of the holdings, the NAV of the *holdings* has a muted response and does not fully adjust to the new ETF price. This muted response results in a premium.

2.2 Data and Basket Construction

We use three sources for ETF and bond data: ETF Global, Bloomberg, and Thomson Reuters Datastream.

[5]There are other assumptions in the textbook example of ETF arbitrage: no transaction costs, no creation/redemption costs, no inventory holdings costs, etc.

We utilize ETF Global for ETF-level data (prices, NAV, flows, assets under management or AUM, and shares outstanding) as well as holdings data. The benefit of the ETF Global data is that it covers the universe of U.S. bond ETFs. We make three sets of sample restrictions. First, we focus initially on the period between September 2017 to February 2020, just before the start of the COVID pandemic in the US, as a baseline. The COVID and post-COVID periods are examined separately in Appendix A. Second, we restrict our sample to the set of ETFs holding at least some U.S. corporate or government bonds with AUM over $100 million (M) at the start of our sample to focus on funds that generate significant primary and secondary market activity. Third, we exclude ETFs from issuers that do not consistently update holdings on a daily basis, as well as leveraged and inverse ETFs, which typically hold derivative contracts instead of bonds.

We employ a novel approach to compute realized creation and redemption baskets. We infer baskets based on changes in ETF holdings on days with creation or redemption activity. To the best of out knowledge, our paper is the first to employ this idea and analyze realized as opposed to announced baskets (the latter are provided by the Depository Trust and Clearing Corporation (DTCC) and mentioned in Pan and Zeng (2021)). There are significant differences between announced and realized baskets on a daily basis, which is consistent with the negotiation that takes place between ETF sponsors and APs to determine what is included in the realized basket. For example, BlackRock refers to an internal system that handles custom baskets, through which APs propose baskets and ETF managers can accept, modify, or reject these proposals.[6] Our methodology implicitly assumes that changes in holdings are mainly driven by creation/redemption activity. Anecdotal evidence from our conversations with several major ETF sponsors suggests that this is a reasonable assumption.

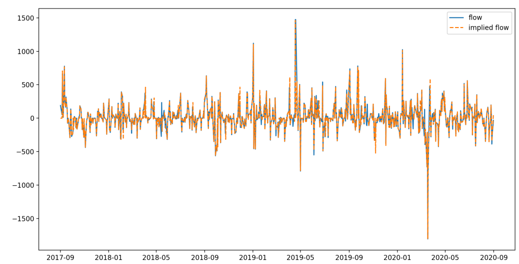
We also study the composition of realized baskets by merging the bonds from holdings with data on individual bond characteristics and returns. This process requires precise holdings data. To ensure our methodology is accurate, we manually check the holdings data and inferred baskets for a subsample of 10 of the largest U.S. bond ETFs. We use hand-collected Bloomberg data to cross-check holdings data from ETF Global, and verify the data with ETF prices and returns data from the Center for Research in Security Prices (CRSP).[7]

[6] Alluded to in a BlackRock white paper (BlackRock, 2021) and discussed during a webinar marking the public release of the paper.

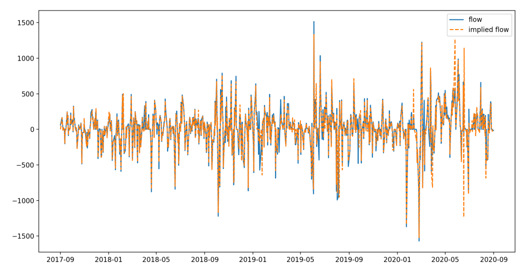
[7] A common correction we manually make is to handle the case where ETF Global or Bloomberg data appears to be shifted by a day. These cases differ by ETF and require manual adjustments to ensure the data

Figure 2.2: Flows vs. Implied Flows

(a) Treasury Bond ETF



(b) High-Yield Bond ETF



Notes. The Figure shows actual and implied flows for two representative ETFs: a Treasury bond ETF (Panel (a)) and a high-yield corporate bond ETF (Panel (b)). Flows are in millions of dollars. See Section 2.2 for more details on the methodology and data.

We then merge the individual bonds in holdings and baskets with bond data from Thomson Reuters Datastream and use this subsample for the analysis of basket characteristics (Section 3.2).

We verify that our method of inferring baskets for the subsample of 10 bond ETFs is sensible by analyzing creation/redemption flows implied by our baskets. Figure 2.2 illustrates the close relation between implied flows and actual flows for a Treasury bond ETF and a high-yield corporate bond ETF. The average correlation between implied and actual flows for the subsample of the 10 bond ETFs is 0.88.[8] We view the close relationship between

is accurate. We also manually correct erroneous data points from this subsample. For example, some data gaps imply that an ETF completely sells its bond holdings and buys them back in exactly the same amount in subsequent days. We verify that these outliers are not the result of actual primary market activity, since the implausibly large outflows followed by identical inflows do not match with the actual recorded flow in Bloomberg and ETF Global.

[8]The correlation is not exactly 1 since the prices in Thomson Reuters Datastream might be different from the ones used to compute actual flows in Bloomberg and ETF Global. This is so because many corporate bonds do not trade on a daily basis and their prices are based on quotes. These quotes can differ across the different data sources, which would give rise to different dollar flows and lead to imperfect correlations.

Table 1: Summary Statistics on Creation/Redemption Flows

	Treasury	Broad Debt	Short Term	Inv. Grade	High Yield
Share of days with creation, %	38.00	46.38	32.13	40.40	43.07
Share of days with redemption, %	20.91	9.18	12.96	16.00	29.53
Creation size, $M	77.30	54.99	41.03	68.50	94.79
Redemption size, $M	-86.32	-63.42	-70.23	-139.76	-152.97
Creation size, percentage of AUM, %	0.84	0.23	0.32	0.45	0.99
Redemption size, percentage of AUM, %	-0.95	-0.30	-0.48	-0.89	-1.53

Notes. The sample includes bond ETFs with over $100M as of September 2017, and the numbers are averages across all bond ETFs within a category over all days from September 2017 to February 2020. "Treasury" is all Treasury ETFs, "Broad Debt" is all bond ETFs that track the aggregate bond market and hold different types of bonds (including government bonds, corporate bonds, and mortgage-backed securities). "Short Term" is ETFs holding short-term (1-5 yr) investment grade corporate bonds, "Inv. Grade" is ETFs holding investment grade bonds with maturity longer than 5 years, and "High Yield" is ETFs holding high-yield corporate bonds. See Section 2.2 for more details on the data.

implied and actual flows as support that our methodology for inferring baskets is sensible, especially given that corporate bond prices and returns may be stale and noisy based on the variety of bond dealers that provide different quotes.

Table 1 presents summary statistics on creation and redemption activity for the U.S bond ETFs in our ETF Global sample. The table shows that creations are roughly two-to-five times more likely to occur than redemptions. The average size of actual creations and redemptions is typically below 1% of AUM for most ETFs. Redemptions are usually larger than creations.

3 New Facts on Bond ETFs

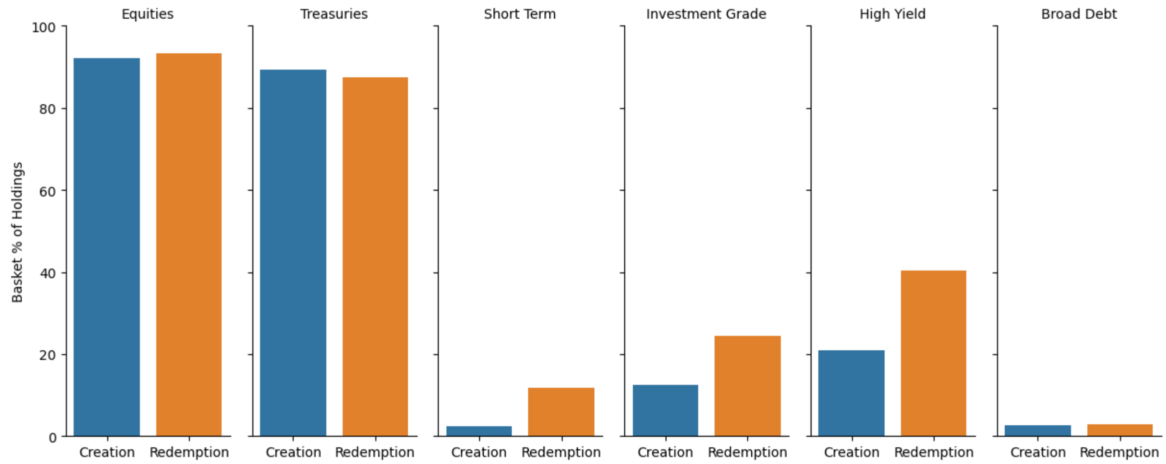
In this section, we present new facts about bond ETF arbitrage. These facts illustrate how bond ETF arbitrage in reality differs from the textbook arbitrage described in Section 2.1.

3.1 ETF Basket Composition

3.1.1 Baskets are a Fraction of Holdings

Figure 3.1 presents the mean basket size as a fraction of ETF holdings for several types of ETFs in our ETF Global sample. The figure shows a striking difference in the basket fraction between equity/Treasury ETFs and bond ETFs holding U.S. corporate bonds. Equity and Treasury ETFs have a mean basket fraction of around 93% and 88% respectively, and a median basket fraction of 100% and 94%. These ETFs are largely consistent with the

Figure 3.1: Basket as Fraction of Holdings



Notes. For each creation and redemption day and for each ETF, we calculate the fraction of bonds held by the ETF that are included in the creation or redemption basket. We then compute the average of these fractions for each ETF and day in the category from September 2017 to February 2020. The bond ETF sample includes bond ETFs with over $100M as of September 2017. "Equities" is several large equity ETFs tracking the S&P 500, Russell 2000, and Nasdaq 100. "Treasury" is all Treasury ETFs, "Broad Debt" is all bond ETFs that track the aggregate bond market and hold different types of bonds (including government bonds, corporate bonds, and mortgage-backed securities). "Short Term" is ETFs holding short-term (1-5 yr) investment grade corporate bonds, "Investment Grade" is ETFs holding investment grade bonds with maturity longer than 5 years, and "High Yield" is ETFs holding high-yield corporate bonds. See Section 2.2 for more details on the data and methodology for computing creation and redemption baskets.

textbook arbitrage example provided in Section 2.1 in that their baskets are almost identical to holdings. In contrast, baskets of bond ETFs holding more illiquid securities are much less representative of holdings. For example, short term, investment grade, and high yield bond ETFs have on average around 12%, 25%, and 40% of holdings in redemption baskets. Moreover, the fraction of holdings in creation baskets is even smaller: short term, investment grade, and high yield bond ETFs have around 3%, 12% and 20% of holdings in creation baskets. Broad debt ETFs place only around 3% of holdings in creation and redemption baskets.

There are at least two features of the bond market that can help explain these empirical facts. First, the bond market is a more concentrated, dealer-dominated market and is generally much less liquid than the equity market (Bessembinder et al., 2020). Second, the minimum trading size of bonds is several orders of magnitude larger than that of equities. Bessembinder et al. (2020) report that the median trade size in bond markets is more than $1 million, much larger than in equity markets (which is typically 100 shares, e.g., $5,000 for a $50 stock). These bond market frictions make it more difficult and costly for APs to deliver (liquidate) the full set of bond holdings in a creation (redemption) basket, and may lead to persistent ETF premiums and discounts as described in Section 2.1.

3.1.2 Basket Composition is not Persistent

Figure 3.2 shows the persistence of creation and redemption baskets for ETFs in our ETF Global sample. Specifically, we compute the probability that a bond in a creation (redemption) basket on a given day is subsequently included in the basket on the next creation (redemption) day (Panel (a)), and over a range of such days (Panel (b)). Whereas persistence in equity and Treasury ETF baskets over the next day is very high and close to 100%, other bond ETFs have relatively low persistence. Short term, investment grade, high yield, and broad debt bond ETFs have persistence of 32%, 28%, 29%, and 12% for creation, respectively. Redemption baskets for these four bond ETFs are somewhat more persistent, particularly for investment grade and broad debt ETFs.

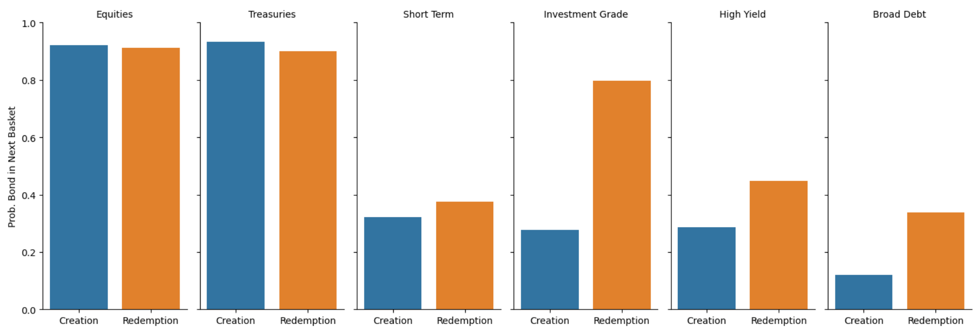
The composition of creation baskets often depends on the availability of bonds in the underlying market or in dealer inventories, which can vary substantially. In contrast, redemption baskets always draw on bonds that are already part of ETF holdings and are thus less dependent on APs' inventory holdings or market conditions. This is consistent with lower persistence of creation baskets relative to redemption baskets.[9]

Panel (b) shows that bonds in a creation basket are more likely to be included in a creation basket several days after creation activity than on the next creation day (similarly for redemption baskets). This fact suggests that ETFs rotate bonds in and out of baskets, perhaps to transmit arbitrage price pressure to more non-basket bonds. The fact that bonds have large minimum trading sizes (as outlined above) can also explain lower basket persistence at short intervals but greater persistence at longer intervals. For example, a creation basket may significantly overweight a few bonds (due to minimum trading increments) and underweight or exclude all other bonds. In subsequent creation days, the overweighted bonds are excluded from the basket and other underweighted bonds are included. Through this process, the ETF can mimic holdings with baskets aggregated over many days, even if

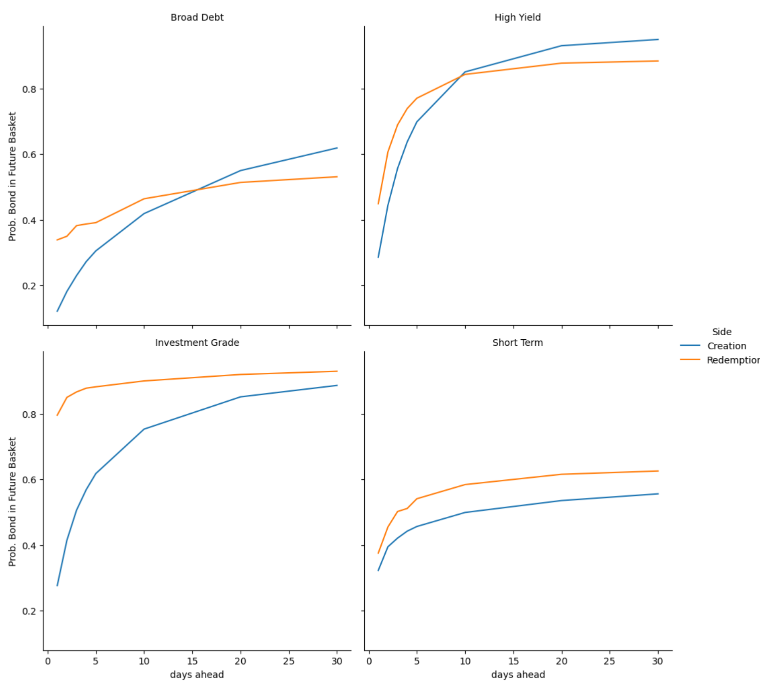
[9] Anecdotal evidence from our conversations with major ETF sponsors suggests that they have an incentive to maintain long-term relationships with APs, and baskets may be the outcome of a formal or informal negotiation process. This may be another reason for differences between the composition of baskets and holdings. For example, if an AP cannot deliver a particular bond, it could propose an alternative bond that the AP has in her inventory or that is easier/cheaper to transact in the market. Even if this alternative bond is not part of the ETF benchmark, the ETF sponsor might accept the proposal if the bond keeps tracking error low (is a suitable substitute) and helps maintain a good relationship with the AP, whose market-making function provides a valuable service to the ETF sponsor.

Figure 3.2: Basket Persistence

(a) Persistence (1 day)



(b) Persistence Over Several Days



Notes. We measure persistence as the probability that a bond in a basket on a given day is in the basket of the same type (creation or redemption) over the next creation/redemption day (Panel (a)), or a range of such days (Panel (b)). The sample includes bond ETFs with over \$100M as of September 2017 and includes the period between September 2017 and February 2020. “Equities” is several large equity ETFs tracking the S&P 500, Russell 2000, and Nasdaq 100. “Treasury” is all Treasury ETFs, “Broad Debt” is all bond ETFs that track the aggregate bond market and hold different types of bonds (including government bonds, corporate bonds, and mortgage-backed securities). “Short Term” is ETFs holding short-term (1-5 yr) investment grade corporate bonds, “Investment Grade” is ETFs holding investment grade bonds with maturity longer than 5 years, and “High Yield” is ETFs holding high-yield corporate bonds. See Section 2.2 for more details on the data and methodology for computing creation and redemption baskets.

baskets on any given day are a small fraction of holdings.

3.2 ETF Basket Characteristics

3.2.1 Basket Duration, Liquidity, and Ratings

Figure 3.3 compares average duration, bid-ask spread, and ratings for bonds in baskets vs. holdings. For each of these three variables, we compute the weighted average value for the creation/redemption basket and for ETF holdings for each day and ETF. We then report the average difference between baskets and holdings across all days and ETFs for each ETF category in our more precise subsample (see Section 2.2 for more details on the sample). We omit equity and Treasury ETFs since their baskets are very similar to holdings.

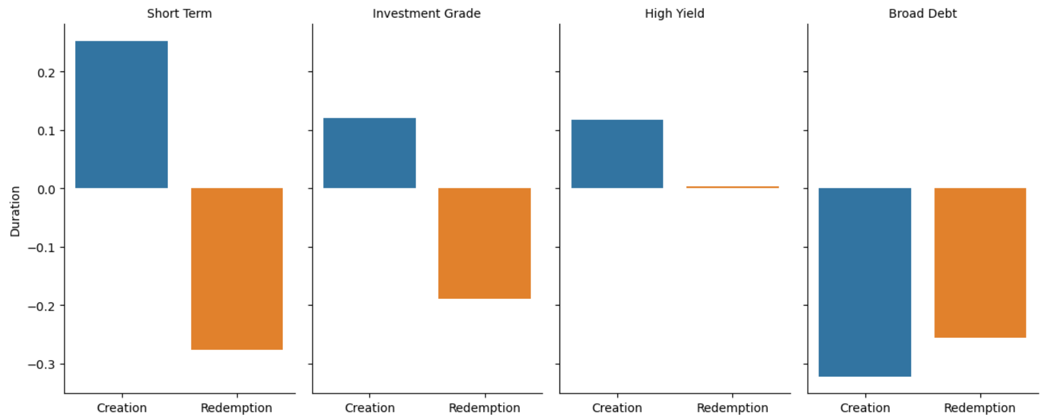
Panel (a) shows that relative to holdings, creation baskets have longer duration, whereas redemption baskets have shorter duration, with the exception of broad debt ETFs. These findings are consistent with ETFs using the creation/redemption process for portfolio management. For example, since bonds have a finite maturity, ETF sponsors must rebalance their portfolio and substitute expiring/shorter-duration bonds with longer-duration bonds (similar to calendar rebalancing for VIX and commodity ETFs in Todorov (2019)). Anecdotal evidence from our conversations with ETF sponsors suggests that they use the creation/redemption process to dispose off bonds that fall below the target maturity and acquire bonds with longer maturities. Since creation/redemption is done in-kind, this is a more cost-effective way for ETFs to attain these targets as opposed to trading bonds directly.

Panel (b) shows that redemption baskets typically have smaller bid-ask spreads than holdings, and creation baskets typically have larger spreads. This is consistent with the idea that ETFs take advantage of the in-kind nature of creations and redemptions: in-kind transfers mean that ETFs do not pay bid-ask spreads associated with trading bonds. Thus, ETFs may be less sensitive to holding illiquid bonds. On the other hand, APs, who are typically bond dealers, do transact in the bond market and as a result may be more sensitive to bond liquidity. Thus, APs may prefer to place illiquid bonds in creation baskets and receive more liquid bonds in redemption baskets.

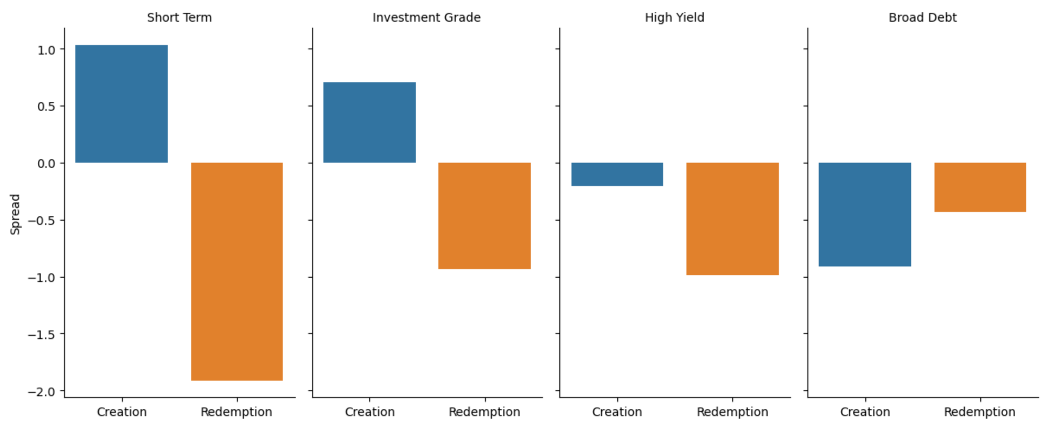
Panel (c) shows that there is very little difference in ratings between baskets and holdings, with the exception of short-term ETF redemption baskets. One caveat is that our data on ratings is limited to ratings at the time of issuance and does not capture time-variation in

Figure 3.3: Duration, Bid-Ask Spreads, and Ratings of Baskets

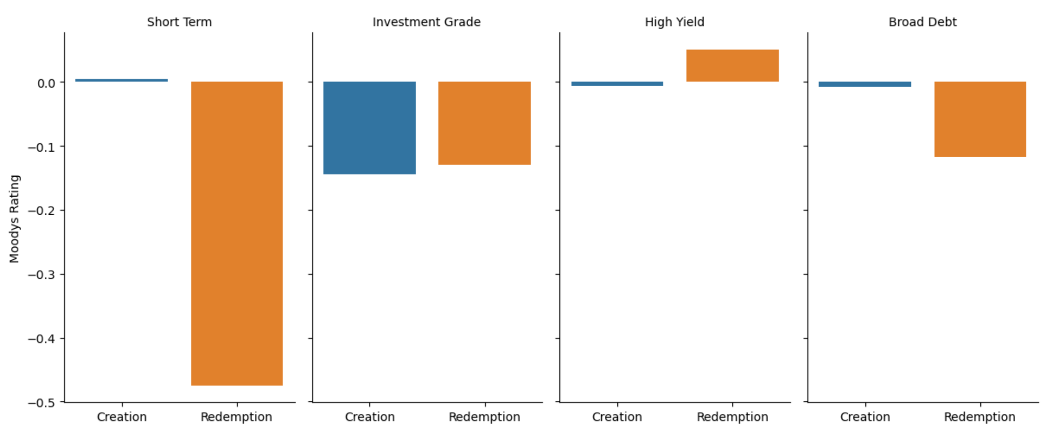
(a) Duration



(b) Bid-Ask Spreads



(c) Ratings



Notes. Panel (a) presents the difference between the weighted-average duration (in years) of bonds in the creation/redemption basket and that of holdings, averaged over days and ETFs within an ETF category. Panel (b) presents the same difference for the weighted-average bid-ask spread (in basis points), and Panel (c) presents differences in bond ratings (converted to a numerical scale; higher numbers correspond to higher ratings). We caveat that our data on bond ratings is static and does not capture changes in ratings. We use a sample of 10 ETFs that are manually checked for accuracy of daily holdings and are merged with bond data from Thomson Reuters Datastream. "Broad Debt" is all bond ETFs that track the aggregate bond market and hold different types of bonds (including government bonds, corporate bonds, and mortgage-backed securities). "Short Term" is ETFs holding short-term (1-5 yr) investment grade corporate bonds, "Investment Grade" is ETFs holding investment grade bonds with maturity longer than 5 years, and "High Yield" is ETFs holding high-yield corporate bonds. See Section 2.2 for more details on the data and methodology for computing creation and redemption baskets.

ratings.

3.2.2 Basket Returns

Figure 3.4 shows the return on the basket (relative to holdings) on the day of creation/redemption (Panel (a)), over the 5 days prior to creation/redemption (Panel (b)), and over the 5 days after creation/redemption (Panel (c)). The figure shows that ETFs tend to "buy" high and "sell" low: creation baskets have larger returns than redemption baskets on the day of creation. This pattern also holds for the returns in the 5 days before creation/redemption (Panel (b)). In the period after creation/redemption (Panel (c)), there is reversal except for broad debt ETFs: created bonds perform worse than redeemed bonds. These facts suggest one of two possibilities. APs may be making a profit on the creation/redemption process: they obtain bonds that perform better in the future compared to the ones that they give to the ETF (especially, for high-yield and short-term bond ETFs). Alternatively, this pattern of returns may be related to price pressure from arbitrage trading activity before creation/redemption and then reversal after creation/redemption.

4 Implications for ETF Premiums and Discounts

In this section, we present the implications of fractional and impersistent baskets on bond ETF premiums.[10]

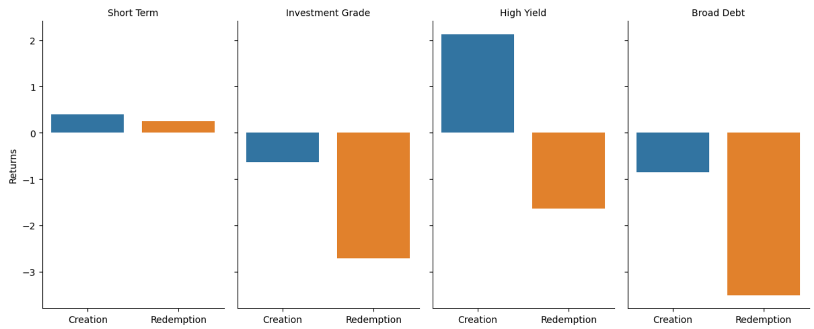
4.1 Conceptual Framework

To fix ideas, consider the simple example shown in Figure 4.1, which depicts an ETF that holds four bonds. In the example, the ETF price increases, which creates a discrepancy between the ETF price and NAV, or a positive ETF premium (Panel (a)). This results in an arbitrage opportunity for APs. However, as shown in Section 3.1.1, only a fraction of the bonds – bond B1 in the example – is included in the creation basket. As a result, APs conduct the arbitrage by buying B1 and selling the ETF share until the prices of the two are aligned. That is, arbitrage works perfectly in aligning the ETF price and the NAV of the *creation basket*, but works imperfectly in aligning the ETF and the NAV of *holdings*. As a

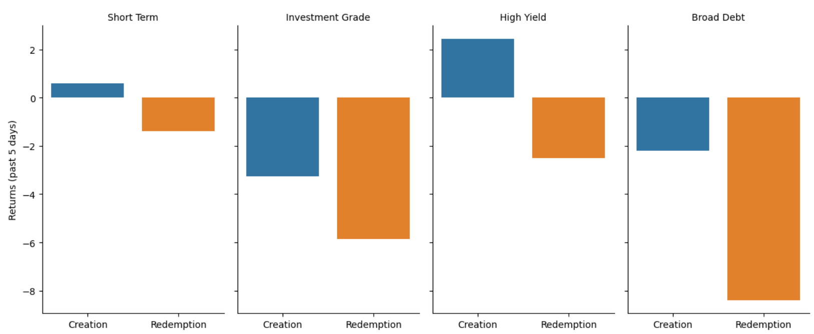
[10]We occasionally use the term premium to describe both positive premiums and negative premiums (i.e., discounts).

Figure 3.4: Basket Returns

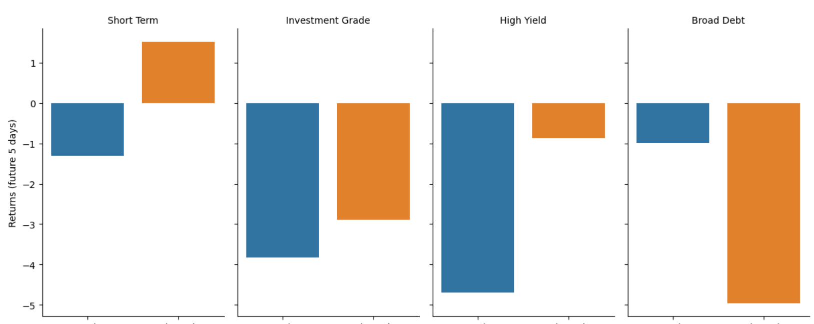
(a) Returns, Creation/Redemption Day



(b) Returns, 5 Days Before Creation/Redemption



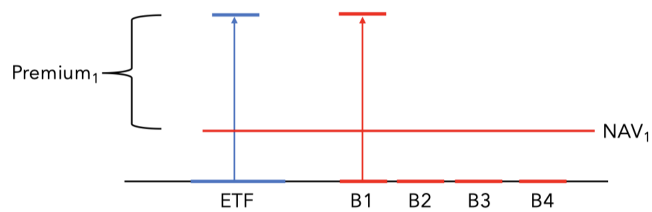
(c) Returns, 5 Days After Creation/Redemption



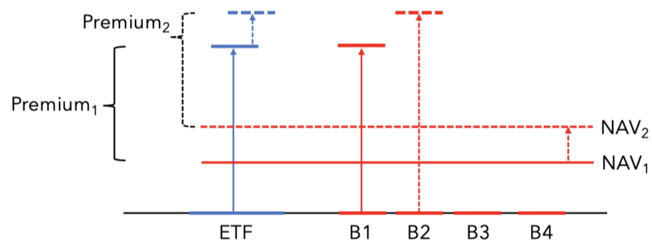
Notes. Panel (a) show the return for baskets relative to holdings on the day of creation/redemption, and Panels (b) and (c) show the cumulative return in the 5 days before and after creation/redemption. All returns are in basis points and are averaged over all days and ETFs within an ETF category. We use a sample of 10 ETFs that are manually checked for accuracy of daily holdings and are merged with bond data from Thomson Reuters Datastream. "Broad Debt" is all bond ETFs that track the aggregate bond market and hold different types of bonds (including government bonds, corporate bonds, and mortgage-backed securities). "Short Term" is ETFs holding short-term (1-5 yr) investment grade corporate bonds, "Investment Grade" is ETFs holding investment grade bonds with maturity longer than 5 years, and "High Yield" is ETFs holding high-yield corporate bonds. See Section 2.2 for more details on the data and methodology for computing creation and redemption baskets.

Figure 4.1: Example: Bond ETF Arbitrage

(a) Day 1: Fractional Basket.



(b) Day 2: Fractional Basket without Persistence.



Notes. Panel (a) shows that arbitrage does not eliminate an ETF premium when the creation/redemption basket is a subset of holdings. The figure shows that arbitrage works to eliminate the discrepancy between the ETF and the NAV of the basket, but there is still a discrepancy between the ETF and the NAV of holdings. Panel (b) extends the example in Panel (a) by another day, where the ETF price continues to increase and the basket is still a subset of holdings but its composition changes. In the extended example, B1 is no longer in the the basket and has been replaced by B2. Just as in Panel (a), arbitrage eliminates the discrepancy between ETF price and the creation basket, but does not eliminate the premium.

result, the ETF premium remains positive because the rest of the holdings (bonds B2, B3, and B4) are not included in the basket and are not involved in the arbitrage trade. Thus, in this example the non-basket bond prices remain unchanged. Continuing the example, Panel (b) depicts the next day, on which the ETF price rises further. In this case, the creation basket now contains a different bond, B2, capturing impersistence of baskets as shown in Section 3.1.2. Similar to Panel (a), APs take advantage of the arbitrage opportunity by buying B2 and selling the ETF until the discrepancy is eliminated. Again, arbitrage works properly in eliminating the discrepancy between ETF price and the NAV of the creation basket, but fractional baskets prevent the NAV of holdings from responding fully. As a result, the premium persists.

We now derive a simple identity to motivate our empirical strategy. Let $P_{e,t}$ and $NAV_{e,t}$ represent the price and NAV of ETF e at time t. The premium is $\pi_{e,t} = P_{e,t} - NAV_{e,t}$, and the percentage premium is $\pi^*_{e,t} = \pi_{e,t}/NAV_{e,t}$. Then, the following identity holds:

$$\pi_{e,t} = \pi_{e,t-1} + \Delta P_{e,t} - \Delta NAV_{e,t}. \tag{4.1}$$

Table 2: ETF Premiums and Discounts

	Equities	Treasury	Short Term	Inv. Grade	High Yield	Broad Debt
Mean π_e (bps)	-0.11	1.60	2.80	7.37	17.69	2.77
Std. π_e (bps)	11.73	21.51	48.53	54.55	40.80	31.99
$\pi_e > 0$ (% days)	75.80	75.20	82.28	73.16	78.99	53.03
Mean π_e (>0, bps)	3.90	5.58	9.41	19.46	28.31	9.12
Mean π_e (<0, bps)	-4.72	-16.93	-34.21	-27.01	-22.74	-20.37

Notes. This table shows summary stats for ETF premiums. The equity ETF sample contains several large equity ETFs tracking the S&P 500, Russell 2000, and Nasdaq 100 over all days from September 2017 to February 2020. The bond ETF groups include all bond ETFs holding at least some corporate bonds or Treasuries with over $100M as of September 2017. The numbers are averages across all ETFs within a category over all days from September 2017 to February 2020. "Treasury" is all Treasury ETFs, "Broad Debt" is all bond ETFs that track the aggregate bond market and hold different types of bonds (including government bonds, corporate bonds, and mortgage-backed securities). "Short Term" is ETFs holding short-term (1-5 yr) investment grade corporate bonds, "Inv. Grade" is ETFs holding investment grade bonds with maturity longer than 5 years, and "High Yield" is ETFs holding high-yield corporate bonds. See Section 2.2 for more details on the data.

We can also represent this expression in returns as

$$\pi^*_{e,t} \approx \pi^*_{e,t-1} + \left(r_{e,t} - r_{NAV_e,t}\right). \tag{4.2}$$

Equation 4.2 shows that the premium today is a function of the past premium, ETF return, and NAV return.[11] That is, the change in premium is based on how much the ETF price moves relative to how much NAV moves. If ETF returns move one-to-one with NAV returns, premiums are not persistent.

Table 2 presents summary statistics on premiums. Premiums generally increase with the illiquidity of the asset and decrease with the share of basket in holdings: they are largest for high-yield corporate bonds and smallest for Treasury and equity ETFs. Premiums are three-to-four times more likely to be positive than negative for most bond ETFs. Discounts (negative premiums) are generally larger in magnitude than positive premiums.

4.2 Persistent Premiums

Equation 4.2 shows that changes in premiums can only emerge from differences between ETF and NAV returns. To assess the degree to which NAV returns follow ETF returns, we estimate the following regression:

$$r_{NAV_e,t} = \alpha + \beta r_{e,t} + \varepsilon_{e,t}. \tag{4.3}$$

[11] Equation 4.2 is exact if we multiply the difference in ETF and NAV returns by a factor $\frac{P_{e,t-1}}{NAV_{e,t}}$, which is approximately one empirically.

Table 3: Premiums

(a) NAV Returns Imperfectly Follow ETF Returns

	Treasury	Short Term	Inv. Grade	High Yield	Broad Debt
β	0.941	0.783	0.897	0.545	0.805
s.e.	0.010	0.031	0.019	0.021	0.020
R^2	0.889	0.688	0.770	0.544	0.708
obs.	5008	882	1708	2502	1277

(b) Premiums Follow an AR(1)

	Treasury	Short Term	Inv. Grade	High Yield	Broad Debt
AR(1) coefficient	0.086	0.604	0.572	0.584	0.468
Average s.e.	0.039	0.040	0.034	0.032	0.032

Notes. The table shows the coefficients from regression of NAV returns on ETF returns for all ETFs within a category (Panel (a)) and the average AR(1) coefficients for bond ETF premiums averaged over all bonds within a category. The sample includes bond ETFs holding U.S. corporate bonds or Treasuries with over \$100M as of September 2017. "Treasury" is all Treasury ETFs, "Broad Debt" is all bond ETFs that track the aggregate bond market and hold different types of bonds (including government bonds, corporate bonds, and mortgage-backed securities). "Short Term" is ETFs holding short-term (1-5 yr) investment grade corporate bonds, "Inv. Grade" is ETFs holding investment grade bonds with maturity longer than 5 years, and "High Yield" is ETFs holding high-yield corporate bonds. See Section 2.2 for more details on the data.

If NAV lags behind ETF returns, $0 < \beta < 1$. Alternatively, if NAV returns are on average equal to ETF returns, $\beta = 1$. Moreover, $\beta = 1$ and an $R^2 = 1$ suggest that premiums are very rare.

To analyze the persistence in premiums, we also estimate an AR(1) model:

$$\pi_{e,t} = \alpha + \psi \pi_{e,t-1} + \varepsilon_{e,t}. \tag{4.4}$$

If NAV returns follow ETF returns imperfectly (i.e., $0 < \beta < 1$), we expect $0 < \psi < 1$. If ETF returns are closely matched by NAV returns (i.e., $\beta = 1$), we expect ψ close to 0.

Table 3 presents the regression and AR(1) estimates for our subsample of 10 ETFs. Panel (a) presents the results from regressing NAV returns on ETF returns. Notably, β and the R^2 are largest for Treasury ETFs, for which baskets are essentially identical to holdings. For these funds, ETF returns are closely matched by NAV returns. Other bond ETFs, especially high-yield bond ETFs, have smaller β and R^2, which suggests that premiums are likely to persist for these funds. Panel (b) presents the AR(1) coefficients on ETF premiums and shows that premiums are more persistent for non-Treasury bond ETFs, consistent with the idea that ETFs with fractional baskets have persistent premiums.

4.3 NAV response to Large Premiums/Discounts

The facts above show that bond ETFs that hold less liquid assets have imperfect NAV adjustment, which contributes to persistent premiums. However, premiums could also be created and eliminated by ETF price overreaction and reversal as opposed to NAV adjustment only. For instance, a premium could materialize because the ETF price increases due to temporary price pressure; once that price pressure dissipates, the ETF price declines and eliminates the premium. In other words, premiums may materialize and persist because of ETF overreaction as opposed to NAV “catching up” to the ETF.

To test whether persistent premiums are related to ETF overreaction, we examine events with significant premiums and discounts in an event-study setting. We use our subsample of 10 ETFs, and study days with premiums below the 5th percentile (i.e., discounts) and above the 95th percentile for categories of ETFs that have persistent premiums: short term, investment grade, high-yield corporate bond ETFs and broad debt ETFs. We ensure there is no overlap between episodes within an ETF.

We examine the average ETF price and NAV in the 5 days before and after large premium/discount days (“event days”). We normalize the ETF price and NAV to 1 starting from 5 days before the event day and average across events. We also examine average cumulative fund flows in these episodes. Large premiums and discounts are unlikely to be exogenous since factors that cause large premiums and discounts may also be related to arbitrage frictions. The purpose of the event study is to examine the typical pattern in ETF prices in the days following large premiums or discounts and not to establish causality.

Figure 4.2 presents the results. Panel (a) shows that (positive) premiums are generated by an imperfect NAV response and there is little evidence that ETF overreaction contributes to premiums. After the initial spike in premiums on day 0, ETF prices are relatively stable on days 1-5 and do not show a significant reversal. Most of the adjustment in premiums is generated by NAV changes on days 1-5. Panel (a) also shows that flows aim to correct premiums: positive flows mean that there are ETF creations. This fact suggests that fractional baskets could help explain persistent differences between NAV and the ETF price.

Panel (b) presents the results for discounts (i.e., negative premiums) and paints a different picture. On the one hand, discounts appear to persist due to imperfect NAV adjustment, similar to premiums. On the other hand, part of the discount is corrected by ETF reversal as illustrated by the upward slope in ETF prices on days 1-5. This means that discounts are related to initial ETF overreaction, which contributes to deeper discounts than what is justified by imperfect NAV adjustment. Panel (b) also shows that arbitrage flows aim to correct discounts: negative flows mean that there are ETF redemptions. These results suggest that ETF arbitrage works similarly for premiums and discounts and results in imperfect NAV adjustment. However, significant discounts are in part generated by overreaction in ETF prices, in contrast to ETF premiums which do not appear to be driven by ETF overreaction.

4.4 Discussion

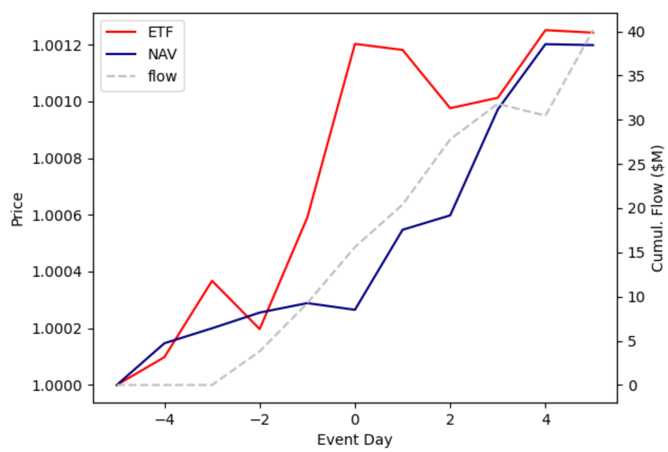
This section collectively paints a picture of delayed NAV reaction to ETF price changes, which leads to persistent premiums and slow adjustment of NAV. One interpretation of these facts is that slow NAV adjustment and persistent premiums are a *feature* of ETFs as opposed to a "design flaw" of ETFs holding illiquid assets. Since rapid selling of illiquid assets can result in fire sales and potentially in contagion (Shleifer and Vishny, 2011; Coval and Stafford, 2007), it may be optimal for bond ETFs to avoid selling illiquid assets at times when their prices are depressed. The mechanism of slow NAV adjustment due to fractional baskets could then act as a buffer to absorb fire sale shocks and would not transmit the selling pressure from the ETF market to underlying holdings.

This feature of bond ETFs stands in sharp contrast to other investment vehicles like mutual funds (MFs). Redemptions can be a significant cost for MFs holding illiquid assets. The funds must either sell illiquid assets to meet redemptions (which can depress NAV) or hold cash/liquid assets as a buffer to satisfy redemptions and avoid fire sales (which increases tracking error). In contrast, ETF redemptions due to fire sales of ETF shares in the secondary market would lead to lower ETF prices but a delayed response in NAV due to its slow adjustment. As a result, fire sales can occur in the more liquid ETF market and increase ETF discounts, but would not affect ETF holdings immediately.

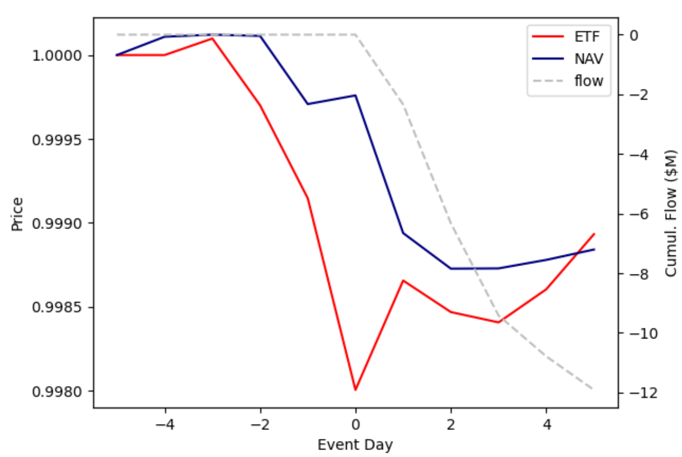
ETF investors who panic or are forced to redeem for liquidity reasons then do so at a cost to themselves, which is reflected in the discount. Non-redeeming investors, investors in the underlying assets, and firms who issue bonds, do not bear this cost and are, effectively, more insulated from fire sales in an ETF than in a MF. As a result, ETFs are able to hold less in cash/liquid assets, and spread the selling pressure in illiquid assets over multiple periods, which protects existing investors. This is the sense in which ETFs offer a natural "liquidity buffer" for those seeking redemptions in periods of market distress. In the next section, we

Figure 4.2: Premium and Discount Event Study

(a) Premiums



(b) Discounts



Notes. This figure plots the average response in ETF prices and NAV for all short term, investment grade, high-yield, and broad debt bond ETFs from September 2017 to February 2020. We identify the top and bottom 5% of premiums (the bottom 5% represents discounts) for each bond ETF, and examine average ETF price and NAV 5 days before and after the large-in-magnitude premium day. We normalize all prices to the first day of each episode window, and report averages across all days and ETFs for these episodes. We also report average cumulative flows averaged over all episodes. Prices are plotted in solid red (ETF) and blue (NAV), and correspond to the left axis which shows normalized prices. The dashed grey line plots cumulative flows and corresponds to the right axis.

present a simple model to illustrate these effects.

5 Model

We build a simple model to explain how ETFs can mitigate fire sales in relatively illiquid assets (e.g., corporate bonds) by allowing an AP to act as a shock absorber. The AP in our model is both an arbitrageur in the ETF and a dealer holding inventory in the underlying assets. The AP balances mark-to-market losses on her inventory with additional holding costs. We show that this type of AP acts as a buffer during fire sale episodes since she endogenously holds bonds instead of selling them at fire sale prices. That is, the AP acts as a stabilizing force since she has "skin in the game."

5.1 Model Setup

There are three securities in our model: two bonds, denoted A and B, and an ETF that holds an equally-weighted portfolio of the two bonds. Initially, the price of the bonds and the ETF are all equal to P_0. The NAV of the ETF is given by the weighted average of the holdings, i.e., initially it is equal to the ETF price P_e ($NAV = P_e = P_0$). We focus on ETF redemptions, and assume that the ETF redemption basket contains only one of the two bonds, motivated by the fractional baskets we document in Section 3.1.1.[12]

There are two main agents in the model. The first is an uninformed investor who exogenously needs to sell the ETF in the secondary market. The second is a representative risk-neutral AP who provides liquidity in the ETF market by trading with the investor. The AP is representative in the sense that she acts as if there is perfect competition among many APs to provide liquidity to the uninformed investor in the secondary market. Specifically, the AP trades with the investor at the maximum ETF price that yields zero profits.[13] In addition to the secondary ETF market, there are two other markets: a primary ETF market where the AP redeems ETF shares for the redemption basket, and a bond market where individual bonds are traded.

[12]We also account for baskets that have the same weights as the ETF holdings in Proposition 2 below. We also study the case of creations in Section B.3 of the Appendix.

[13]It is relatively trivial to include additional revenue from providing liquidity in the secondary market if there is imperfect competition among APs.

There is a single trading period, in which the investor submits a market order in the secondary ETF market to sell $q < 0.5$ units of the ETF, where q is the amount sold as a fraction of the total AUM of the ETF.[14] The AP sets the ETF price at which she is willing to buy the q units of the ETF from the investor. We also assume that the AP then immediately redeems all of newly purchased ETF units for the redemption basket and adds these bonds to her inventory position.[15]

The AP initially holds an equal amount of inventory in each bond A and B, i.e., $z_A = z_B = z > 0$. We assume the initial inventory is ex-ante optimal given the inventory needs of the AP as a bond dealer. The AP buys q units of the ETF in the secondary market at price P_e, determined below, and receives bonds in the redemption basket initially worth NAV_{basket}. The AP then chooses to keep a fraction γ of the redeemed bonds in her inventory and sells the remaining fraction $1 - \gamma$ in the bond market.

The AP incurs three types of costs associated with holding and trading bonds. First, she faces *price impact costs* when trading bonds on the bond market. When she sells a bond, the price of that bond falls immediately to $P_i - C((1 - \gamma) \cdot q)$ where i denotes the bond (A or B) and the cost function $C(x)$ is linear in the quantity of bonds traded. We capture the illiquidity of bonds in the spirit of Shleifer and Vishny (1992) by assuming that the bond market can only absorb a certain quantity of bonds up to a threshold τ. Trading more than τ leads to a fire sale in the sense that the impact on bond prices is much greater . Specifically, the price impact cost is given as

$$C(x) = \begin{cases} cx & , x \leq \tau \\ fx & , x > \tau \end{cases}, \tag{5.1}$$

where $f > c > 0$, and f is the marginal cost in the case of a fire sale. Thus, selling more than τ leads to a discontinuous jump in the price impact cost. This assumption captures the idea that if the AP trades bonds in "typical" amounts, she is able to find "specialist" counterparties who charge lower transaction costs, i.e., are willing to pay higher prices. However, if the AP needs to trade bonds in large amounts (i.e., above τ), she needs to trade

[14] In other words, the initial AUM of the ETF is the numeraire.

[15] In reality, APs may balance inventory between the ETF and the underlying assets. For example, in the case where the AP buys shares of the ETF in the secondary market, the AP may redeem a fraction of the shares for the redemption basket. In principle, we could capture this tradeoff by modeling not only bond holding costs (discussed further below) but also ETF holdings costs.

with "non-specialists" who offer lower prices for the bond (and thus the AP realizes greater price impact costs).[16]

The idea builds on Shleifer and Vishny (1992), where, in times of fire sales and when specialists are constrained, distressed firms need to sell their assets to outsiders who have lower valuations. In our model, we assume that there is only one specialist buyer: the AP cannot split the large amount to several buckets of size τ and sell them to different specialist buyers. The discontinuity in the cost function at τ comes from the assumption that the price impact cost of all units (including inframarginal units below τ) is f. There are two reasons for this. First, this specification captures the possibility that specialist buyers are constrained in times of fire sales ($x > \tau$) and the AP has to sell the whole amount to non-specialists. Second, the specification is a reduced form way of capturing a jump at τ that comes from greater fixed costs associated with units sold to non-specialists.[17]

The second cost the AP faces is a *holding cost* of additional inventory. We assume the cost of increasing holdings by x is $\frac{\lambda}{2}x^2$ with $\lambda > 0$; if $x < 0$, there are no additional costs. Since we assume that the AP's initial inventory level z is optimal, these costs arise due to positive *deviations* from this level. This specification of holding costs is similar to the one in Pan and Zeng (2021), and captures the idea of corporate bond inventory costs (Goldstein and Hotchkiss (2020) and Bessembinder et al. (2018)) and balance sheet capacity costs (Andersen et al. (2019)). We make an additional restriction that λ is bounded above by $\bar{\lambda}$ (the formal expression is provided in Appendix Section B.1). If λ is above this bound, holding costs are so large that the AP always prefers to sell, even at fire sale prices, rather than to hold redeemed bonds. Since fire sales are usually thought to be relatively large, this restriction allows us to focus on cases where fire sales impose a significant cost. The case where λ is above $\bar{\lambda}$ may represent a scenario where APs are so constrained that they are unwilling to act as a buffer between the ETF and the bond market.

The third type of cost that the AP incurs is a *mark-to-market cost*, which arises from adjustments to the valuation of the AP's bond inventory. This is the sense in which the

[16]This idea is consistent with BlackRock's own description of the types of APs that participate in the primary market. In a webinar to discuss ETF primary market activity held on May 10, 2021, BlackRock stated that dealers specializing in particular types of bonds were more likely to conduct primary market activity than non-specialized dealers, especially in recent years (BlackRock, 2021).

[17]Another way to interpret this function is that the marginal cost of additional units sold above τ is decreasing. An alternative specification is to make the marginal cost of all bonds sold after τ constant, in which case the cost would be $f + c$. We adopt the simpler version of cost presented in the main text since the alternative specification adds significant complexity without changing the main intuition.

AP internalizes the cost of lower bond prices. Since the AP holds some initial inventory in bonds, lower bond prices decrease the mark-to-market value of the AP's bond holdings. This cost represents a decline in the AP's borrowing capacity (or equivalently, an increase in leverage and a decrease in capital ratios). We model this cost as the total change in value of the initial bond inventory plus the newly acquired bonds that are held after redemption, or $(z + \gamma q)\Delta P_i$.

5.2 ETF Redemption

In the spirit of fractional baskets, we assume that the redemption basket only contains bond A. In the case of ETF redemption of q shares, the AP earns an arbitrage profit of $NAV_{basket} - P_e = P_0 - C((1-\gamma)q) - P_e$ per unit. That is, the arbitrage profit is the difference between the selling price of the bonds (initial NAV of the basket inclusive of price impact costs) and the purchase price of the bonds (ETF price).[18] The APs' total payoff is:

$$\underbrace{(NAV_{basket} - P_e) \cdot q}_{\text{arbitrage profit}} - \underbrace{\frac{\lambda}{2} \cdot (\gamma q)^2}_{\text{holding cost}} - \underbrace{(z + \gamma q) \cdot \Delta P_A}_{\text{mark-to-market cost}} \tag{5.2}$$

where $\Delta P_A = C\left((1-\gamma)\, q\right)$.

It is useful to analyze AP's payoff by initially assuming that price impact costs are c regardless of the traded quantity. Specifically, denote γ_c^* as the equilibrium fraction of bonds held (the fraction sold is then $1 - \gamma_c^*$) by the AP who acts as if price impact costs are fixed at c. γ_c^* is useful in analyzing our two cases. The first case is a "typical" redemption case, in which the equilibrium fraction of bonds sold is below the fire-sales threshold τ. This case represents the region where the selling fraction $1 - \gamma_c^*$ yields a selling quantity that is below τ and there are no fire sales. The second case is the fire sale case: if the AP were to sell the equilibrium fraction of bonds acting as if the price impact cost was c, she would sell more than τ and incur a cost of f rather than c. That is, the typical redemption selling amount would lead to a fire sale and the AP would need to take into account the significant increase in price impact costs when choosing γ. We describe each case below.

[18] Given that creation/redemption fees are typically very small, we assume they are zero for simplicity. The logic is unchanged with constant non-zero fees.

Case 1: No Fire Sale Region $(1-\gamma_c^*)q \leq \tau$

In this region, the AP incurs price impact costs of $c < f$ when selling bonds. Since the representative AP is disciplined by perfect competition, we set her payoff from equation 5.2 to zero. Rearranging equation 5.2 yields an expression for the ETF price P_e as a function of the AP's choice of γ:

$$P_e(\gamma) = P_0 - c(q+z) + cz\gamma - \left(\frac{\lambda}{2} - c\right) q\gamma^2. \tag{5.3}$$

In keeping with competition, the AP selects γ such that she maximizes the ETF price.[19]

There are two key conditions for the equilibrium holding fraction γ_c^* to be interior, i.e., γ_c^* between 0 and 1. First, holding costs $\left(\frac{\lambda}{2}\right)$ must be greater than the non-fire sale price impact cost (c). Second, AP's inventory z must be below the threshold $q\frac{(\lambda-2c)}{c}$. If $\frac{\lambda}{2} \leq c$ (holding costs are too low) or $z \geq \frac{\lambda-2c}{c}q$ (inventory is too large), the AP holds all bonds (i.e., $\gamma_c^* = 1$). Otherwise, γ_c^* is interior and is given by

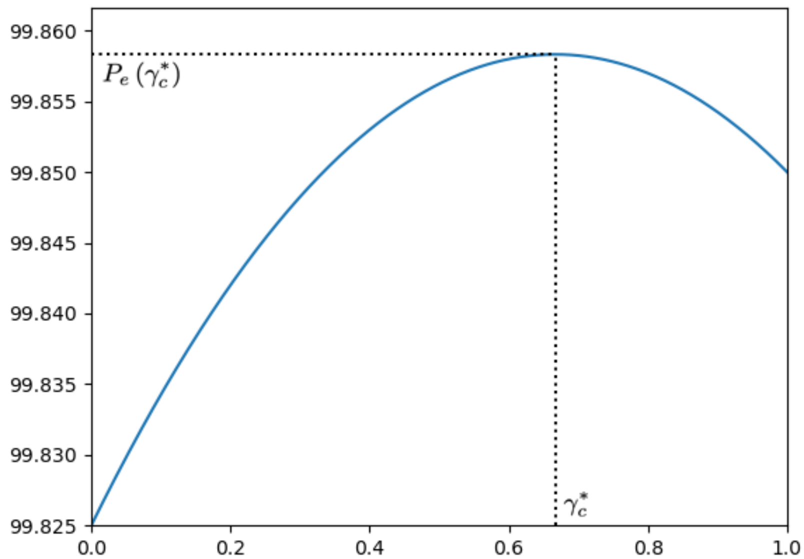
$$\gamma_c^* = \frac{cz}{(\lambda - 2c)\,q}, \tag{5.4}$$

and is increasing in the ratio of the initial inventory z to the redeemed quantity q. This captures the tradeoff that the AP faces: reducing mark-to-market losses (which are increasing in z) vs. reducing excess holding costs (which are increasing in q). The larger the inventory relative to the redeemed amount, the larger the fraction of redeemed bonds that the AP decides to hold to avoid the cost from lower bond prices. Figure 5.1 provides an example of an interior γ_c^* and the resulting ETF price in the case of a typical (non-fire sale) redemption.

The characterization of the equilibrium in this non-fire sale case depends on q being sufficiently small such that the equilibrium selling quantity $(1-\gamma_c^*)\,q$ is smaller than the fire-sale threshold τ. If this condition is not satisfied, there is a fire sale and the price impact costs jump from c to f. As a result, the ETF price given by γ_c^* is not achievable. Also, note that the quantity redeemed q can be greater than τ but a fire sale can still be avoided if $(1-\gamma_c^*)\,q < \tau$. This is possible since the AP sells only a fraction of the redeemed bonds.

[19] The logic is consistent with market clearing in a competitive market in which an investor sells q in the secondary ETF market and APs compete to purchase q, which yields the highest price for the investor and no profits for APs. It is straightforward to extend the model with non-zero AP profits by adding an additional term that is constant or scales with q but not γ in equation 5.3.

Figure 5.1: Typical Redemption



Notes. This figure plots the ETF price $P_e(\gamma)$ as a function of γ (the share of redeemed bonds that the AP puts on her balance sheet) and shows the equilibrium holding fraction when price impact costs are fixed at c: γ_c^*. For this example, we use $q = 0.075$, $z = 0.1$, $c = 1$, $\lambda = 4$, $P_0 = 100$, and $\tau = 0.1$. See the main text for more details on the shape of the function and the conditions for an interior γ_c^* .

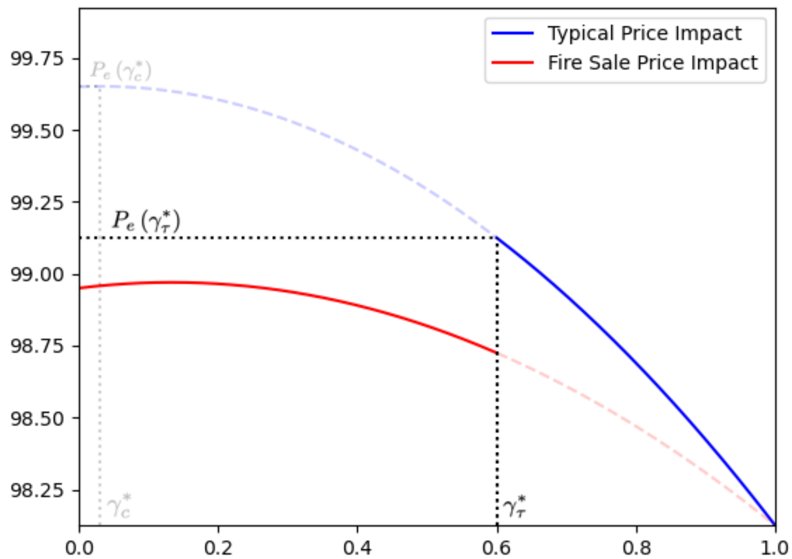
Case 2: Fire Sale Region $(1 - \gamma_c^*)q > \tau$

We now consider the fire sale case when $(1 - \gamma_c^*)q > \tau$. In this fire sale region, the AP must adjust her holding quantity from γ_c^* to an alternative quantity which we denote γ_τ^*. The AP does this because γ_c^* is unachievable when it implies a selling quantity greater than τ. The alternative quantity γ_τ^* must take into account the increase in price impact costs from c to f when selling more than τ.

Since for each possible value of $\gamma \in [0, 1)$, the ETF price is greater with price impact costs c than with costs f, the AP chooses γ_τ^* to be the smallest holding fraction that satisfies $(1 - \gamma_\tau^*)\, q \leq \tau$, or $\gamma_\tau^* = 1 - \frac{\tau}{q}$. That is, the AP maximizes the ETF price by selling up to the threshold τ in order to avoid the discontinuous jump in price impact costs, and the resulting drop in the ETF price. The intuition behind this choice is simple. Since the AP holds initial inventory in the redeemed bonds, she internalizes the cost of fire sales. In addition, even in the absence of inventory, competition pushes the AP to avoid a fire sale because she is able to quote a higher ETF price.

Figure 5.2 provides an example of the case where the AP adjusts her holding fraction

Figure 5.2: Fire Sale Redemption



Notes. This figure plots the ETF price $P_e(\gamma)$ as a function of γ (the share of redeemed bonds that the AP keeps on her balance sheet) in the case when $(1-\gamma_c^*)q > \tau$. The solid lines represents the ETF price, which is discontinuous when the selling quantity $(1-\gamma)q$ is greater than the fire sale threshold τ. The equilibrium holding fraction in the fire sale scenario γ_τ^* is the minimum amount held such that the price impact cost is still $c < f$. For this example, we use $q = 0.25$, $z = 0.1$, $c = 1$, $f = 3$, $\lambda = 15$, $\tau = 0.1$, and $P_0 = 100$. See the main text for more details.

to avoid a fire sale. The graph shows that γ_c^* would yield the highest ETF price but is not achievable because selling $(1-\gamma_c^*)q$ triggers a jump in price impact costs from c to f and a drop in the ETF price. As a result, the AP chooses γ_τ^*, the smallest γ possible to avoid a fire sale and retain price impact costs of c.[20]

Equilibrium

Combining the fire-sale and non-fire-sale regions allows us to describe the equilibrium.

Proposition 1. *(Equilibrium AP Bond Sales) When bond holding costs λ are below $\bar{\lambda}$, the equilibrium fraction of the redeemed bonds that are held by the AP in inventory, γ^*, is given by the following:*

[20]The assumption that $\lambda < \bar{\lambda}$ can also be explained in the context of Figure 5.2. If $\lambda = \bar{\lambda}$, then the maximum of the curve representing the fire sale price impact is exactly equal to the ETF price from selecting γ_τ^*. As λ tends towards infinity, the two curves mimic each other since holding costs become the dominant factor and price impact costs become irrelevant.

1. *In the case where* $(1-\gamma_c^*)q \leq \tau$ *(a "typical" redemption),* $\gamma^* = \gamma_c^*$ *where*

$$\gamma_c^* = \frac{cz}{(\lambda - 2c)\, q}.$$

2. *In the case where* $(1-\gamma_c^*)q > \tau$ *(a "fire sale" redemption),* $\gamma^* = \gamma_\tau^*$ *where*

$$\gamma_\tau^* = 1 - \frac{\tau}{q}.$$

The ETF price is

$$P_e^* = P_0 - c(q+z) + \gamma^* cz - (\gamma^*)^2 \left(\frac{\lambda}{2} - c\right) q,$$

the NAV of the ETF is

$$NAV^* = P_0 - \frac{0.5-q}{1-q}(1-\gamma^*)cq, \tag{5.5}$$

and the ETF discount is

$$-\pi^* = -\left(P_e^* - NAV^*\right) = (\gamma^*)^2(\frac{\lambda}{2} - c)q + (1-\gamma^*)zc + \frac{0.5 + \gamma^*(0.5-q)}{1-q}cq. \tag{5.6}$$

Discussion of the Equilibrium In this subsection we discuss three important observations that follow from Proposition 1.

First, the AP's optimal holding fraction γ^* always avoids selling quantities that lead to a fire sale. The intuition for this result is straightforward – the AP internalizes the cost of a fire sale because she holds inventory in the redeemed bonds and triggering a fire sale would result in costly mark-to-market losses.

Second, we assume that competition disciplines the AP and results in her setting the highest ETF price possible but this does not necessarily minimize the ETF discount. This is contrary to conventional wisdom. In the textbook arbitrage case, which implicitly assumes no inventory and no transaction costs, the AP eliminates ETF discounts (when $P_e < NAV$). Our model highlights that transaction costs affect the mark-to-market value of the AP's inventory and may prevent the AP from eliminating the discount completely.

Third, the AP essentially acts as a liquidation agent on behalf of ETF investors. The AP sells the redeemed bonds in a manner that protects existing ETF investors because she is also a holder of the bonds. Moreover, since the ETF discount is a reflection of the AP's costs, redeeming ETF investors ultimately pay these costs through a lower ETF price.[21]

The fact that the AP acts as a shock absorber suggests that ETFs can be an important mechanism to insulate illiquid bonds from fire sales. Perhaps the most significant economic consequence is that firms could potentially refinance easier and at lower cost in periods of stress, since bond markets can continue to function normally and bond yields would be lower than in the case when large ETF redemptions propagate to the underlying bonds. Another benefit of the AP acting as a shock absorber is that bond investors are also insulated from large ETF redemptions. In effect, the AP acts in their interest since she also holds bonds in inventory and takes into account the costs associated with lower mark-to-market values.

Full Baskets and Mutual Fund Comparison It is important to analyze the case when baskets are identical to holdings (full baskets). We present the equilibrium with full baskets below, and compare it to Proposition 1. The full baskets case is also useful to compare the ETF with a stylized mutual fund.

Proposition 2. *(Full Baskets) In the case of full baskets, the equilibrium ETF price is larger and the ETF discount and NAV are smaller when compared to the case with fractional baskets.*

The ETF puts equal amounts of bonds A and B in the basket ($q_A = q_B = \frac{q}{2}$) and the AP chooses $\gamma_A = \gamma_B = \gamma^*$. The ETF price is

$$P_e^* = P_0 - \frac{c(q+z) - \gamma^* cz + (\gamma^*)^2(\frac{\lambda}{2} - c)q}{2},$$

the NAV of the ETF is

$$NAV^* = P_0 - \frac{c(1-\gamma^*)q}{2}, \tag{5.7}$$

and the ETF discount is

$$-\pi^* = \frac{1}{2}[(\gamma^*)^2(\frac{\lambda}{2} - c)q + (1-\gamma^*)cz + cq\gamma^*]. \tag{5.8}$$

The proposition states that having full instead of fractional baskets decreases the ETF

[21] In Appendix B.2, we also consider a two-period extension, which shows that the ETF price increases in the second period because the AP is able to liquidate the excessive inventory without triggering a fire sale. This reduces AP's holding costs and narrows the ETF discount. This result is similar to the pattern shown in Figure 4.2, where large discounts are in part due to ETF price overreaction and subsequent reversal.

discount because it raises the ETF price and lowers NAV. In the context of our model, we interpret the discount as a type of "buffer" in the sense that a larger discount represents less selling pressure transmitted from the ETF to the underlying assets. Fractional baskets provide a larger buffer than full baskets for two reasons. First, holding costs are larger with fractional baskets since they are concentrated in one bond versus spread between two, and the cost is quadratic per bond. These larger costs are passed along to redeeming ETF investors in the form of a larger discount. Second, with fractional baskets, the AP concentrates selling in one bond instead of two, and that bond is precisely the one that has a lower weight in the ETF holdings after the redemption. That is, the ETF sheds bonds that are sold by the AP and overweights bonds that are not, which leads to a greater NAV.[22]

In addition, it is useful to understand that the discount depends on holding costs, price impact costs, and the initial inventory. Equations 5.6 and 5.8 show that $-\partial\pi^*/\partial\lambda > 0$, $-\partial\pi^*/\partial c > 0$ and $-\partial\pi^*/\partial z > 0$. First, these facts suggest that greater costs or greater inventory lead to larger discounts. Quite intuitively, larger costs for the AP require greater compensation in the form of lower ETF prices (larger arbitrage profits). Second, to the extent that we allow for differences between bond A and B along the dimensions of price impact costs, holding costs, or initial inventory, we can see that the choice of which bond to place in the (fractional) basket has an impact on the ETF discount. Selecting the bond that has larger costs, or the bond that is more widely held in inventory, leads to larger discounts. In Section 3.2 we show that redemption baskets are more likely to contain more liquid bonds, and the model shows that this could contribute to reducing ETF discounts.[23]

Finally, we use the full baskets case to consider a stylized mutual fund and analyze how that fund handles redemptions in case of fire sales. In the context of our model, we think of a mutual fund as an ETF with full baskets and where the AP has zero inventory ($z = 0$) and always sells all bonds to meet redemptions ($\gamma^* = 0$). In this case, the ETF price is equal to NAV ($P_e^* = NAV^*$) as can be seen from Proposition 2. This is because redemptions of mutual fund shares are directly transmitted to the underlying bonds since there is no intermediary to internalize the cost of price impact, i.e. there is no buffer. The mutual fund price, NAV, and discount are always lower than those of any ETF with an AP who

[22]Section 3.2.2 provides empirical support for this fact since ETFs are more likely to include underperforming bonds in the redemption basket.

[23]There may be other factors that can explain why more liquid bonds tend to be in redemption baskets. See Section 3.2 for more discussion.

holds non-zero inventory. This is simple to see in Figure 5.2, where the mutual fund always achieves the price given by $\gamma = 0$.[24]

6 Conclusion

We document several new facts about bond ETFs. First, bond ETFs holding U.S. corporate bonds utilize "fractional baskets" in the sense that creation/redemption baskets only contain a subset of holdings. This challenges the assumption that baskets are similar if not exactly identical to holdings and suggests that ETF arbitrage may be less effective in eliminating premiums and discounts for bond ETFs. Second, bond ETF baskets tend to exhibit high turnover. Third, baskets differ from holdings along several dimensions. Typically, creation baskets contain bonds with longer duration and wider bid-ask spreads compared to holdings; redemption baskets contain bonds with shorter duration and narrower bid-ask spreads compared to holdings. Fourth, ETFs with fractional baskets have persistent premiums and discounts, which is related to the slow adjustment of NAV to ETF price movements. In addition, large discounts tend to arise in part from ETF price overreaction, resulting in a temporarily depressed ETF price.

We interpret the discrepancies that can arise between ETF price and NAV as a potential benefit of ETFs that manage illiquid assets. Bond ETFs can absorb large redemptions without transmitting the selling pressure to the more fragile bond market because of slow NAV adjustment. To illustrate this effect, we present a simple model of ETFs holding illiquid bonds during fire sales. The model shows that when the AP holds initial inventory, she acts as a buffer between the ETF market and the bond market. The key insight is that the AP internalizes the cost of a fire sale since she incurs mark-to-market losses on her existing inventory. As a result, she optimally decides to avoid a fire sale by adding a larger fraction of redeemed bonds in her inventory.

Our findings suggest that ETFs may be better suited for managing a portfolio of illiquid assets compared to other investment vehicles like mutual funds. Mutual funds may be forced to sell illiquid assets to satisfy large redemptions which could trigger a fire sale, potentially leading to distress in the underlying bond market and impeding fundraising efforts.

[24] To prevent fire sales in reality, mutual funds may hold a larger share of liquid assets or cash. This could increase the tracking error and create cash drag. We interpret this choice as a cost to mutual fund investors related to the risk of a fire sale.

References

Andersen, Leif, Darrell Duffie, and Yang Song, "Funding Value Adjustments," *Journal of Finance*, 2019, *74*, 145–192.

Aramonte, Sirio and Fernando Avalos, "The recent distress in corporate bond markets: cues from ETFs," *BIS Bulletin No 6*, April 2020.

Bao, Jack, Jun Pan, and Jiang Wang, "The Illiquidity of Corporate Bonds," *Journal of Finance*, 2011, *66*, 911–946.

Becker, Bo and Efraim Benmelech, "The Resilience of the U.S. Corporate Bond Market During Financial Crises," *NBER Working Paper*, 2021.

Ben-David, Itzhak, Francesco Franzoni, and Rabih Moussawi, "Do ETFs Increase Volatilty," *The Journal of Finance*, 2018, *73* (6), 2471–2535.

Bessembinder, Hendrik, Chester Spatt, and Kumar Venkataraman, "A Survey of the Microstructure of Fixed-Income Markets," *Journal of Financial and Quantitative Analysis*, 2020, *55.*

_ , Kathleen M., William F. Maxwell, and Danielle Xu, "Measuring Abnormal Bond Performance," *Review of Financial Studies*, 2009, *22*, 4219–4258.

_ , Stacey Jacobsen, William Maxwell, and Kumar Venkataraman, "Capital Commitment and Illiquidity in Corporate Bonds," *Journal of Finance*, 2018, *73*, 1616–1661.

BlackRock, "By the numbers: the bond ETF primary market process," 2021.

Boyarchenko, Nina, Anna Kovner, and Or Shachar, "It's What You Say and What You Buy: A Holistic Evaluation of the Corporate Credit Facilities," *FRB of New York Staff Report No. 935*, 2020.

Brogaard, Jonathan, Davidson Heath, and Da Huang, "The Rise of ETF Trading and the Bifurcation of Liquidity," *Working paper*, 2019.

Coval, Joshua and Erik Stafford, "Asset fire sales (and purchases) in equity markets," *Journal of Financial Economics*, 2007, *86* (2), 479–512.

Da, Zhi and Sophie Shive, "Exchange traded funds and asset return correlations. 24, 136–168," *European Financial Management*, 2018, *24*, 136–168.

D'Amico, Stefania, Vamsidhar Kurakula, and Stephen Lee, "Impacts of the Fed Corporate Credit Facilities through the Lenses of ETFs and CDX," *Working paper*, 2020.

Dannhauser, Caitlin, "The Impact of Innovation: Evidence from Corporate Bond ETFs," *Journal of Financial Economics*, 2018, *125*, 537–560.

Gilchrist, Simon, Bin Wei, Vivian Z Yue, and Egon Zakrajšek, "The Fed takes on corporate credit risk: an analysis of the efficacy of the SMCCF," *BIS Working Paper*, 2021, *963.*

Goldstein, Michael A. and Edith S. Hotchkiss, "Providing liquidity in an illiquid market: Dealer behavior in US corporate bonds," *Journal of Financial Economics*, 2020, *135* (1), 16–40.

Haddad, Valentin, Alan Moreira, and Tyler Muir, "When Selling Becomes Viral: Disruptions in Debt Markets in the Covid-19 Crisis and the Fed's Response," *Review of Financial Studies, forthcoming*, 2020.

Holden, Craig and Jayoung Nam, "Market Accessibility, Corporate Bond ETFs, and Liquidity," *Working paper*, 2019.

Jin, Dunhong, Marcin T. Kacperczyk, Bige Kahraman, and Felix Suntheim, "Swing Pricing and Fragility in Open-end Mutual Funds," *The Review of Financial Studies*, 2021, *Forthcoming.*

Kaminska, Izabella, "A little ETF rule change that could make a big difference," 2020.

Laipply, Stephen and Ananth Madhavan, "Pricing and liquidity of fixed-income ETFs in the COVID-19 virus crisis of 2020," *Journal of Index Investing*, Winter 2020.

Lettau, Martin and Ananth Madhavan, "Exchange-Traded Funds 101 for Economists," *Journal of Economic Perspectives*, 2018, *32*, 135–154.

Lewrick, Ulf and Jochen F. Schanz, "Is the Price Right? Swing Pricing and Investor Redemptions," *BIS Working Paper No. 664*, 2017.

Malamud, Semyon, "A dynamic equilibrium model of ETFs," *Working paper*, 2015.

Pan, Kevin and Yao Zeng, "ETF Arbitrage Under Liquidity Mismatch," *Working Paper*, 2021.

Saglam, Mehmet, Tugkan Tuzun, and Russ Wermers, "Do ETFs Increase Liquidity?," *Working paper*, 2019.

Shim, John J., "Arbitrage Comovement," *Working Paper*, 2020.

Shleifer, Andrei and Robert Vishny, "Liquidation Values and Debt Capacity: A Market Equilibrium Approach," *Journal of Finance*, 1992, *47* (4), 1343–1366.

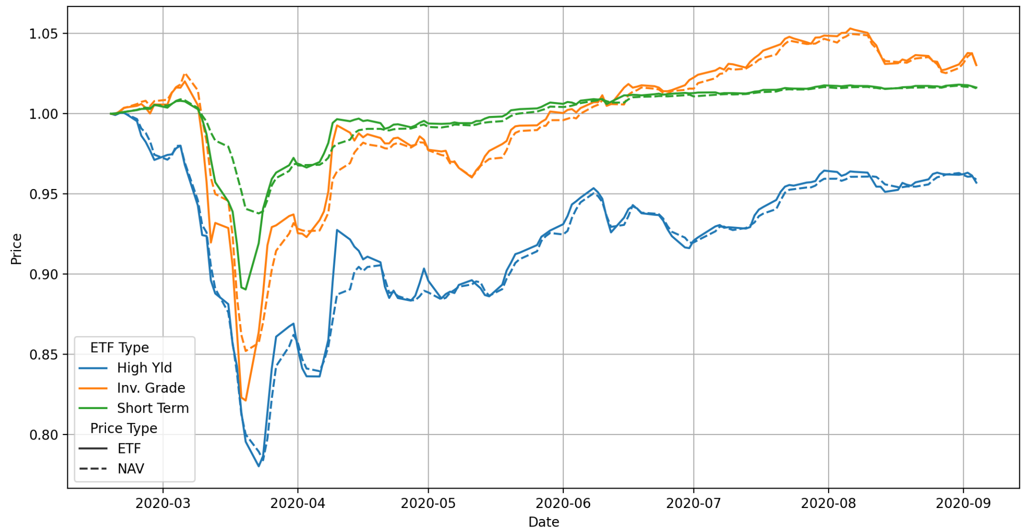
_ **and** _ , "Fire Sales in Finance and Macroeconomics," *Journal of Economic Perspectives*, 2011, *25* (1), 29–48.

Sushko, Vladyslav and Grant Turner, "The implications of passive investing for securities markets," *BIS Quarterly Review, March*, March 2018.

Todorov, Karamfil, "Passive Funds Affect Prices: Evidence from the most ETF-dominated Markets," *Working Paper*, 2019.

_ , "The anatomy of bond ETF arbitrage," *BIS Quarterly Review*, March 2021.

Figure A.1: ETF prices and NAV during COVID and Fed Purchase Periods



Notes. This figure plots ETF price and NAV (normalized to 1 on March 1, 2020) for our ETF subsample (described in Section 2.2) during the COVID period and post-COVID period when the Federal Reserve instituted an ETF Purchase Program. "Short Term" is ETFs holding short-term (1-5 yr) investment grade corporate bonds, "Inv. Grade" is ETFs holding investment grade bonds with maturity longer than 5 years, and "High Yld" is ETFs holding high-yield corporate bonds.

A Bond ETFs during COVID and the Federal Reserve Purchase Program

A.1 Background on COVID and the Fed intervention

In this section, we provide supporting evidence that the facts described in the main text and based on the pre-COVID data only, also hold during the COVID and post-COVID (Fed) periods.

In February-March 2020 the bond market saw a widespread decline as a result of the COVID shock. To stabilize markets, the Fed announced several measures. On March 23rd, the Fed introduced the Primary Market Corporate Credit Facility (PMCCF) and Secondary Market Corporate Credit Facility (SMCCF). The PMCCF involved purchases of investment grade corporate bonds or syndicated loans and the SMCFF was aimed at purchases of investment grade corporate bond ETFs. On April 9th the SMCCF was expanded to include high-yield corporate bonds and ETFs with actual purchases commencing on May 12th (see, e.g., Gilchrist et al. (2021) for a discussion of the effects for bonds). These interventions

Table 4: NAV Adjustment

(a) COVID Period

	Treasury	Short Term	Inv. Grade	High Yield	Broad Debt
$r_{e,t}$	0.812	0.207	0.399	0.568	0.302
s.e.	0.077	0.042	0.066	0.032	0.066
R^2	0.797	0.465	0.595	0.764	0.483
obs	128	48	64	64	47

(b) Federal Reserve Purchase Period

	Treasury	Short Term	Inv. Grade	High Yield	Broad Debt
$r_{e,t}$	0.904	0.155	0.329	0.509	0.444
s.e.	0.043	0.084	0.055	0.028	0.046
R^2	0.802	0.115	0.356	0.604	0.401
obs	1408	528	704	700	525

Notes. The COVID period is from March 1, 2020 to March 23, 2020. The Fed period is from March 24, 2020 until September 4, 2020. The results from Panel (a) should be take with caution given the low number of observations. See Section 2.2 for more details on the data, and Section 4 for more details on the regression specification.

coincided with strong reversals in ETF prices.

Figure A.1 depicts corporate bond ETF prices and NAVs during the COVID and Fed episodes. During the COVID decline in March 2020, bond ETFs experienced significant price drops and large discounts. These discounts were more pronounced for short term and investment grade corporate bonds. This is consistent with delayed adjustment of NAV and overreaction in ETF prices. Perhaps surprisingly, riskier high-yield bonds saw smaller discounts than safer investment grade bonds. The announcement and implementation of the Fed programs helped reverse ETF prices with slow adjustment followed by NAV.

A.2 Empirical Results

We also repeat the results of Section 4.2 by estimating the regressions in equation 4.3 for the COVID and Fed episodes. Table 4 presents the results. Panel (a) and (b) show that NAV adjustment during these episodes was generally even more muted than in the pre-COVID period (see Table 3).

B Theory Appendix

B.1 ETF redemption proofs

B.1.1 Proof of Proposition 1.

First, let us solve for γ_c^*. There are three cases when maximizing the ETF price in equation 5.3 if P_e is concave in γ ($\frac{\lambda}{2} > c$). First, $\gamma_c^* = 0$ when the AP has no inventory: $z = 0$. Second, $\gamma_c^* = 1$ when the AP has a large inventory position: when $z \geq \frac{\lambda-2c}{c}q$. Third, γ_c^* is interior (between 0 and 1) if the AP has an "interior" inventory position, $0 < z < \frac{\lambda-2c}{c}q$.

Since $P_e(\gamma)$ is concave function, for an interior solution, the optimal share of redeemed bonds that the AP should hold γ is found by taking first-order conditions (FOC):

$$2\gamma(c - \frac{\lambda}{2})q + cz = 0$$

$$\gamma_c^* = \frac{c}{\lambda - 2c}\frac{z}{q} \tag{B.1}$$

The optimal γ^* depends then on whether the parameters of the model give $0 < \gamma_c^* < 1$ (interior solution) and $(1 - \gamma_c^*)q < \tau$ (typical redemption). These conditions are satisfied if:

$$\begin{cases} cz < (\lambda - 2c)q \\ q - \tau < \frac{cz}{\lambda - 2c} \end{cases} \Leftrightarrow (\lambda - 2c)(q - \tau) < cz < (\lambda - 2c)q \tag{B.2}$$

If ($\frac{\lambda}{2} < c$), then $P_e(\gamma)$ is convex, $\gamma_c^* < 0$ and since the FOC gives the minimum, the function is increasing from 0 to 1. Hence, the optimal solution is always $\gamma^* = 1$.

Now, assume there is a fire sale: $(1 - \gamma_c^*)q > \tau$. For the AP to sell exactly fraction $\frac{\tau}{q}$ and to choose $\gamma_\tau^* = 1 - \frac{\tau}{q}$, we need $P_e(c, 1 - \frac{\tau}{q}) > P_e(f, \gamma)$. The condition is satisfied for any γ if $P_e(c, 1 - \frac{\tau}{q})$ is always greater than the max of $P_e(f, \gamma)$. There are two cases: $f > \frac{\lambda}{2}$ and $f \leq \frac{\lambda}{2}$

In the first case when $f > \frac{\lambda}{2}$, $P_e(f, \gamma)$ is convex in γ and the optimal solution is corner solution at $\gamma^* = 1$, since the function is increasing from 0 to 1. Given that $P_e(c, \gamma)$ is decreasing from $1 - \frac{\tau}{q}$ to 1 due to its concavity (since $\gamma_\tau^* = 1 - \frac{\tau}{q} > \gamma_c^*$), $P_e(c, 1 - \frac{\tau}{q}) > P_e(c, 1) = P_e(f, 1) = P_0 - \frac{\lambda}{2}q$. More formally:

$$P_e(c, 1 - \frac{\tau}{q}) > P_e(c, 1) = P_e(f, 1) = P_0 - \frac{\lambda}{2}q.$$

$$\Leftrightarrow -c(q+z) - \frac{(q-\tau)^2}{q}(\frac{\lambda}{2}-c) + \frac{(q-\tau)cz}{q} + \frac{\lambda}{2}q > 0$$
$$\Leftrightarrow (q - \frac{(q-\tau)^2}{q})(\frac{\lambda}{2}-c) - \frac{\tau cz}{q} > 0$$
$$\Leftrightarrow (\frac{2q\tau-\tau^2}{q})(\frac{\lambda}{2}-c) - \frac{\tau cz}{q} > 0$$
$$\Leftrightarrow (2q-\tau)(\frac{\lambda}{2}-c) > cz$$
$$\Leftrightarrow z < \frac{(\lambda-2c)}{c}(q-\frac{\tau}{2})$$

The latter condition is satisfied as $z < \frac{\lambda-2c}{c}(q-\tau)$ since $1-\frac{\tau}{q} > \gamma_c^* = \frac{c}{\lambda-2c}\frac{z}{q}$. Hence, the optimal choice for the AP is $\gamma_\tau^* = 1-\frac{\tau}{q}$.

In the second case when $f \leq \frac{\lambda}{2}$, $P_e(f,\gamma)$ is a concave function of γ and then its maximum is interior. We then need

$$P_e(c, 1-\frac{\tau}{q}) > P_e(f, \gamma_f^*), \tag{B.3}$$

, where $\gamma_f^* = \frac{f}{\lambda-2f}\frac{z}{q}$ in order for γ_τ^* to be an equilibrium. We denote as $\bar{\lambda}$ the maximum possible λ that satisfies this condition. That is, as long as holding costs are at or below $\bar{\lambda}$, the AP avoids a fire sale. $\bar{\lambda}$ is given by the positive root of the following quadratic equation:

$$\lambda^2(q-\tau)^2 - \lambda\left(2f(q-\tau)^2 + Y\right) + 2fY + (fz)^2 = 0$$

where

$$Y = 2q(f-c)(q+z) + 2(q-\tau)cz + 2c(q-\tau)^2 .$$

Then

$$\bar{\lambda} = \frac{2f(q-\tau)^2 + Y + \sqrt{[2f(q-\tau)^2+Y]^2 - 4(q-\tau)^2[2fY+(fz)^2]}}{2(q-\tau)^2}.$$

If the condition in equation B.3 is satisfied, the AP's optimal choice is $\gamma_\tau^* = 1-\frac{\tau}{q}$. There might be cases when holdings costs λ are so large relative to fire sales costs f (i.e., $\lambda > \bar{\lambda}$) that the condition is not satisfied, but these cases are perhaps less interesting from a practical point of view since fire sales costs are usually assumed to be large (e.g., Shleifer and Vishny

(1992)). If holding costs are above $\bar{\lambda}$, then the equilibrium γ^* is given by γ_f^*.

If $P_e(c, 1-\frac{\tau}{q}) > P_e(f, \gamma)$ for any γ, the ETF price is the left-most point on the blue curve before the discontinuity in Figure 5.2. Since the AP does not pick γ_c^*, the ETF price is lower than the one in the optimal case: $P_e(c, 1-\frac{\tau}{q}) < P_e(c, \gamma_c^*)$. This is so because the AP has to hold more than the optimal quantity of redeemed bonds, and therefore incurs additional holding costs. Thus, she can quote a lower ETF price.

Let us now find the discount (negative premium) as a function of γ. After redeeming q ETF shares for q units of bond A, the NAV of the ETF becomes:

$$NAV = \frac{1}{1-q}((0.5-q)(P_0-(1-\gamma)cq)+0.5P_0) = P_0 - \frac{0.5-q}{1-q}(1-\gamma)cq \tag{B.4}$$

The discount $-\pi$ is then:

$$-\pi = -(P_e - NAV) = \gamma^2(\frac{\lambda}{2}-c)q + (1-\gamma)zc + \frac{0.5+\gamma(0.5-q)}{1-q}cq$$

B.1.2 Proof of Proposition 2

Let us analyze the case when baskets are identical to holdings (full baskets). To illustrate the effects on discounts parsimoniously, assume the ETF puts equal amounts of bonds A and B in the basket $q_A = q_B = \frac{q}{2}$, which corresponds to the proportion of the two bonds in holdings. The AP chooses $\gamma_A = \gamma_B = \gamma$, and the redemption is typical. Then, using $P_A = P_B = P_0$, we can write AP's payoff as:

$$(0.5(P_0 - c(1-\gamma)\frac{q}{2}) + 0.5(P_0 - c(1-\gamma)\frac{q}{2}) - P_e)q - 2\frac{\lambda}{2}\cdot(\frac{\gamma q}{2})^2 - (z + 2\frac{\gamma q}{2})c(1-\gamma)\frac{q}{2}$$

$$(P_0 - \frac{c}{2}(1-\gamma)q - P_e)q - \frac{\lambda}{4}(\gamma q)^2 - (z+\gamma q)\frac{c}{2}(1-\gamma)q$$

This expression is exactly the same as in the case with fractional baskets except that the multipliers of λ and c are now divided by two. The ETF price is then:

$$P_e = P_0 - \frac{c(q+z) + \gamma^2(\frac{\lambda}{2}-c)q - \gamma cz}{2} \tag{B.5}$$

Equation B.5 shows that the ETF price is larger than in the fractional baskets case for any γ. The intuition is that with full baskets, both the the holding costs are smaller (due to convexity, spreading the redeemed amount on two parts decreases total costs), the price impact costs are smaller (since the arbitrage profit is concave in q), and the AP can quote a larger price.

NAV and discount are:

$$NAV = \frac{1}{1-q}((0.5-\frac{q}{2})(P_0-\frac{c(1-\gamma)q}{2})+(0.5-\frac{q}{2})(P_0-\frac{c(1-\gamma)q}{2})) = P_0-\frac{c(1-\gamma)q}{2} \quad \text{(B.6)}$$

$$-\pi = -(P_e - NAV) = \frac{1}{2}[\gamma^2(\frac{\lambda}{2}-c)q + cz(1-\gamma) + cq\gamma] \quad \text{(B.7)}$$

Since $0 < q < 0.5$, equation B.6 shows that NAV is lower compared to the case of fractional baskets, because with full baskets prices of both bonds adjust. Since the ETF price is higher, the discount is smaller as seen from equation B.7.

B.2 Two-period setting with fire sales

Consider two periods with redemptions. In the first period, there is a fire sale: the redemption amount is above the threshold, and $(1-\gamma^*)q_1 > \tau$. If $f > \frac{\lambda}{2}$, or $f \leq \frac{\lambda}{2}$ and Equation B.3 is satisfied, the AP chooses to sell τ and then she has to hold part of the redeemed bonds from period 1 to period 2 to avoid fire sales and to avoid paying cost f when trading bonds.

If the redemption amount q_2 (in the same bond for simplicity) in the second period is small, the AP can sell the inefficiently held amount of bonds completely while paying price impact cost c. This would allow her to raise the ETF price in the second period, potentially above the one in the first period. Denote the fraction of q_2 held after redemption in the second period as $\gamma_2 \leq 1$. If $P_{e,2}(\gamma_2^*) > P_{e,1}(1-\tau/q)$, the ETF price drops initially and then recovers as the AP gradually sells the redeemed bonds.

The AP's payoff in the second period is:

$$(P_0 - c\tau - c(1-\gamma_2)q_2 - P_{e,2})q_2 - \underbrace{\frac{\lambda}{2}(q_1 - \tau + \gamma_2 q_2)^2}_{\text{amount of bonds held after deciding how much to sell}} - (z + q_1 - \tau + \gamma_2 q_2)c(1-\gamma_2)q_2$$

Assuming zero profit as before gives the ETF price:

$$P_{e,2} = P_0 - c(1-\gamma_2)q_2 - c\tau - \frac{\lambda}{2}\gamma_2^2 q_2 - \frac{\lambda}{2q_2}(q_1-\tau)^2 - \lambda(q_1-\tau)\gamma_2 - c(z+q_1-\tau+\gamma_2 q_2) + c\gamma_2(z+q_1-\tau+\gamma_2 q_2)$$

$$P_{e,2} = P_0 - \gamma_2^2(\frac{\lambda}{2} - c)q_2 + \gamma_2(cz + (c-\lambda)(q_1-\tau)) - \frac{\lambda}{2q_{2i}}(q_1-\tau)^2 - c(z+q_1+q_2) \quad \text{(B.8)}$$

Note that the functional form of $P_{e,2}$ is now different compared to $P_{e,1}$. Taking FOC gives the optimal γ:

$$\gamma_2^* = \frac{cz + (c-\lambda)(q_1-\tau)}{(\lambda-2c)q_2} = \gamma_c^* + \frac{c-\lambda}{(\lambda-2c)}\frac{q_1-\tau}{q_2} \quad \text{(B.9)}$$

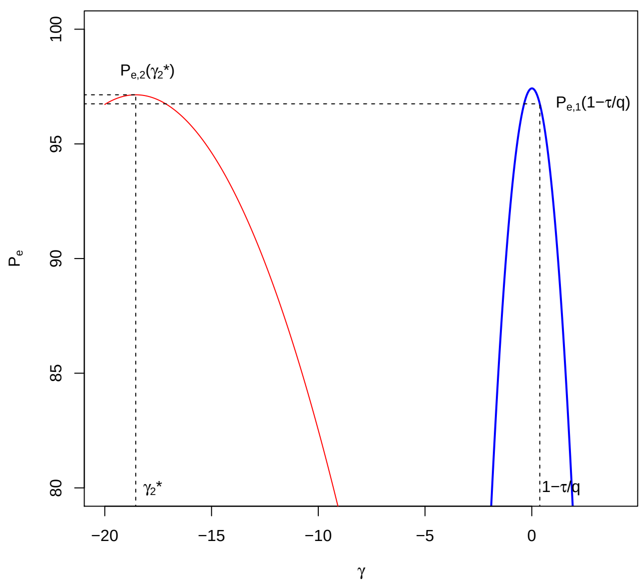
The optimal gamma is similar to the expression from B.1, but now there is an additional term. This term captures the tradeoff that the AP is facing: on the one hand, selling the inefficiently held bonds from the first period $(q_1 - \tau)$ decreases bond prices and as a result the arbitrage profit and the mark-to-market value of AP's inventory. On the other hand, keeping the bonds on the balance sheet incurs holding costs. As $\lambda > 2c$, the second term is negative, and $\gamma_2^* < \gamma_c^*$. For small q_2, γ_2^* can even be below zero because the AP can sell more than 100% of q_2 by selling part of the inefficiently held bonds from the first period. In other words, $1 - \gamma_2^* > 1$.

If $P_{e,2}(\gamma_2^*) > P_{e,1}(1-\tau/q)$, the ETF price increases in the second period (see Figure B.1). The intuition is as follows: the AP had to hold more than the optimal quantity of bonds in the first period, which decreased $P_{e,1}$ due to holding costs. However, in the second period, the AP is able to offload the unnecessary quantity and to save on the balance sheet costs. This allows her to increase $P_{e,2}$. The reversal is more likely to happen if the AP holds a low initial inventory in the redeemed bond. For a larger initial inventory, the price in the second period can be lower, in which case the ETF price decreases further instead of reverting.

B.2.1 Two-period setting with uncertainty

The two-period setting can be easily extended by allowing q_2 to be random. In that case, the price can revert if the expected q_2 is different from the realized one. Suppose that with

Figure B.1: Reversal: intuition. The figure shows the ETF price as a function of γ in the first and second periods. Parameters: $\tau = 0.25, q_1 = 0.4, q_2 = 0.01, z = 0.008, c = 10, f = 20, \lambda = 60, P_0 = 100$.



probability h there is typical redemption, whereas with probability l there are fire sales and the AP has to hold the inefficiently held bonds from the first period, for yet another period. By changing these probabilities, the model can generate a steeper drop in the ETF price in the first period in case of fire sales. This is because now the AP expects to hold $q - \tau$ bonds for $(1+l)$ periods compared to the 1 period from the baseline case. The only difference to the previous analysis is that now the balance sheet cost term is $\frac{\lambda(1+l)}{2} \cdot (\gamma q)^2$ and the ETF price in the first period is lower since $P_{e,1}$ is decreasing in λ: $P_{e,1}(c, \lambda(1+l), 1-\frac{\tau}{q}) < P_{e,1}(c, \lambda, 1-\frac{\tau}{q})$.

B.3 Creation (q>0)

Now consider the case when there is a creation of q units: the AP buys bonds and exchanges them for ETF shares. γ is now the share of bonds that the AP takes from her own balance sheet, whereas $1-\gamma$ is the share bought on the bond market. In contrast to redemptions, the price impact cost is now continuous, without a jump at $x = \tau$: there are no "fire purchases". We first consider the case when there are no balance sheet costs in case of creation. Buying bonds on the market is now beneficial since it increases AP's balance sheet mark-to-market value. The representative AP acts now as if driven by competition on the *min* ETF (since this is what the AP gets paid). AP's payoff is:

$$\underbrace{(P_e - NAV_{basket}) \cdot q}_{\text{arbitrage profit}} + \underbrace{(z - \gamma q) \cdot \Delta P_A}_{\text{mark-to-market gain}}$$

$$(P_e - P_0 - 1_{\gamma<1} \cdot c \cdot (1-\gamma)q) \cdot q + 1_{\gamma<1} \cdot (z - \gamma q) \cdot c \cdot (1-\gamma)q \tag{B.10}$$

If the AP takes everything from her inventory ($\gamma = 1$), her payoff is $(P_e - P_0)q$. The implicit assumption is then that the AP has enough inventory: $z > q$.

If the AP buys everything on the market ($\gamma = 0$), her payoff is $(P_e - P_0)q + cq(z - q)$. If $z > q$, she benefits from the mark-to-market gain and is willing to buy everything on the market (this is easy to see by comparing with the price from the case when $\gamma = 1$).

If the AP takes some of the bonds from her inventory ($0 < \gamma < 1$) and buys the rest, her payoff is:

$$(P_e - P_0 - c \cdot (1-\gamma)q) \cdot q + (z - \gamma q) \cdot c \cdot (1-\gamma)q$$

$$cq(z-q) + \gamma^2 cq^2 - \gamma cqz + q(P_e - P_0) \tag{B.11}$$

The price is:

$$P_e = \begin{cases} P_0 & ,\gamma = 1 \\ P_0 - c(z-q) & ,\gamma = 0 \\ P_0 - c(z-q) - \gamma^2 cq + \gamma cz & ,0 < \gamma < 1 \end{cases} \tag{B.12}$$

The function is concave and since the AP is minimizing, the solution is a corner one: either $\gamma = 0$ or $\gamma = 1$. If $z \geq q$, the solution is $\gamma = 0$ since $P_e(0) < P_e(1)$. The intuition is that for a large initial inventory, the AP makes a larger mark-to-market gain by buying bonds on the market, which allows her to accept a lower ETF price and a lower profit from creating ETF shares. In fact, this is the only possible case because if $z < q$, the AP cannot set $\gamma = 1$ since she does not have enough inventory to create q shares.

NAV is then (using bond A as the basket, and remembering that $P_A = P_B = P_0$ in the initial period):

$$NAV(\gamma = 0) = \frac{1}{1+q}(0.5P_0 + (0.5+q)(P_0 + cq)) = P_0 + \frac{0.5+q}{1+q}cq \tag{B.13}$$

The premium is:

$$\pi(\gamma = 0) = P_e(\gamma = 0) - NAV(\gamma = 0) = \frac{cq}{2(1+q)} - cz \tag{B.14}$$

The premium can be even negative if the AP has a large inventory: the intuition is that the AP is ready to accept even negative ETF profit if buying bonds on the market boosts her mark-to-market value.

B.3.1 Premium when baskets are identical to holdings

To illustrate the effects on premiums parsimoniously, assume again as in the case of redemptions that $q_A = q_B = q/2$ and the AP chooses $\gamma_A = \gamma_B = \gamma$. Then, using $P_A = P_B = P_0$, the AP's payoff is:

$$(P_e - 0.5(P_0 + c \cdot (1-\gamma)q/2) + 0.5(P_0 + c \cdot (1-\gamma)q/2))q + (z - 2\frac{\gamma q}{2}) \cdot c \cdot (1-\gamma)\frac{q}{2}$$

$$(P_e - P_0 - c/2 \cdot (1-\gamma)q) \cdot q - (z - \gamma q) \cdot c/2 \cdot (1-\gamma)q$$

This expression is exactly the same as in the case with fractional baskets except that the multipliers of c is now twice smaller. The optimal $\gamma = 0$ and the ETF price is then $P_e(\gamma = 0) = P_0 - \frac{c(z-q)}{2}$. The ETF price is larger than in the fractional baskets case.

NAV and premium are:

$$NAV(\gamma = 0) = \frac{1}{1+q}((0.5 + \frac{q}{2})(P_0 + \frac{cq}{2}) + (0.5 + \frac{q}{2})(P_0 + \frac{cq}{2})) = P_0 + \frac{cq}{2} \tag{B.15}$$

$$\pi_1(\gamma = 0) = P_e(\gamma = 0) - NAV(\gamma = 0) = -\frac{cz}{2} \tag{B.16}$$

Since $0 < q < 0.5$, equation B.15 shows that NAV is larger compared to the case of fractional baskets, since now prices of both bonds adjust. The premium is smaller as seen from equation B.16 and is always negative since the AP benefits from pushing bond prices higher, which increases the mark-to-market value and allows her to accept even negative ETF profit.

B.3.2 Creation with balance sheet costs

Let us now consider the case when there are balance sheet costs to creation. Assume that taking bonds from the balance sheet also incurs a cost since the AP deviates from the target inventory: $\lambda > 0$. AP's payoff is now:

$$\underbrace{(P_e - NAV_{basket}) \cdot q}_{\text{arbitrage profit}} - \underbrace{\frac{\lambda}{2} \cdot (\gamma q)^2}_{\text{balance sheet cost}} + \underbrace{(z - \gamma q) \cdot \Delta P_0}_{\text{mark-to-market gain}}$$

$$(P_e - P_0 - 1_{\gamma<1} \cdot c \cdot (1-\gamma)q) \cdot q - \frac{\lambda}{2} \cdot (\gamma q)^2 + 1_{\gamma<1} \cdot (z - \gamma q) \cdot c \cdot (1-\gamma)q \tag{B.17}$$

If the AP takes everything from her inventory ($\gamma = 1$), her payoff is $(P_e - P_0)q - \frac{\lambda}{2} \cdot q$. The implicit assumption is then that the AP has enough inventory: $z > q$.

If the AP buys everything on the market ($\gamma = 0$), her payoff is $(P_e - P_0)q + cq(z - q)$. If $z > q$, she benefits from the mark-to-market gain more than she loses on the arbitrage profit and she is willing to buy everything on the market.

If the AP takes some of the bonds from her inventory ($0 < \gamma < 1$) and buys the rest, her payoff is given by:

$$(P_e - P_0 - c \cdot (1-\gamma)q) \cdot q - \frac{\lambda}{2} \cdot (\gamma q)^2 + (z - \gamma q) \cdot c \cdot (1-\gamma)q$$

$$\gamma^2 (c - \frac{\lambda}{2}) q^2 - \gamma c q z + c q (z - q) + q(P_e - P_0) \tag{B.18}$$

$$P_e = \begin{cases} P_0 - \frac{\lambda}{2} \cdot q & , \gamma = 1 \\ P_0 - c(z - q) & , \gamma = 0 \\ P_0 - c(z - q) - \gamma^2 (c - \frac{\lambda}{2}) q + \gamma c z & , 0 < \gamma < 1 \end{cases} \tag{B.19}$$

The solution to the minimization problem is interior ($0 < \gamma^* < 1$) if $\lambda < 2c$, since then the function is convex. In other words, the creation balance sheet cost multiplier is smaller than the price impact cost, which is the **opposite** condition to the one for the redemption case.

The optimal γ is then found by taking FOC:

$$2\gamma(c - \frac{\lambda}{2})q = cz$$

$$\gamma^* = \frac{c}{2c - \lambda}\frac{z}{q} \tag{B.20}$$

The optimal γ depends then on whether the parameters of the model give $0 < \gamma^* < 1$. A necessary condition for this to be satisfied is $\lambda < 2c$: for creation, the balance sheet cost multiplier has to be smaller than $2c$. In other words, the holding costs are asymmetric for creation and redemption, which makes sense intuitively. The cost of putting additional bonds on the inventory is larger than the cost of taking bonds from the inventory, since the former cost consumes balance sheet space and is costlier from a regulations' prospective. AP's inventory needs to satisfy $z < q\frac{2c-\lambda}{c}$ for $\gamma^* < 1$.

Previous volumes in this series

974 November 2021	The natural rate of interest through a hall of mirrors	Phurichai Rungcharoenkitkul and Fabian Winkler
973 October 2021	What does digital money mean for emerging market and developing economies?	Erik Feyen, Jon Frost, Harish Natarajan and Tara Rice
972 October 2021	Non-bank financial intermediaries and financial stability	Sirio Aramonte, Andreas Schrimpf and Hyun Song Shin
971 October 2021	Do term premiums matter? Transmission via exchange rate dynamics	Mitsuru Katagiri and Koji Takahashi
970 October 2021	Big techs in finance: on the new nexus between data privacy and competition	Frederic Boissay, Torsten Ehlers, Leonardo Gambacorta and Hyun Song Shin
969 October 2021	Inter-agency coordination bodies and the speed of prudential policy responses to the Covid-19 pandemic	Michael Brei and Blaise Gadanecz
968 October 2021	Indebted demand	Atif Mian, Ludwig Straub and Amir Sufi
967 October 2021	Joined at the hip: monetary and fiscal policy in a liquidity-driven world	Guillermo A Calvo and Andrés Velasco
966 October 2021	The Treasury market in spring 2020 and the response of the Federal Reserve	Annette Vissing-Jørgensen
965 October 2021	Technological capacity and firms' recovery from Covid-19	Sebastian Doerr, Magdalena Erdem, Guido Franco, Leonardo Gambacorta and Anamaria Illes
964 September 2021	Macroeconomic policy under a managed float: a simple integrated framework	Pierre-Richard Agénor and Luiz A Pereira da Silva
963 September 2021	The Fed takes on corporate credit risk: an analysis of the efficacy of the SMCCF	Simon Gilchrist, Bin Wei, Vivian Z Yue and Egon Zakrajšek
962 August 2021	Global lending conditions and international coordination of financial regulation policies	Enisse Kharroubi
961 August 2021	Private equity buyouts and firm exports: evidence from UK firms	Paul Lavery, Jose-Maria Serena, Marina-Eliza Spaliara and Serafeim Tsoukas

All volumes are available on our website www.bis.org.