Steering a Ship in Illiquid Waters: Active Management of Passive Funds

Naz Koont
Stanford Graduate School of Business, USA

Yiming Ma
Columbia Business School, NBER, USA

Lubos Pastor
University of Chicago Booth School of Business, NBER, CEPR, USA

Yao Zeng
The Wharton School, University of Pennsylvania, NBER, USA

Abstract

Exchange-traded funds (ETFs) are typically viewed as passive index trackers. In contrast, we show that corporate bond ETFs actively manage their portfolios, trading off index tracking against liquidity transformation. In our model, ETFs optimally choose creation and redemption baskets that include cash and only a subset of index assets, especially if those assets are illiquid. Our evidence supports the model. We find that ETFs dynamically adjust their baskets to correct portfolio imbalances while facilitating ETF arbitrage. Basket inclusion improves bond liquidity in general, but worsens it in periods of large imbalance between creations and redemptions, such as the COVID-19 crisis. (*JEL* G12, G23)

We are grateful for comments from Livia Amato, Itzhak Ben-David, Alex Chinco, Zhi Da, Caitlin Dannhauser, Davidson Heath, Wei Jiang, Xu Lu, Thomas Marta, Stefan Nagel, Anna Pavlova, Adam Reed, Christopher Reilly, Marco Sammon, John Shim, Clemens Sialm, Robert Stambaugh, Stijn Van Nieuwerburgh, Laura Veldkamp, Russ Wermers, Liyan Yang, conference participants at the 2024 Spring NBER LTAM meeting, 2024 ESSFM, 2023 AFA, 2023 FIRS, 2023 Workshop on Non-Bank Financial Institutions, 2023 Spring Q Group, 2022 Conference on Financial Economics and Accounting, 2022 Fixed Income and Financial Institution Conference, 2022 Four Corners Index Investing Jamboree, 2022 Minnesota Corporate Finance Conference, 2022 Toronto Junior Finance/Macro Conference, 2022 MIT Sloan Junior Finance Conference, and seminar participants at Chicago, Columbia, Florida, HKUST, Indiana, LSU, New York Fed, Notre Dame, SAIF, TMX Market Forum, UNC, USC, UT Dallas, and Wharton. We thank market participants at Barclays, BlackRock, State Street, and Vanguard—Shreya Adiraju, Joshua Barrickman, Matthew Bartolini, Stephen Laipply, Ananth Madhavan, Jay Mauro, Jeff Meli—for many helpful conversations. We thank Livia Amato for excellent research assistance. Pastor serves as an independent director and trustee of Vanguard. The views expressed here do not necessarily reflect those of The Vanguard Group or the Vanguard Funds. This research was funded in part by the Fama-Miller Center for Research in Finance at the University of Chicago Booth School of Business. Send correspondence to Lubos Pastor, lubos.pastor@chicagobooth.edu.

Exchange-traded funds (ETFs) are among the most important financial intermediaries. Their assets under management have grown quickly since their first appearance in 1993, reaching $7.2 trillion by the end of 2021 in the United States alone. This amount is about half of total assets of U.S. equity mutual funds, and it exceeds total assets of U.S. fixed income mutual funds. Given their large and growing footprint, ETFs seem relatively understudied.

The vast majority of ETFs track passive indexes. To manage index deviations, ETFs rely on authorized participants (APs) to conduct arbitrage trades, in which APs create and redeem ETF shares in exchange for baskets of securities called the "creation basket" and the "redemption basket," respectively. These baskets are chosen by the ETF.

We study how ETFs use creation and redemption baskets to manage their portfolios. We make three main contributions. First, we show that, despite their passive image, ETFs are remarkably active in their portfolio management. They often use baskets that deviate substantially from the underlying index, by including cash and only a subset of index securities, and they adjust those baskets dynamically in a predictable fashion. Second, we build a theoretical model in which ETFs' active basket management emerges endogenously. The model makes predictions about basket composition and its interactions with the liquidity of the underlying securities. We find support for these predictions in the data. Third, we show that ETFs' active management has state-contingent effects on the liquidity of the underlying securities. Being included in an ETF basket typically makes a security more liquid, but this relation flips in periods of large imbalance between creations and redemptions.

Our first two contributions indicate that ETFs are active to facilitate liquidity transformation. ETF shares tend to be more liquid than the underlying securities, in part because APs' arbitrage trades tend to absorb shocks to investors' demand for ETF shares. When investors sell ETF shares, APs can buy and redeem them; when investors buy ETF shares, APs can create and sell them. By absorbing ETF investors' trades, APs reduce the price impact of those trades. APs' arbitrage trading thus provides liquidity to investors who must trade ETFs at short notice. This liquidity provision requires APs to trade basket securities and thus incur transaction costs. To help reduce those costs, ETFs adjust their baskets to make them cheaper to trade. These basket adjustments help ETFs transform liquidity, but they also constrain ETFs' index-tracking capacity. We argue that active basket management is necessary for ETFs to balance index tracking and liquidity transformation.

Our third contribution, specific to bond ETFs, shows that their active management has significant consequences for bond market liquidity. In normal times, a bond's inclusion in an ETF basket makes the bond more liquid because shocks to investors' demand for ETF shares are largely idiosyncratic. A random mix of creations and redemptions across ETFs increases the trading activity in basket bonds, improving their liquidity. That is not the case, however, in periods when investors' liquidity shocks are systematic, resulting in imbalances between creations and redemptions. For example, large redemptions move redemption-basket bonds to APs' balance sheets. The APs, who also tend to act as market makers in these bonds, may become reluctant to purchase more of the same bonds, reducing their liquidity. The effect of basket inclusion on bond liquidity is thus state-dependent: positive in normal times but negative when there is large imbalance between creations and redemptions.

Our empirical analysis focuses on U.S. corporate bond ETFs. We choose this category to emphasize the role of liquidity transformation, which is enhanced by the relative illiquidity of corporate bonds. Corporate bond ETFs have been growing fast, with assets tripling over the past six years before reaching over $342 billion at the end of 2021. We analyze the baskets, portfolios, and indexes of bond ETFs, along with the characteristics of the underlying bonds. Our analysis uses both proprietary and nonproprietary data from 2017 to 2020.

We begin by establishing two simple facts about bond ETF baskets. First, these baskets include large amounts of cash. The average creation (redemption) basket contains 4.6% (7.8%) of its assets in cash, based on the baskets pre-announced by the ETF at the start of a trading day. The cash proportions are even larger, 11.6% (8.2%) for creation (redemption) baskets, based on baskets imputed from ETF holdings. Second, ETF baskets are concentrated—they include only a small subset of the bonds that appear in the underlying index (see also Shim and Todorov 2023). Both facts impair the ETF's index tracking.

To rationalize these facts, we build a model that highlights ETFs' dual role of index tracking and liquidity transformation. The model features two types of investors, patient and impatient, and an AP. The impatient investors face a liquidity shock that forces them to either consume or save at short notice. Investors who need to consume sell ETF shares; those who need to save buy them. Any imbalance between the buys and sells is exploited by the AP via an arbitrage trade. When there are more sells than buys, the AP buys ETF shares in the market and redeems them in exchange for the redemption basket. When there are more buys than sells, the AP sells ETF shares after creating them in exchange for the creation basket. Either way, the AP improves the ETF's liquidity by mitigating the price impact of impatient investors. This liquidity improvement comes at a cost to the AP, who incurs transaction costs by trading basket securities. To reduce those costs, the ETF finds it optimal to use baskets that include some cash and only a subset of index securities, as we observe in the data. However, such baskets deviate from the index, which is generally costly to patient investors. Therefore, when designing its basket, the ETF balances the benefits of liquidity transformation against the costs of imperfect index tracking.

Our model also predicts that ETFs have more basket cash and larger tracking errors when their securities are less liquid. Intuitively, when liquidity transformation is more costly, ETFs find it optimal to reduce this cost by tolerating looser index tracking. We test these predictions empirically. We find that ETFs investing in less liquid corporate bonds indeed have more cash in their baskets and larger tracking errors, consistent with the model.

We make two sets of predictions about how an ETF should manage its portfolio over time. First, the ETF should dynamically adjust its creation and redemption baskets to steer its portfolio toward the index. Second, these adjustments should be smaller when liquidity transformation is more costly. For example, suppose the ETF's portfolio currently overweights a given security relative to the index. The ETF should add some of this security to the redemption basket and remove it from the creation basket. Similarly, if a security is underweighted, it should be added to the creation basket and removed from the redemption basket. Either way, these basket modifications should be attenuated when the ETF invests in less liquid securities. The same ideas apply to the cash balance. If the ETF currently holds more cash than usual, it should add cash to the redemption basket and remove it from the creation basket, but less so if the underlying index is less liquid.

We find empirical support for these predictions in our sample of corporate bond ETFs. One estimation challenge is that the over/underweighting of securities in ETF portfolios can be driven by unobserved fundamentals. To address this challenge, we make use of monthly index rebalancing. Bond indexes rebalance at month-ends to reflect changes in bond characteristics during the month. Bond index weights jump at the rebalancing dates, providing plausibly exogenous variation in over/underweighting that allows us to trace out the ETFs' response. We find that when index rebalancing makes a bond more overweighted (underweighted) in an ETF's portfolio, the ETF includes more of that bond in its redemption (creation) basket. ETFs thus revise their baskets to steer their portfolios toward the index, as predicted. We also find these basket adjustments are less pronounced for ETFs tracking less liquid indexes, again consistent with our predictions.

Having analyzed the nature and causes of ETFs' active portfolio management, we study its consequences for bond market liquidity. We find that ETFs' efforts to improve the liquidity of their shares significantly affect the liquidity of the underlying bonds. Specifically, a bond's inclusion in an ETF basket makes the bond more liquid in normal times but less liquid in times of large imbalance between creations and redemptions.

We reach these conclusions after estimating the effect of basket inclusion on liquidity in two ways. First, we relate bond liquidity to two measures of basket inclusion in the presence of numerous controls and fixed effects. Second, we exploit the monthly bond index rebalancing once again to construct a novel instrument for basket inclusion. For each bond, the instrument measures the jump in portfolio overweighting at the rebalancing date added up across all ETFs experiencing creations or redemptions. The instrumented inclusion of that bond in an ETF basket should then reflect a plausibly exogenous shock. Repeating our estimation using two-stage least squares (2SLS), we find that a bond's inclusion in an ETF basket, creation or redemption, improves the bond's liquidity in the full sample.

This result flips in periods of large imbalance between creations and redemptions. For example, bond ETFs faced large net redemptions in March 2020, after the onset of the COVID-19 pandemic (Falato, Goldstein, and Hortacsu 2021). When we rerun our baseline regression for the spring of 2020, we find that being included in a redemption basket worsens rather than improves a bond's liquidity. Similarly, when we interact basket inclusion with basket imbalance, the inclusion's effect on liquidity is negative when the imbalance is sufficiently large. The effect of active basket management on bond liquidity is thus state-dependent.

1 Related Literature

The ETF literature has traditionally examined equity ETFs (see Ben-David, Franzoni, and Moussawi 2017 and Lettau and Madhavan 2018 for surveys). Given our focus on liquidity transformation, we instead analyze bond ETFs, for which this transformation is far more extensive. We show that bond ETFs' role in liquidity transformation turns passive ETFs into active managers of their portfolios. ETFs balance this role against their better-known index-tracking role, with strong implications for the underlying bonds.

We merge data on ETF creation and redemption baskets, ETF holdings, index holdings, and the underlying bonds for a large, representative sample of bond ETFs, assembling the most comprehensive ETF database to date, to our knowledge. ETF basket data are important for understanding ETF arbitrage, are rarely analyzed, and their analysis represents the frontier of the ETF literature. Our work complements several ongoing studies that do examine bond ETF baskets. Shim and Todorov (2023) show that these baskets tend to cover only a small fraction of ETF holdings. They explore the implications of concentrated baskets for ETF premiums and discounts, a focus different from ours. Reilly (2022) finds that APs engaging in ETF arbitrage tend to deliver underperforming bonds in creation baskets. His focus on the information asymmetry between ETFs and APs is also different from ours. Several other studies also focus on the role of APs. Pan and Zeng (2019) find that APs strategically use creations and redemptions to manage their own bond inventories. Dannhauser and Karmaziene (2023) show that bonds selected into creation baskets have higher dealer inventory costs. Du (2023) shows that bond ETFs primarily create with a small subset of dealers that hold larger inventories despite concerns of adverse selection and agency problems. Gorbatikov and Sikorskaya (2022) analyze the concentration and market power of APs. Helmke (2023) shows theoretically that AP balance sheet costs can in principle make ETFs less liquid than mutual funds. Shim and Todorov (2023) build a model in which the AP acts as a buffer between the ETF market and the bond market. We depart from these studies by highlighting the role of cash in ETFs' basket management, by showing that this management of both cash and basket bond composition is surprisingly active, and by studying the effects of this active management on bond liquidity.

We are not the first to note that ETFs are somewhat active. Some ETFs track nontraditional indexes, while others exhibit large index deviations (Akey, Robertson, and Simutin 2021; Easly et al. forthcoming; Ben-David et al. 2023; Brogaard, Heath, and Huang 2024). Some ETFs engage in securities lending or cross-subsidization of affiliated financial institutions (Cheng, Massa, and Zhang 2019). Some equity index ETFs and smart beta products engage in factor investing (Cong, Huang, and Xu 2024). We contribute to this literature in two ways. First, while the studies mentioned earlier analyze equity ETFs, we focus on bond ETFs and the related liquidity transformation. Second, and more important, we emphasize a different notion of ETFs' activeness—their active basket management, including the dynamic management of their cash balances and individual bonds. By examining ETFs' active portfolio management and its underlying trade-offs, we also relate to the literature on fund activeness (e.g., Kacperczyk, Sialm, and Zheng 2008; Cremers and Petajisto 2009) and the trade-offs among active funds' characteristics (Pastor, Stambaugh, and Taylor 2020).

Our work also relates to the literature on the asset pricing implications of ETF ownership. For equity ETFs, this literature shows that ETF ownership has positive effects on stock volatility (Ben-David, Franzoni, and Moussawi 2018), return co-movement (Da and Shive 2018), and liquidity co-movement (Agarwal et al. 2018), but an unclear effect on stock liquidity (Israeli, Lee, and Sridharan 2017; Saglam, Tuzun, and Wermers 2021). Equity ETF returns have little impact on the underlying stock returns (Box et al. 2021). For bond ETFs, Dannhauser and Hoseinzade (2022) find that the selling pressure on ETFs induced by the 2013 taper tantrum was transmitted to the bonds owned by those ETFs. Dannhauser and Dathan (2023) link ETF participation to the pricing and liquidity in the primary market. The evidence on the effect of ETF ownership on the liquidity of the constituent bonds is mixed. While Dannhauser (2017) finds that ETF ownership has a weak or even negative effect on bond liquidity, Holden and Nam (forthcoming) and Marta (2024) find positive effects. Instead of ETF ownership, we analyze ETF baskets. They offer a more direct perspective on ETF activity because they contain securities that are actually transacted rather than just owned by ETFs. We show that ETFs' active basket management has significant state-dependent effects on the liquidity of the underlying bonds. To show causality, we design a novel instrument for basket inclusion. Using our instrument, Finnerty, Reisel, and Zhong (forthcoming) report that basket inclusion improves bond liquidity. While they use a different ETF sample, including two BlackRock ETFs, different liquidity measures, and fewer controls and fixed effects, their findings are similar to ours, except that we also find the opposite effect on liquidity in periods of large creation-redemption imbalance.

Finally, we contribute to the broad literature on liquidity transformation by financial intermediaries. Starting with Diamond and Dybvig (1983), this literature has traditionally focused on banks. More recently, this literature has come to emphasize the liquidity provision by shadow banks (e.g., Kacperczyk and Schnabl 2013; Sunderam 2015), including mutual funds (e.g., Chen, Goldstein, and Jiang 2010; Goldstein, Jiang and Ng 2017; Chernenko and Sunderam 2017; Choi et al. 2020; Anand, Jotikasthira, and Venkataraman 2021; Jin et al. 2022; Ma, Xiao, and Zeng 2023). We analyze the liquidity transformation by ETFs, one of the fastest-growing financial intermediaries. Unlike banks or mutual funds, ETFs do not need cash to meet redemptions. Nonetheless, we show that ETFs use substantial amounts of cash in their baskets to facilitate arbitrage by APs.

2 Data

To construct our sample, we identify passive U.S. corporate bond ETFs in the ETF Global database. We exclude ETFs that use active strategies, total bond market ETFs, and ETFs that primarily invest in Treasury bonds and bills, mortgage-backed securities, international bonds, municipal bonds, and loans. Our sample includes 118 ETFs from January 1, 2017, to December 31, 2020.

For each ETF, we obtain daily data on fund portfolio holdings, shares, and prices from ETF Global. We also obtain index holding data from Bloomberg for about half of the ETFs in our sample. In panels A and B of Table 1, we report the summary statistics for all ETFs in our sample, as well as for the subset of ETFs whose index holding data are available. The two sets of ETFs look similar based on most observable characteristics. For example, the average ETF in our sample holds 753 bonds in its portfolio. The average portfolio size is slightly larger, 835, when computed only across the ETFs with index holding data. For the latter ETFs, the average number of bonds in the index is 1,153.

Before describing our data on ETF baskets, we provide some background on how those baskets are used. ETFs are investment funds that issue shares backed by a portfolio of securities. ETF shares trade in the secondary market. In the primary market, investors known as APs can create and redeem ETF shares in-kind for the underlying securities.[1] The APs can profit from arbitrage trading across the primary and secondary markets. When ETF shares are relatively cheap in the secondary market, APs can buy them, redeem them for a basket of securities called the redemption basket, and sell the securities at a profit. When ETF shares are expensive, APs can deliver securities in the creation basket to the ETF issuer and sell the newly created ETF shares. Figure 1 illustrates the process.

ETFs disclose their desired creation and redemption baskets ahead of each trading day. During the day, an AP can request basket modifications, such as the omission of hard-to-locate securities or the inclusion of securities that are of interest to the AP. The final basket composition is a result of negotiations between the AP and the ETF. The use of custom baskets that are not fully representative of ETF portfolio holdings is permitted under the Securities and Exchange Commission's rule 6c-11 as long as ETF issuers adopt and enforce policies governing the construction of custom baskets and act in the best interests of the ETF and its shareholders. Custom baskets were common also before the introduction of rule 6c-11 in 2019, due to a widespread use of exemptive orders that allowed non-pro-rata baskets. We do not see any significant change in basket representativeness following the adoption of rule 6c-11, which suggests that prior restrictions on basket composition were not binding.

2.1 Realized ETF baskets

We impute ETFs' realized creation (CR) and redemption (RD) baskets from changes in ETF holdings on days with CR or RD activity (Shim and Todorov 2023). We identify CR (RD) days as those on which there is a positive (negative) change in the number of ETF shares. We then use daily changes in the number of bonds held to determine the composition of the average ETF basket on that day. This imputation assumes that ETFs' portfolio changes reflect only APs' CR and RD activities. This assumption could in principle be violated in several ways. First, ETFs could trade in the secondary market instead of relying solely on CR and RD activities to manage their portfolios. Such trades are uncommon, as we explain in the following paragraph. Second, different APs may negotiate different baskets with an

[1] For bond ETFs, APs are mostly large dealer banks such as JPMorgan Chase and Goldman Sachs, which also operate bond trading desks and typically act as market makers in the underlying bonds.

ETF on the same day. Third, there may be RD baskets on days with net creations and CR baskets on days with net redemptions. For these reasons, each imputed/realized basket is best interpreted as the average net basket for the given ETF on the given day.

ETFs rarely trade directly in the secondary market because such trades could create tax liabilities for investors, unlike the in-kind exchanges that occur during CR and RD events. To provide some evidence, we compare the extent of portfolio changes—the number of bonds with changes in their shares held by the ETF—on days with and without net creations or redemptions, at the ETF-day level. On days without CR/RD activity, the median number of bonds with changes in shares is only 1, and even the 90th percentile is only 4 (see panel C of Table 1). In contrast, the portfolio changes on days with CR/RD activity are much larger: the median number of bonds with changes in shares is 89, and the 90th percentile is 478. The amount of secondary-market trading by ETFs thus seems fairly small.

On most CR/RD days, realized baskets coincide with changes in the number of ETF shares. However, the daily portfolio change recorded in the ETF Global database occasionally leads or lags the corresponding change in ETF shares by one day. Such instances are easy to spot because of the large number of zero-share-change days in the sample. To correct these apparent data errors, we shift the dates of such realized baskets by one day to align them with the dates of ETF share changes. We do not perform such shifts, and do not impute any baskets, on days when the ETF portfolio changes are very small, affecting only one or two bonds. Such small portfolio changes on days with no changes in the number of ETF shares can result from occasional secondary-market transactions by the ETF, or from corporate events by bond issuers leading to changes in bond identifiers.

2.2 Reported ETF baskets

We obtain proprietary data on ETFs' reported baskets from the Depository Trust and Clearing Corporation (DTCC). Pursuant to the Securities Exchange Act of 1934, Rule 19(b), ETF issuers announce the CR and RD baskets to all APs at the close of each trading day for use on the following trading day. These baskets represent specific portfolios—names and quantities—of securities that ETF issuers intend to exchange for ETF shares during the CR/RD process. ETF issuers distribute this information via the DTCC's National Securities Clearing Corporation, a clearinghouse service that automates the CR/RD process.

The main advantage of this dataset is that baskets are reported on the eve of each trading day, including days with no CR or RD events. On any given day, there are many ETFs with no CR or RD activity, and thus no realized baskets. As we show in panel D of Table 1, on an average day, only 7.7% (19.2%) of ETFs have redemptions (creations), and different ETFs have realized baskets on different days. The reported basket data are therefore particularly useful for cross-sectional analysis comparing ETF baskets on a given day.

The main disadvantage is that reported baskets may differ from the actual baskets used in the CR/RD process. APs can negotiate with the ETF issuer about basket composition, arriving at a realized basket that may differ from the preannounced reported basket. In addition, in our sample, reported CR baskets are always the same as reported RD baskets. We thus rely primarily on realized baskets in our analysis, while using reported baskets only as a complementary dataset when examining the cross-section.[2]

The average ETF in our sample has 424 bonds in the reported basket (see panel A of Table 1). This number exceeds the average numbers of bonds in realized CR and RD baskets (104 and 147, respectively), indicating that reported baskets tend to be larger than realized baskets. In fact, realized baskets tend to be subsets of their reported counterparts—in our sample, only about 1% of ETF-bond-day observations are in realized baskets but not in reported ones. In contrast, reported baskets tend to be smaller than ETF portfolios as well as index portfolios, which shows that even intended ETF baskets are not designed to fully replicate the underlying index.

2.3 Individual bonds

We obtain bond price and trading data from Enhanced TRACE and bond-level characteristics from Mergent FISD. We clean the TRACE data by following the methodology described in Dick-Nielsen (2014). We use the TRACE data to calculate three daily measures of market illiquidity (IL) for each bond: the effective tick size (*Tick*; see Holden (2009); Goyenko, Holden, and Trzcinka (2009)), the imputed round-trip cost (*IRC*; see Feldhutter (2012)), and the interquartile range (*IQR*; see (Song and Zhou 2007; Pu 2009)). We describe these measures and their construction in detail in the Internet Appendix. Panel E of Table 1 provides bond-level summary statistics for the three measures as well as other bond characteristics. When we tabulate our results later in the paper, we use IL1, IL2, and IL3 to denote *Tick*, *IRC*, and *IQR*, respectively. Unless we note otherwise, we winsorize all fund- and bond-level variables at the 1% level.

[2] Vanguard ETFs do not disclose daily portfolio holdings, precluding the imputation of realized baskets. We proxy for Vanguard ETFs' realized baskets by reported baskets on days with ETF share changes.

3 Stylized Facts

In this section, we present stylized facts related to cash holdings and basket concentration.

3.1 Cash in ETF baskets and portfolios

The first fact is that corporate bond ETF baskets contain substantial amounts of cash. To identify cash and cash equivalents in ETF baskets, we manually classify the securities in realized baskets by their asset class and security description in ETF Global. We include any cash-like money market instruments, such as securities labeled as "Cash," "Currency," "US Dollars," as well as money market funds and short-term Treasury ETFs.[3]

We compute the proportions of cash in both realized and reported baskets. For each realized basket, we divide the imputed amount of cash by total basket value on the same day. For each reported basket, we divide the amount of cash reported to DTCC by total basket value. We compute time-series averages of these "cash ratios" fund by fund and show their empirical distributions in Table 2.

Table 2 shows that cash accounts for a significant proportion of realized ETF baskets: 11.6% of CR baskets and 8.2% of RD baskets, on average. The average proportion of cash in reported baskets is a bit smaller, ranging from 4.6% to 7.8% depending on whether we average across all days, days with RD activity, or days with CR activity.[4]

These averages mask large dispersion in the cash ratios across ETF baskets. The cross-sectional distributions of the basket cash ratios have long right tails. Their medians are dwarfed by the means, and the 90th percentiles of imputed cash ratios exceed 34%. Equal-weighted averages are larger than their value-weighted counterparts, indicating that smaller ETFs, such as high-yield bond ones, tend to use more cash in their baskets. These results show that cash plays a significant role in ETF baskets.[5]

[3] The most popular Treasury ETFs are "BlackRock Cash Funds: Treasury, SL Agency" and "Invesco Premier U.S. Government Money Portfolio." We do not include direct Treasury holdings among cash equivalents because Treasury bonds can be illiquid and risky. This treatment is conservative; if we were to include Treasuries as cash, the proportions of cash in ETF baskets and holdings would be even larger.

[4] The fact that CR baskets contain more cash than RD baskets is consistent with the observation that ETF issuers tend to offer more flexibility to APs in the customization of CR baskets. For example, Dannhauser and Hoseinzade (2022) cite a letter to the Securities and Exchange Commission, in which BlackRock states that they allow APs to engage in an iterative process in determining the components of CR baskets, whereas BlackRock's proposed custom RD baskets are generally non-negotiable as redemption does not require APs to source securities.

[5] About a quarter of realized RD baskets, and a smaller fraction of realized CR baskets, have negative amounts of cash. Negative cash can result from the use of credit lines, failed trades, accounting errors, and

Next, we turn our attention from ETF baskets to ETF holdings. Table 2 shows that the average ETF holds 1.7% of its portfolio in cash. ETFs typically keep their cash in money market sweep vehicles. Most of the cash holdings seem discretionary. Only a small part comes from receipts of coupon payments that have not yet been reinvested. The time-series correlation between cash from coupon payments and cash holdings at the ETF-day level is only 1%, presumably because coupon payments can be fully anticipated, and thus easily managed, by the ETF. Proceeds from matured bonds also matter little because most bonds leave the ETF portfolio before they mature. When we add proceeds from matured bonds to coupon payments and recompute the above correlation, it still rounds to 1%. Even though ETFs' cash holdings are nontrivial, it is clear from Table 2 that ETFs hold substantially more cash in their baskets than in their portfolios. The disproportionate use of cash in ETF baskets is consistent with ETFs' desire to incentivize the CR and RD activity by APs. Conversations with ETF managers confirm that cash is often used as a "plug," filling in for bonds that would be costly for the AP to source or sell in the market.[6]

ETFs do not need to hold cash to meet withdrawals, unlike banks and mutual funds, as ETFs do not engage in cash transactions with investors. Furthermore, the indexes tracked by bond ETFs include little or no cash. Any index cash results from intra-month coupon and principal payments that are not reinvested between monthly rebalance dates. Index cash builds up during the month as the payments are received, before dropping to zero at the next month-end rebalancing date. The amounts are small. For example, for the ICE BoFA US Corporate Bond Index, the average daily cash value in 2021 is 11 basis points (bps), and the daily cash levels range from zero to 29 bps. The numbers are only slightly larger for indexes of high-yield bonds, for which coupon payments account for a larger fraction of the return. For example, cash in the ICE BoFA US High Yield Master II Index has a mean of 14 bps and a maximum of 38 bps in 2021.[7] These index cash amounts are at least an order of magnitude smaller than the cash amounts in ETF baskets reported in Table 2. As a result, any nontrivial cash in ETF portfolios is costly in terms of index tracking.

3.2 Basket concentration

Our second stylized fact is that ETF baskets are highly concentrated compared to index portfolios. This fact may seem surprising, just like cash in baskets, given the stated objective

basket price adjustments, as revealed to us in private communication with ETF managers. Some of the negative amounts can also result from the imputation errors described earlier.

[6] Dannhauser and Hoseinzade (2022) note that bond ETFs hold some cash in their portfolios, but they do not discuss the amount of cash in ETF baskets, which is substantially larger, as we show here.

[7] We are grateful to Matthew Bartolini of State Street for his helpful guidance in this matter.

of passive ETFs to track the underlying index. A natural way for an ETF to ensure index-tracking would be to use CR and RD baskets representative of the index. Instead, ETFs use baskets that contain fewer securities, resulting in imperfect index tracking.

To assess a given basket's concentration on a given day, we divide the number of bonds in the basket by the number of bonds in the underlying index. We compute these "basket ratios" for both CR and RD baskets, reported as well as realized. We compute time-series averages of these ratios and report their cross-sectional distributions in Table 3.

Table 3 shows that realized baskets are highly concentrated relative to their indexes. The average realized basket ratios for CR and RD baskets are 24.6% and 29.8%, respectively. The medians are even lower, at 19.4% and 22.0%. These findings are consistent with those of Shim and Todorov (2023), who also examine realized baskets. Reported baskets, which are not examined by Shim and Todorov (2023), are less concentrated. The average reported basket ratio is about 76%—much larger than the averages for realized baskets, but much smaller than one. We thus see two dimensions of ETFs' active basket choice. First, ETFs pre-announce baskets that differ substantially from the index. Second, ETFs allow the APs to use realized baskets that differ substantially from the pre-announced baskets.

Table 3 also reports the cross-sectional distribution of analogous ratios that capture ETF portfolio concentration rather than basket concentration. These are ratios of the number of bonds in the ETF portfolio to the number of bonds in the index. The mean and median of these ratios are 81.3% and 88.0%, respectively. These values are larger than their counterparts for any of the previously discussed baskets, suggesting that ETFs actively update their baskets to keep their portfolios fairly close to the index.[8]

When ETFs choose individual bonds for inclusion in their baskets, they do not sample index bonds evenly. Rather, some bonds are persistently included, whereas others rarely appear. To show this fact, we calculate the basket inclusion probability for each bond-ETF pair as the number of times the bond appears in the ETF's CR or RD basket divided by the total number of baskets in which this bond could have appeared. We find that basket inclusion probabilities differ substantially across bonds. The median bond has only an 8% (13%) likelihood of appearing in its ETF's CR (RD) basket, whereas for the 99th percentile bond, this likelihood is 72% (88%). The Internet Appendix offers more details.

Just like cash in baskets, concentrated baskets help ETFs incentivize AP arbitrage. Through this channel, the use of concentrated baskets helps ETFs achieve their objective of

[8]For some ETFs, basket ratios exceed one. About 90% of the securities that are held by ETFs but absent from the index are fixed-income securities. The remaining securities are mostly cash equivalents.

liquidity transformation. We formalize these ideas in the following section.

4 Model

Motivated by the stylized facts uncovered in Section 3, we develop a simple model to shed light on the underlying economic mechanism. In the model, an ETF's optimal basket choice reflects a trade-off between the ETF's dual objectives of index tracking and liquidity transformation. Besides predicting the two main facts from Section 3, the model makes additional predictions regarding ETFs' cash and basket management. We test those predictions empirically in Section 5. All the proofs are in the Internet Appendix.

4.1 Model setup

The economy has three dates, $t = 0, 1, 2$, with no time discount. A unit measure of ex ante identical infinitesimal agents is born at $t = 0$, each endowed with one unit of a consumption good called cash. Cash is riskless, is liquid, and serves as the numeraire. The agents jointly form a representative ETF at $t = 0$, each holding an equal share of it. ETF shares trade in a competitive market at $t = 1$ at the market price p_E, which is determined in equilibrium. The ETF matures at $t = 2$ at the value v_E, also determined in equilibrium. Beyond the ETF, agents cannot access the underlying security market or any other investment technology to transfer wealth across time. This assumption captures the difficulty investors face in accessing illiquid security markets on their own. All agents have the same mean-variance utility function, $u(c) = \mathrm{E}(c) - (\rho/2)\mathrm{Var}(c)$, over their lifetime consumption c.[9]

Each agent privately learns their preferences at $t = 1$, becoming one of three types:

1. Impatient consumer: With probability π_c, the agent is subject to an idiosyncratic consumption shock at $t = 1$. The agent sells their ETF share at $t = 1$ and consumes the proceeds, obtaining utility $u(p_E)$ at that time.

2. Impatient saver: With probability π_s, the agent is subject to an idiosyncratic saving shock at $t = 1$. The agent buys an additional ETF share at $t = 1$, paying the price p_E per share. The agent holds two ETF shares until $t = 2$, at which time the agent consumes the ETFs' matured value $2v_E$, enjoying total utility $u(2v_E - p_E)$.

[9] The mean-variance assumption simplifies the exposition and facilitates analytical comparative statics. In the Internet Appendix, we show that the model's key predictions remain unchanged when we replace mean variance with constant absolute risk aversion (CARA) utility and solve the model numerically.

3. Patient investor: With probability $1 - \pi_c - \pi_s$, the agent faces no shock at $t = 1$. The agent consumes the matured value of one ETF share at $t = 2$, receiving utility $u(v_E)$.

The modeling of both consumption and saving shocks allows us to consider ETF creations and redemptions in a unified way, highlighting liquidity transformation in both cases. Consumption shocks correspond to liquidity shocks in Diamond and Dybvig (1983). Saving shocks are similar because they, too, reflect agents' idiosyncratic and immediate demand for liquidity (e.g., due to receiving a windfall that is not immediately consumed).

In the knife-edge case of $\pi_c = \pi_s$, the masses of impatient consumers and savers are equal, so the two groups trade with each other: impatient consumers sell their ETF shares to impatient savers. When $\pi_c \neq \pi_s$, though, the imbalance between the demand and supply for ETF shares in the secondary market must be met by their creation or redemption in the primary market. For this purpose, we introduce a deep-pocketed, risk-neutral, representative arbitrageur, or "AP."[10] When $\pi_c < \pi_s$, the AP meets the excess demand for ETF shares by creating new shares at $t = 1$ and immediately selling them to impatient savers at the price p_E. When $\pi_c > \pi_s$, the AP meets the excess supply of ETF shares by buying them from impatient consumers for p_E and immediately redeeming them at $t = 1$. The AP creates and redeems ETF shares in kind, by exchanging those shares for a basket of securities chosen by the ETF at $t = 0$. Our modeling of the AP's activity closely mirrors the real-world CR/RD process, which is unique to ETFs and available only to authorized participants.[11]

By meeting excess demand and supply in the secondary market for ETF shares, the AP makes these shares more liquid, supporting the liquidity transformation function of ETFs. The AP's presence in the market effectively reduces the price impact of the liquidity-demanding trades of impatient agents. The extent of the AP's liquidity provision can be inferred from the ETF's equilibrium market price p_E, as we explain later.

At $t = 0$, the newly formed ETF is endowed with an equal-weighted portfolio of N risky securities, which are identical ex ante. We normalize the N securities' values at $t = 0$ and $t = 1$ to unity. Each security's value at $t = 2$ is distributed as i.i.d. normal, $N(\mu, \sigma^2)$, with

[10] For simplicity, we abstract from the market structure of the AP sector, which is analyzed by Gorbatikov and Sikorskaya (2022).

[11] Our model does not apply to closed-end funds, for example, whose arbitrageurs do not have access to the primary market for fund shares. For simplicity, we assume that the AP transacts with the ETF sponsor to redeem any shares bought and create any shares sold in the secondary market. Relaxing this assumption would introduce one more decision for the AP—whether to hold ETF shares or transact in the primary market—which is not relevant to our analysis, as long as the AP can trade in the primary market. We note that the arbitrage process by APs on the primary market is also present in USD-backed stablecoins, except that stablecoin redemptions and creations are in-cash rather than in-kind (e.g., Ma, Zeng, and Zhang (2024)).

$\mu > 1$. We label the equal-weighted portfolio of the N securities as the "index."[12]

The N securities are not only risky but also illiquid, in that sourcing and trading them is costly to the AP at $t = 1$. There are two types of costs, fixed and variable.[13] First, sourcing a unique security entails a fixed cost $\lambda > 0$, so that a basket with I unique securities entails a total cost of λI. This fixed cost proxies for the search-and-matching costs typical of over-the-counter security markets. Second, regardless of the basket count, the AP incurs a variable cost of $\frac{1}{2}\phi s^2$, where $\phi > 0$, when holding s security units as a result of ETF creations or redemptions. The value of s can be positive or negative, capturing long or short security positions after redemptions or creations, respectively. This variable cost captures the AP's transaction costs resulting from liquidating the inventory associated with creations or redemptions, as well as any balance-sheet costs of carrying that inventory.[14]

The ETF manages its portfolio by designing the basket of securities that the AP can exchange for ETF shares. The same basket is used for both creations and redemptions. The ETF cannot trade securities directly; it can change its portfolio composition only through basket design. Consistent with this assumption, real-world ETFs rarely trade on their own because in-kind exchange of ETF shares for security baskets is more tax-efficient.

The ETF's basket design decision at $t = 0$ involves two choices. First, the ETF chooses α, the basket cash weight, where $0 \leq \alpha \leq 1$. The ETF basket holds α in cash and $1 - \alpha$ in risky securities. Second, the ETF chooses I, the number of risky securities in the basket, where $I \leq N$. Since all securities are identical ex ante, the choice among them boils down to the choice of I. We refer to I as the basket count. The ETF chooses α and I to maximize agents' expected aggregate welfare:

$$\max_{\alpha, I} \mathrm{E}[W] \equiv \underbrace{\pi_c \, u(p_E)}_{\substack{\text{impatient} \\ \text{consumers}}} + \underbrace{\pi_s \, E[u(2v_E - p_E)]}_{\substack{\text{impatient} \\ \text{savers}}} + \underbrace{(1 - \pi_c - \pi_s) \, \mathrm{E}[u(v_E)]}_{\substack{\text{patient} \\ \text{investors}}} . \tag{1}$$

By maximizing the welfare of its shareholders, the ETF maximizes its appeal to clients, which should enhance its fee revenue.[15] We do not endogenize management fees. We also assume

[12] In the Internet Appendix, we prove all of our theoretical results in a more general setting, in which the ETF's endowment portfolio can include not only the index portfolio but also some amount of cash.

[13] While we separate the two types of liquidity costs in our model, we cannot do so empirically due to the lack of data. We use the three bond liquidity measures described in Section 2.3 to proxy for both λ and ϕ. In practice, λ and ϕ are likely to be highly correlated because bonds that are harder to locate (i.e., higher λ) also tend to be costlier to trade (i.e., higher ϕ). Indeed, Hendershott et al. (2021) find a negative relation between quoting activity and trading costs in proprietary corporate bond data.

[14] We interpret the AP's costs broadly to include also frictions such as the minimum lot size.

[15] Welfare maximization can be interpreted as the maximization of the ETF's expected fee revenue under a fixed management fee. Suppose that at time $t = 0$, agent i faces a participation cost of η_i, whose cumulative distribution function across agents is $G(\eta)$, which is independent of agent type. The agent participates in

π_c and π_s are known, for simplicity. Abstracting from agency frictions and uncertainty about secondary-market imbalances allows us to sharpen our focus on liquidity transformation.

4.2 Equilibrium ETF pricing and liquidity transformation

We first solve for the equilibrium ETF price at $t = 1$, p_E, taking the ETF's basket cash weight α and basket count I as given. We solve for α and I later, in Section 4.3.

The ETF's market-clearing price ensures that the AP is willing to clear the imbalance between the demand and supply of ETF shares by the impatient agents. If there is excess demand for ETF shares (i.e., $\pi_s - \pi_c > 0$), the AP must create $\pi_s - \pi_c$ new ETF shares and sell them to impatient savers. If there is excess supply (i.e., $\pi_c - \pi_s > 0$), the AP must redeem $\pi_c - \pi_s$ shares after buying them from impatient consumers. Let x denote the number of ETF shares bought and redeemed by the AP at $t = 1$ (where $x > 0$ for redemptions and $x < 0$ for creations). The AP chooses x to maximize its profit function, $\Gamma(x)$:

$$\max_x \Gamma(x) \equiv [\alpha + (1-\alpha)\mu - p_E]\, x - \frac{1}{2}\phi\,[(1-\alpha)x]^2 - \lambda I\,, \tag{2}$$

subject to the participation constraint $\Gamma(x) \geq 0$. The first term in Equation (2) captures the AP's expected arbitrage profit. Suppose $x > 0$. After buying x ETF shares from impatient consumers at the price of p_E per share, the AP exchanges these shares for x units of the ETF basket. This basket contains α units of cash and $\frac{1-\alpha}{I}$ units of each of I risky securities, so its expected payoff is $\alpha + (1-\alpha)\mu$. The AP's profit per ETF share is therefore $\alpha + (1-\alpha)\mu - p_E$, yielding the first term. If $x < 0$, the same logic applies, except that the AP creates ETF shares and sells them for p_E. The second term in Equation (2) reflects the AP's variable costs associated with carrying and transacting $1-\alpha$ units of risky basket securities for each of the x ETF shares created or redeemed. Finally, the last term represents the fixed costs associated with sourcing I distinct basket securities in the market.

From the first-order condition corresponding to Equation (2), we obtain

$$x^* = \frac{\alpha + (1-\alpha)\mu - p_E}{\phi(1-\alpha)^2}\,. \tag{3}$$

When ETF shares are cheap relative to the basket, so that $p_E < \alpha + (1-\alpha)\mu$, the AP buys x^* ETF shares, redeems them, and sells the basket securities. When the ETF is expensive relative to the basket, so that $p_E > \alpha + (1-\alpha)\mu$, the AP purchases the basket, creates x^*

the ETF market if $\eta_i < \mathrm{E}[W]$. So, the mass of agents who participate is $G(\mathrm{E}[W])$. If a proportional fee of ϕ is levied at $t = 0$, then the ETF maximizes $\phi G(\mathrm{E}[W])$, which is equivalent to the problem in Equation (1).

ETF shares, and sells them. Either way, the AP makes money, before transaction costs. The AP will execute this trade only if the profit net of transaction costs is also positive.

The ETF's equilibrium price equates supply and demand for ETF shares. We obtain this price by combining Equation (3) with the market-clearing condition $x = \pi_c - \pi_s$.

Proposition 1. The ETF's market-clearing price at time 1 is given by

$$p_E = \underbrace{\alpha + (1-\alpha)\mu}_{\text{basket NAV}} + \underbrace{\phi(1-\alpha)^2(\pi_s - \pi_c)}_{\text{premium/discount}} . \tag{4}$$

The ETF market clears at time 1 if and only if $I \leq \bar{I}$, where $\bar{I}$ is the basket count that makes the AP break even (i.e., for which the supremum of $\Gamma(x)$ in Equation (2) is zero).

The first term in Equation (4) is the basket's expected payoff, which is also its net asset value from the perspective of the risk-neutral AP. The second term is the ETF premium or discount, as perceived by the AP. From the AP's viewpoint, the ETF trades at a premium if and only if $\pi_s > \pi_c$, which indicates excess demand for ETF shares in the secondary market and leads to net ETF creation in the primary market. Similarly, the ETF trades at a discount if $\pi_s < \pi_c$, which indicates excess supply of ETF shares and net redemption.

We interpret the ETF premium/discount as the equilibrium price impact of impatient traders. A discount means that excess supply of ETF shares by impatient consumers has pushed p_E down, reducing those consumers' utility, $u(p_E)$. A premium means that excess demand by impatient savers has pushed p_E up, increasing their disutility. The magnitude of the premium/discount thus captures the degree of liquidity transformation performed by the AP. The lower the magnitude, the lower is the price impact of impatient traders in the secondary ETF market, and the higher is the ETF liquidity transformation.[16]

From Equation (4), the magnitude of the ETF premium/discount increases with ϕ but decreases with α. Both effects are intuitive under the price impact interpretation. A larger cash weight α indicates a more liquid basket, which reduces the AP's variable transaction costs. Given the lower cost of providing liquidity, the AP chooses to provide more of it, resulting in a smaller price impact in the ETF market. In contrast, a larger ϕ means that the risky securities are more costly for the AP to trade. The AP then provides less liquidity and the price impact is larger. This model prediction is consistent with the empirical finding that ETF premiums and discounts tend to be larger for ETFs holding less liquid securities (e.g., Petajisto 2017; Pan and Zeng 2019; Shim and Todorov 2023).

[16] The ETF premia/discounts observed in the data can also reflect stale pricing of the underlying securities. However, such considerations are absent from Equation (4) because stale pricing is absent from our model.

The price impact is infinite, and no liquidity transformation takes place, if the AP's participation constraint is not satisfied. In that case, any excess demand or supply from impatient agents remains unmet and the ETF market fails to clear. As noted in Proposition 1, the basket count must be sufficiently low to incentivize the AP to trade. We thus see that not only a larger cash weight but also a lower basket count facilitate liquidity transformation. However, they come at a cost, as we explain in the following section.

4.3 Trade-off between liquidity transformation and index tracking

Recall that the ETF's decision at $t = 0$ involves choosing the basket cash weight, α, and basket count, I. The ETF faces a trade-off: increasing α or decreasing I improves liquidity transformation, as explained earlier, but it comes at the expense of index tracking. To understand this, recall that the ETF is endowed with the index portfolio containing all N securities. If the basket includes only a subset of those securities (i.e., $I < N$), then the ETF's portfolio after creations or redemptions will deviate from the index. Similarly, if the basket includes cash, so will the ETF's post-CR/RD portfolio, again deviating from the (cash-free) index. While index tracking per se does not enter agents' preferences, it is related to the welfare of patient investors, as we explain here.

We solve the ETF's optimal basket design problem at $t = 0$. Substituting the results from Section 4.2 into Equation (1), we re-express the objective function as

$$\max_{\alpha, I} \mathrm{E}[W] \;=\; (\pi_c - \pi_s) p_E \;+\; V(\mu_E, \sigma_E^2) \,, \tag{5}$$

where p_E is in Equation (4), μ_E and σ_E^2 are the mean and variance of the ETF portfolio's time-2 payoff, and $V(\mu_E, \sigma_E^2)$ is a function increasing in μ_E and decreasing in σ_E^2. Explicit formulas for μ_E, σ_E^2, and $V(\mu_E, \sigma_E^2)$ are in the Internet Appendix.

Equation (5) sheds more light on the trade-off between liquidity transformation and index-tracking. The first term captures liquidity transformation; the second index tracking. The first term, $(\pi_c - \pi_s)p_E$, echoes our prior discussion of price impact. When there is excess demand for ETF shares (i.e., $\pi_c - \pi_s < 0$), the ETF wants to push p_E down to reduce the price impact of the impatient savers' buying pressure. When there is excess supply (i.e., $\pi_c - \pi_s > 0$), the ETF wants to push p_E up to counter the price impact of the selling pressure. The value of p_E is determined by the amount of liquidity that the AP is willing to provide. The ETF can incentivize the AP to provide more liquidity by using more cash or fewer securities in the basket (i.e., higher α or lower I).

However, higher α or lower I imply a larger index deviation at $t = 2$, as explained earlier. This larger index deviation affects the second term in Equation (5). In particular, σ_E^2 increases when I decreases, holding α constant. With fewer securities in the basket, the ETF portfolio will have a larger variance than the equal-weighted index at $t = 2$, due to reduced diversification. While I has no effect on μ_E, α affects both μ_E and σ_E^2. A larger α implies more (less) cash in the ETF portfolio at $t = 2$ after creations (redemptions). More cash in the portfolio implies a lower expected payoff, μ_E, because risky securities have higher expected returns than cash ($\mu > 1$), but also lower variance, σ_E^2, because cash is risk-free. The net effect of α on the welfare of mean-variance investors is thus ambiguous. However, the poorer index tracking caused by a lower I unambiguously reduces the welfare of patient investors because a decrease in I increases σ_E^2, which in turn reduces $V(\mu_E, \sigma_E^2)$.

Solving the ETF's basket design problem in Equation (5), we obtain the following results.

Proposition 2. If the illiquidity parameters ϕ and λ are sufficiently large, then

(i) The ETF optimally chooses a positive basket cash weight, $\alpha^* > 0$.

(ii) The ETF optimally chooses a concentrated basket, $I^* < N$.

(iii) The optimal basket cash weight, α^*, increases with ϕ.

(iv) The index-tracking error of the ETF's equilibrium security portfolio increases with λ.

Parts (i) and (ii) of Proposition 2 rationalize the two stylized facts presented in Section 3. Parts (iii) and (iv) show that the tension between liquidity transformation and index tracking is exacerbated by the illiquidity of the underlying securities. This illiquidity makes it costlier for the AP to provide liquidity in the ETF market. The result is lower ETF liquidity, which increases the price impact of impatient investors, reducing their welfare. To restore the balance between the welfare of patient and impatient investors, the ETF incentivizes more liquidity provision by the AP. Specifically, larger variable transaction costs (ϕ) lead the ETF to raise its basket cash weight, whereas larger fixed transaction costs (λ) induce it to use a more concentrated basket, resulting in a larger tracking error.

Parts (iii) and (iv) of Proposition 2 generate additional testable predictions that go beyond the stylized facts from Section 3. Part (iii) directly motivates the following hypothesis.

Hypothesis 1. The proportion of cash in an ETF's baskets increases when the underlying securities are less liquid.

Part (iv) of Proposition 2 directly motivates the following hypothesis.

Hypothesis 2. The tracking error of an ETF's security portfolio is larger when the underlying securities are less liquid.

Proposition 2 also helps us make predictions about how ETFs adjust their portfolios over time, building on the existing literature on dynamic portfolio choice. Garleanu and Pedersen (2013) analyze dynamic trading behavior in the presence of transaction costs. They show that the optimal portfolio strategy involves patient trading that brings the current portfolio closer to its moving target. Their analysis does not consider intermediated trading, whereas a key feature of our model is an endogenous wedge between an ETF's portfolio and its basket used by an intermediary—namely, the AP. This wedge is consistent with our observation in Section 3 that baskets often deviate from portfolios. Many such deviations are driven by the ETF's intention to adjust its portfolio back to the underlying index. ETF portfolios often deviate from the index, for reasons such as index rebalancing. Thus, we can naturally hypothesize that when an ETF's portfolio deviates from the index, the ETF favors baskets that push the portfolio back to the index.

Specifically, suppose an ETF holds too much of a given security relative to the underlying index. The ETF could in principle sell some of this security in the secondary market, but such transactions are not commonly done in practice because they trigger tax liabilities. Instead, it can be more efficient for the ETF to increase the security's weight in the RD basket. Upon delivering this RD basket to an AP in exchange for ETF shares, the ETF reduces its holdings of this security in a tax-free manner (because the exchange is in-kind). The same objective can be achieved by reducing the weight of this security in the CR basket, because new creations then bring less of this security to the ETF's portfolio.

The same logic, in reverse, applies when an ETF holds too little of a given security relative to the index. The ETF can then increase the security's weight in the CR basket, and reduce it in the RD basket, to steer the ETF portfolio back to the index. The logic applies not only to securities but also to cash holdings. We thus obtain the following hypotheses about the active management of ETF baskets over time.

Hypothesis 3. When cash is overweighted in an ETF's portfolio, the ETF increases its weight in the RD basket and decreases it in the CR basket. Such adjustments are attenuated when the underlying securities are less liquid.

Hypothesis 4. When a given security is overweighted in an ETF's portfolio, the ETF increases its weight in the RD basket and reduces it in the CR basket. Such adjustments are attenuated when the underlying securities are less liquid.

Both Hypotheses 3 and 4 are composed of two parts. The first part of each hypothesis naturally follows from the preceding arguments. The second part, which pertains to liquidity, is directly motivated by our model. While the model does not explicitly incorporate dynamic basket adjustments, it does explain how liquidity transformation comes at the expense of such adjustments, that is, at the expense of index tracking. Proposition 2 shows that ETFs tolerate larger index deviations when index securities are less liquid. Suppose an ETF's portfolio is currently overweighting certain securities, or cash, relative to the target. To improve index tracking, the ETF will adjust its baskets to reduce this overweighting. However, these basket adjustments will be smaller if the underlying securities are costlier to trade, because the ETF needs the AP to be willing to trade those securities to exchange them for the baskets. Instead of making aggressive basket adjustments that would correct index deviations immediately, the ETF tolerates some tracking error to deliver more liquidity transformation. This prediction echoes the result of Garleanu and Pedersen (2013) that when trading is costly, it is optimal to trade only partially toward the portfolio target.

Returning to the underlying theoretical results, the Internet Appendix presents formulas for α^* and I^*, the solutions to the ETF's optimal basket design problem. Their main properties are summarized in Proposition 2. In addition, basket count I^* has an ambiguous dependence on liquidity (increasing in ϕ, decreasing in λ). Shim and Todorov (2023) empirically analyze the relations between bond ETF basket count and variables such as bond liquidity and ETF flows. While the basket cash proportion, α^*, displays sensible comparative statics, its magnitude can be unrealistically large for some parameter values, as we show in the Internet Appendix. This is an artifact of mean-variance utility. When we use CARA utility instead, this undesirable property disappears. The CARA specification generates smaller cash proportions while preserving all of our key theoretical predictions. However, this quantitative benefit comes at the cost of losing analytical tractability. Therefore, we keep the tractable mean-variance setting as the baseline and present the CARA extension in the Internet Appendix.

Our simple model could also be extended in other ways. By adding more time periods, one could analyze basket dynamics more formally than we do in this section. By relaxing the assumption that the risky securities are identical ex ante, one could look for additional implications about within-basket heterogeneity. By explicitly modeling the AP's potential role as a market maker in the underlying securities, one could explore the links between optimal basket composition and the AP's current inventory. As these extensions would significantly complicate the model, we leave them for future research.

5 Cash and Basket Management

In this section, we test the four hypotheses developed in Section 4. We test Hypothesis 1 in Section 5.1, Hypothesis 2 in Section 5.2, etc. Our evidence supports all four hypotheses. We find that ETFs tracking less liquid indexes use more cash and have larger tracking errors. We also find that ETFs actively steer their portfolios toward their indexes, and that their capacity to do so is constrained by the illiquidity of their holdings.

5.1 Cash and liquidity

According to Hypothesis 1, ETFs tracking less liquid indexes should have more cash in their baskets. Intuitively, trading less liquid bonds imposes larger transaction costs on APs, so that more basket cash is needed to encourage arbitrage trading by APs.

To test the hypothesis, we relate $Cash_{jt}$, the proportion of cash in the reported basket of ETF j on day t (i.e., the cash ratio), to $Illiquidity_{jt}$, the average illiquidity of the non-cash securities in the ETF's portfolio, value-weighted by the amount of bonds held. We use portfolio illiquidity to proxy for index illiquidity for our cross-sectional analysis because index data are available for only about half of our ETFs, as noted earlier. We exclude ETF-days with cash ratios exceeding 20%, to suppress outliers. We drop ETFs that track indexes with target maturity dates (or target-date ETFs) because such ETFs tend to increase their cash holdings as the maturity date approaches.

Figure 2 shows scatterplots of $Cash_{jt}$ against $Illiquidity_{jt}$ in panels A, C, and E. The three panels correspond to our three measures of liquidity (see Section 2.3). To reduce noise in the data, we sort observations by $Illiquidity_{jt}$ into 20 equal-size bins. We then compute the average values of $Cash_{jt}$ and $Illiquidity_{jt}$ within each bin and plot one against the other. In each panel, the relation between $Cash_{jt}$ and $Illiquidity_{jt}$ is clearly positive, indicating that ETFs tracking less liquid indexes tend to have larger cash ratios.

Each panel of Figure 2 also plots the line of best fit from the regression

$$Cash_{jt} = \beta\, Illiquidity_{jt} + \omega_{It} + \kappa s_{jt} + \epsilon_{jt}\,, \tag{6}$$

where ω_{It} is an issuer-time fixed effect and s_{jt} is ETF j's bid-ask spread in the secondary market. By including an issuer-time fixed effect, we remove time-varying differences in cash management across ETF issuers. We are effectively estimating a cross-sectional correlation, comparing ETFs from the same issuer on the same day. By including s_{jt}, we control for the secondary-market liquidity of ETF shares.

The β estimates—slopes of the lines plotted in panels A, C, and E—range from 0.22 to 0.38. These estimates imply that a one-standard-deviation increase in $Illiquidity_{jt}$, which is standardized to unit variance, is associated with an increase in the cash ratio of 22 to 38 bps, depending on the liquidity measure.[17] These estimates are economically significant relative to the 109 bps median cash ratio for reported baskets. All three estimates are also statistically significant at the 1% level, as we show in the Internet Appendix.[18]

We prioritize reported baskets over realized ones in this test because reported baskets are available every day, whereas realized baskets can be imputed only on days with CR or RD activity. Such activity is often induced by buying or selling pressure, which could be correlated with ETF-level shocks that simultaneously affect basket cash decisions. This could introduce bias when comparing cash across ETFs on the same day, because any given day has realized-basket observations only for a non-randomly-selected subset of ETFs. Another disadvantage of realized baskets is a smaller number of observations, as noted earlier.

With these caveats, we also estimate regression (6) using cash in realized baskets. We exclude realized baskets containing fewer than 10 securities, both here and in subsequent analysis, because such small baskets are more likely to be driven by imputation errors or trades occurring outside the AP process. We find results similar to those based on reported baskets, with stronger economic significance but weaker statistical significance. As we show in the Internet Appendix, the β estimates are positive in all six regression specifications (three liquidity measures times two baskets, CR and RD). In five of the six specifications, the β estimate is larger based on realized baskets, indicating stronger economic significance. However, only two of the six β estimates are statistically significant at the 5% level.[19] Overall, based on both reported and realized baskets, we find that ETFs holding less liquid bonds use more cash in their baskets, as predicted by Hypothesis 1.

While Hypothesis 1 applies to basket cash, it also carries implications for cash in ETFs' portfolio holdings. In periods of sustained net ETF share creation, more cash in baskets translates into more cash in holdings. Therefore, in such periods, ETFs tracking less liquid indexes should hold more cash not only in their baskets but also in their portfolios. In our

[17] To help interpret a one-standard-deviation increase in $Illiquidity_{jt}$, in March 2020, bond illiquidity rose to levels between two and five standard deviations above average, depending on the liquidity measure.

[18] The results are not driven by small ETFs. If we exclude the smallest 10% of ETFs from the analysis, all of the results, not only here but throughout Section 5, continue to hold.

[19] When we include not only issuer-time but also investment-grade-time fixed effects, the results become somewhat weaker. This is not surprising because the latter fixed effects absorb some of the variation of interest as investment-grade bonds tend to be more liquid than high-yield bonds. Nonetheless, even within investment-grade and high-yield ETFs, we observe positive slopes in five out of six panels. See the Internet Appendix.

sample period of 2017 through 2020, bond ETFs have indeed experienced a large amount of net creation, as we document later in Section 6.3. Therefore, Hypothesis 1 seems relevant also to cash in ETF portfolios, albeit to a lesser extent.

To examine this version of Hypothesis 1, we repeat the previous analysis replacing cash in baskets by cash in holdings, so that $Cash_{jt}$ becomes the proportion of cash in the portfolio of ETF j on day t. We include only ETF-days on which the ETF experienced no creations or redemptions to better capture the ETF's equilibrium level of cash holdings. The results are reported in panels B, D, and F of Figure 2. The β estimates range from 0.12 to 0.17, indicating that a one-standard-deviation increase in $Illiquidity_{jt}$ is associated with an increase in the cash ratio of 12 to 17 bps. These estimates are not as large as those in panels A, C, and E (visually, the slopes in the right-hand panels are not as steep as those in the left-hand panels), but they are sizable relative to the median cash ratio of 83 bps. Moreover, all three estimates are significant at the 1% level. ETFs holding less liquid bonds hold more cash in their portfolios, as predicted.

5.2 Tracking error and liquidity

According to Hypothesis 2, ETFs tracking less liquid indexes should track less closely. As trading illiquid bonds is costly to APs, ETFs holding such bonds tolerate larger tracking errors to accommodate APs' demands and thereby incentivize APs' arbitrage activity.

For each ETF j and day t, we compute $TrackingError_{jt}$ as the standard deviation of the difference between the ETF's daily return and the index return over the past month. As in our prior holdings-based analysis, we exclude ETF-days with creations and redemptions and target-date ETFs. We sort ETF-day observations into 20 equal-size bins by $Illiquidity_{jt}$, compute the average values of $TrackingError_{jt}$ and $Illiquidity_{jt}$ within each bin, and plot them in panels A, C, and E of Figure 3. The three panels correspond to the three liquidity measures used before. Each panel also plots the line of best fit from the regression

$$TrackingError_{jt} = \beta\, Illiquidity_{jt} + \omega_{It} + \kappa s_{jt} + \epsilon_{jt}\,, \quad (7)$$

where ω_{It} is an issuer-time fixed effect and s_{jt} is ETF j's bid-ask spread in the secondary market.[20] As in Equation (6), by including ω_{It}, we compare ETFs from the same issuer on the same day, and by including s_{jt}, we control for the ETF's secondary-market liquidity.

[20]We are recycling some of the notation from Equation (6)—β, ω_{It}, s_{jt}, and ϵ_{jt}—even though there is no direct relation between Equations (6) and (7). To economize on notation, we also engage in similar recycling in all subsequent regression equations. Throughout the paper, we use β as generic notation for a regression slope, ω as generic notation for a fixed effect, and ϵ as generic notation for the error term, with various subscripts or superscripts. The various β's, ω's, and ϵ's are not related across regression equations.

Figure 3 shows a strongly positive relation between $TrackingError_{jt}$ and $Illiquidity_{jt}$. In each panel, the fitted regression line slopes upward, indicating larger index deviations for ETFs tracking less liquid indexes. All three β estimates are positive and statistically significant at the 1% level, as we show in the Internet Appendix. The Internet Appendix also shows that the results remain similar after adding investment-grade-time fixed effects.

One potential concern is the role of stale bond prices. $TrackingError_{jt}$ is calculated from the difference between ETF returns and index returns. While ETF returns are computed from ETF prices, which change frequently, index returns are computed from the underlying bond prices, which change infrequently. To remove this mismatch, we use an alternative measure of tracking error that computes ETF returns from NAVs, which are based on the underlying bond prices rather than the ETF share price. In this approach, stale pricing may still be present, but it should affect both ETF and index returns symmetrically. Panels B, D, and F of Figure 3 present results for this alternative measure of tracking error. The relation of interest remains positive and significant at the 1% level. A one-standard-deviation increase in $Illiquidity_{jt}$ is associated with an ETF tracking error that is 25 to 42 percentage points of a standard deviation larger. Overall, ETFs investing in less liquid bonds deviate more from their indexes, consistent with Hypothesis 2.

5.3 Active cash management

In this section, we analyze how ETFs actively manage their basket cash over time. According to Hypothesis 3, ETFs dynamically adjust their baskets to revert to a certain desired level of cash, which may or may not be zero. ETFs holding too much cash respond by adding cash to RD baskets and removing it from CR baskets, and vice versa. This active cash management is constrained by the illiquidity of the underlying securities.

To test Hypothesis 3, we construct three variables. $BasketCash_{jt}$ is the proportion of cash in a CR or RD basket of ETF j on day t. We use realized baskets because they exhibit much more time variation than reported baskets. As in our prior basket-related analysis, we exclude baskets with fewer than 10 securities and target-date ETFs. $\Delta Cash_{jt-1}$ is the difference between the fraction of cash in the ETF's portfolio on day $t-1$ and the average of those fractions over the prior month. Interpreting this average as the equilibrium level of cash, $\Delta Cash_{jt-1}$ captures the ETF's over- or underweighting of cash. Finally, $Illiquidity_j$ is ETF j's average index illiquidity, which is computed by first calculating the value-weighted average illiquidity of the ETF's index constituents and then averaging over time. We estimate how $BasketCash_{jt}$ varies with $\Delta Cash_{jt-1}$ and its interaction with $Illiquidity_j$:

$$BasketCash_{jt} = \beta_1 \, \Delta Cash_{jt-1} + \beta_2 \, \Delta Cash_{jt-1} \times Illiquidity_j + \omega_j + \kappa s_{jt} + \epsilon_{jt} \, , \qquad (8)$$

where ω_j is an ETF fixed effect and s_{jt} is ETF j's bid-ask spread in the secondary market. By including ω_j, we are effectively running a series of time-series regressions, one for each ETF. Controlling for s_{jt} further ensures that our beta estimates are unaffected by any potential correlated variation in the ETF's secondary-market liquidity.

Table 4 reports the estimates of β_1 and β_2. We find that $\beta_1 < 0$ for CR baskets and $\beta_1 > 0$ for RD baskets. These results indicate that when cash is overweighted in an ETF's portfolio (i.e., $\Delta Cash_{jt-1} > 0$), the ETF tends to remove cash from CR baskets and add it to RD baskets. Such basket adjustments help ETFs steer their cash holdings toward their desired long-term average. We also find that $\beta_2 > 0$ for CR baskets and $\beta_2 < 0$ for RD baskets, indicating that the above basket adjustments are attenuated for ETFs tracking less liquid indexes. All of these results are consistent with Hypothesis 3.

The estimates of β_1 and β_2 are statistically significant for both CR and RD baskets and for all three measures of liquidity. They are also economically significant. A one-standard-deviation increase in $\Delta Cash_{jt-1}$ is associated with 3.3 to 3.6 percentage points (pps) less cash in CR baskets and 3.1 to 3.2 pps more cash in RD baskets for an ETF with the average level of illiquidity. When the illiquidity is one standard deviation larger, the effect is mitigated by 1.3 to 1.7 pps for CR baskets and 1.1 to 1.4 pps for RD baskets.

5.4 Active basket management

Finally, we test Hypothesis 4, according to which ETFs actively manage the composition of their baskets to steer their portfolios toward the index. ETFs holding too much of a given bond respond by adding this bond to RD baskets and removing it from CR baskets, and vice versa. These adjustments are smaller when the bonds are less liquid.

To test this hypothesis, we use index rebalancing as a source of plausibly exogenous variation in the over- or underweighting of bonds in ETF portfolios. Using such variation helps alleviate the concern that a bond's over- or underweighting could be driven by unobserved characteristics that could also cause the bond's basket inclusion. We take advantage of the fact that fixed-income indexes rebalance on the last day of each month. While changes in the bond universe and bond characteristics occur throughout the month, index composition is not updated until the month-end to reflect those changes. The monthly jump in index portfolio weights therefore constitutes a plausibly exogenous shock to the over- or underweighting of bonds in ETF portfolios relative to the index. Conversations with ETF managers indicate that the changes in index weights are not fully predictable. Even though some of them are predictable, ETFs incur tracking error if they make anticipatory portfolio adjustments before the month-end. Therefore, ETFs often postpone such adjustments until the month-end. Indeed, the difference between ETF portfolio weights and index weights tends to spike at the monthly index rebalancing dates, as we show in the Internet Appendix. The Internet Appendix also contains a more detailed discussion of the rebalancing of fixed-income indexes.

Let $Deviation_{ijt}$ denote the difference between bond i's weight in ETF j's portfolio and the bond's weight in the index on day t. We refer to this difference as the bond's "overweighting" in the ETF's portfolio relative to the index, with the understanding that negative values represent underweighting.[21] For each index rebalancing day h, we compute each bond's overweighting in excess of its average overweighting over the previous week:

$$\Delta Deviation_{ijh} \equiv Deviation_{ijh} - \frac{1}{5}\sum_{k=1}^{5} Deviation_{ijh-k}\,. \tag{9}$$

We interpret $\Delta Deviation_{ijh}$ as a rebalancing-induced shock to the bond's overweighting. We relate this shock to the bond's basket inclusion by estimating the regression models

$$Basket^{CR}_{ijt} = \beta^{CR}_1\,\Delta Deviation_{ijh} + \beta^{CR}_2\,\Delta Deviation_{ijh} \times Illiquidity_j + \omega_{jt} + \epsilon_{ijt} \tag{10}$$

$$Basket^{RD}_{ijt} = \beta^{RD}_1\,\Delta Deviation_{ijh} + \beta^{RD}_2\,\Delta Deviation_{ijh} \times Illiquidity_j + \omega_{jt} + \epsilon_{ijt}\,, \tag{11}$$

where $Basket^{CR}_{ijt}$ and $Basket^{RD}_{ijt}$ measure bond i's basket inclusion (we consider two measures of basket inclusion, as detailed below), $Illiquidity_j$ is ETF j's index illiquidity, as before, and ω_{jt} represent ETF-day fixed effects. Given these fixed effects, we exploit cross-sectional variation in basket inclusion across bonds within the same ETF on the same day.

Our sample for this analysis consists of observations at the ETF-bond-day level on days on which the ETF has a realized basket.[22] We split the sample by whether this basket is a CR basket (to estimate Equation (10)) or an RD basket (to estimate Equation (11)). For each ETF, we include in the sample observations corresponding to the first 10 baskets after each monthly rebalance date. This choice amounts to using the month's first 10 trading days on which the ETF exhibits net CR or RD activity (recall that we have one realized basket per trading day with CR/RD activity). We use the first few baskets after each index rebalancing because they are the most likely to respond to the rebalancing. We do not use

[21] The value of $Deviation_{ijt}$ is missing if bond i is held by neither ETF j nor its index on day t. If the bond is not held by the ETF but is in the index, $Deviation_{ijt}$ is negative to indicate underweighting. If the bond is held by the ETF but is not in the index, $Deviation_{ijt}$ is positive to indicate overweighting.

[22] In the Internet Appendix, we also implement this analysis differently, by collapsing observations to the ETF-bond-rebalancing-instance level by basket type. We find results similar to those presented in Table 4.

more than 10 baskets because the first 10 baskets are likely to close much of the gap between the portfolio and the index, and we do not want to use baskets unaffected by the rebalancing. If fewer than 10 baskets are available for the given month, we use all available baskets in that month. On average, we use observations from six baskets following a rebalance. There are 1,227 monthly rebalance instances in our sample. In 67% of them, the ETF has 10 or fewer baskets in the subsequent month.

Panel A of Table 5 shows the results when $Basket^{CR}_{ijt}$ ($Basket^{RD}_{ijt}$) is a dummy variable equal to one when bond i is included in ETF j's CR (RD) basket on day t. We see that $\beta^{CR}_1 < 0$ and $\beta^{RD}_1 > 0$, indicating that when a bond becomes more overweighted in an ETF's portfolio after the rebalancing of the ETF's index, the bond is less likely to be included in CR baskets but more likely to be included in RD baskets. These results are consistent with ETFs steering their portfolios back to the index after deviations caused by index rebalancing. We also find that $\beta^{CR}_2 > 0$ and $\beta^{RD}_2 < 0$, indicating that the aforementioned basket adjustments are attenuated when the bonds held by the ETF are less liquid. All of these results are highly statistically significant and fully consistent with Hypothesis 4. The results are also economically significant. For example, a one-standard-deviation increase in rebalancing-induced overweighting for a given bond is associated with a 3% lower likelihood of that bond's presence in CR baskets for an ETF of average index illiquidity. The effect is attenuated by about 0.5% when the illiquidity is one standard deviation higher.

Panel B of Table 5 shows the results when $Basket^{CR}_{ijt}$ ($Basket^{RD}_{ijt}$) is measured by the number of shares of bond i in ETF j's CR (RD) basket. In light of the right skewness in these measures, we estimate Equations (10) and (11) by Poisson regression, following the recommendation of Cohn, Liu, and Wardlaw (2022). Given the heterogeneity in issuance across bonds, we add a control for the log number of shares outstanding of bond i. The results are very similar to those in panel A in that $\beta^{CR}_1 < 0$, $\beta^{RD}_1 > 0$, $\beta^{CR}_2 > 0$, and $\beta^{RD}_2 < 0$. All 12 slope estimates have the predicted signs and 10 of them are significant at the 5% level.

6 The Effect of ETF Basket Inclusion on Bond Liquidity

Having examined the active management of passive ETFs, we ask whether and how this activeness impacts bond liquidity. This is a natural question to ask. Recall that ETFs' activeness facilitates AP arbitrage by reducing the APs' transaction costs. Early theoretical research on index arbitrage predicts that enhanced arbitrage activity improves the liquidity of the underlying securities through increased trading (Fremault 1991; Kumar and Seppi

1994). This mechanism is likely to be particularly strong in the context of bond ETFs whose APs tend to be market makers for the underlying bonds (Bessembinder et al. 2018; Pan and Zeng 2019; He, Khorrami, and Song 2022). Therefore, prior research combined with our own suggests the hypothesis that ETFs' active management enhances the liquidity of the underlying bonds. We now examine this hypothesis, given its strong theoretical motivation.

We estimate the effect of basket inclusion on individual bond liquidity in two ways. In Section 6.1, we estimate the relation between basket inclusion and subsequent bond liquidity in the presence of controls and fixed effects. In Section 6.2, we use an instrument for basket inclusion to estimate its causal effect on bond liquidity by 2SLS. Using both approaches, we find that a bond's inclusion in an ETF basket makes the bond more liquid. However, basket inclusion makes the bond less liquid in periods of large imbalance between creations and redemptions, as we show in Section 6.3.

6.1 Baseline analysis

Our baseline analysis relates an individual bond's illiquidity on a given day to the bond's basket inclusion on the previous day. Let $Illiquidity_{it+1}$ denote the illiquidity of bond i on day $t+1$, for any of our three illiquidity measures. Let $Basket_{it}^{CR}$ denote the extent of bond i's inclusion in ETFs' realized CR baskets on day t. We consider two measures of basket inclusion: first, a dummy variable equal to one if the bond is included in at least one ETF's basket, and second, the number of ETFs that include this bond in their CR baskets on this day.[23] $Basket_{it}^{RD}$ is defined analogously, except for RD baskets. We estimate the model

$$Illiquidity_{it+1} = \beta^{CR}\, Basket_{it}^{CR} + \beta^{RD}\, Basket_{it}^{RD} + Controls_{it} + \omega_{Ft} + \omega_{mt} + \epsilon_{it}\,. \quad (12)$$

The controls include the average $Illiquidity_{it}$ over the prior week to address the possibility that bonds selected into baskets may be more liquid ex ante.[24] We further control for the average basket size of ETFs holding bond i because baskets containing few bonds may reflect imputation errors or adjustments that do not operate through ETF arbitrage. Importantly, we include firm-day fixed effects, ω_{Ft}, and maturity-day fixed effects, ω_{mt}. The former fixed effects control for time-varying fundamental shocks at the firm level, whereas the latter account for time-varying yield-curve effects. By including these fixed effects, we are comparing

[23] In the Internet Appendix, we also consider two measures based on the number of bond shares included in ETF baskets: first, the log of one plus the number of shares of this bond that are included in ETFs' CR baskets on this day, and second, the number of shares of this bond included across all ETF baskets scaled by the overall number of bond shares included in all ETF baskets on this day. The results are similar.

[24] In the same spirit, Brogaard, Heath, and Huang (2024) find that more liquid stocks are more likely to be selected into equity ETF baskets. They also study the effects of primary flows in equity ETFs on the underlying stocks.

near-identical bonds with different basket inclusions on the same day.

Our sample includes all bonds present in TRACE and Mergent that are held in at least one ETF portfolio. We remove observations in which a bond appears in CR baskets for the first time or in RD baskets for the last time. We do this to remove mechanical basket inclusions because a bond newly added to the index can only appear in the ETF portfolio by being part of the CR basket, and a bond newly excluded from the index can only leave the ETF portfolio by being part of the RD basket. The remaining variation in basket inclusion better represents ETFs' discretionary basket management.

Table 6 shows that the estimates of β^{CR} and β^{RD} are both negative, for both basket inclusion measures and all three liquidity measures. All 12 of these estimates are statistically significant at the 1% level. These results are consistent with the hypothesis that inclusion in ETF baskets improves a bond's liquidity.

To interpret the magnitudes of the estimates, we standardize $Illiquidity_{it}$ by its mean and standard deviation, both computed across all of its bond-day observations, and multiply its values by 100 to improve the readability of the coefficients. The estimates imply that being included in a CR (RD) basket of one more ETF is associated with a decrease in the bond's next-day illiquidity by 1.0% to 3.7% (1.8% to 4.3%) of a standard deviation. These relations are likely to be causal given that we are comparing a bond to near-identical bonds on the same day and controlling for the bond's recent illiquidity.

6.2 The instrumental variables approach

In Section 6.1, we address potential concerns regarding the endogeneity of a bond's basket inclusion by controlling for the bond's recent illiquidity and using firm-time and maturity-time fixed effects. Nonetheless, readers may still wonder what the determinants of the remaining variation in basket inclusion are and whether they correlate with liquidity within the same firm and maturity bracket. One determinant, uncovered earlier in this paper, is the bond's over- or underweighting in the ETF's portfolio relative to the index. Recall from Section 5.4 that ETFs are more likely to include underweighted bonds in CR baskets and overweighted bonds in RD baskets. In this section, we use plausibly exogenous variation in this over- or underweighting to construct an instrument for basket inclusion.

We construct this instrument by recognizing, again, that fixed-income indexes rebalance monthly. Even though changes in bond characteristics that trigger index inclusions and exclusions happen throughout the month, index composition is not updated until the monthly rebalancing day. Any effects of changes in bond characteristics on bond liquidity should be incorporated in asset markets at the time when those characteristics are changing, not later when the index is rebalanced. Therefore, the fact that index changes are partially predictable by changes in bond characteristics does not violate the exclusion restriction. This is because shocks to bond over- or underweighting on index rebalancing days should affect ETF basket inclusion without being confounded by continuing changes in bond characteristics that influence bond liquidity throughout the month.

To instrument for the basket inclusion variables $Basket_{it}^{CR}$ and $Basket_{it}^{RD}$ in Equation (12), we construct the variables $CRInstr_{it}$ and $RDInstr_{it}$, respectively, as follows:

$$CRInstr_{it} \equiv \sum_{j \in J_t^{CR}} \Delta Deviation_{ijh} \tag{13}$$

$$RDInstr_{it} \equiv \sum_{j \in J_t^{RD}} \Delta Deviation_{ijh}\,, \tag{14}$$

where $\Delta Deviation_{ijh}$, which is defined in Equation (9), is a shock to bond i's overweighting in ETF j's portfolio induced by the rebalancing of the ETF's index on day h. For any day t following day h, we compute bond i's instrument $CRInstr_{it}$ by adding up the bond's overweighting shocks over the set of all ETFs that have CR baskets, J_t^{CR}. Similarly, we compute $RDInstr_{it}$ by summing the same shocks over the set J_t^{RD} of ETFs with RD baskets.[25] As before, we use only the first 10 baskets per ETF after each rebalancing.

Intuitively, bonds with high values of $CRInstr_{it}$ or $RDInstr_{it}$ have become more overweighted in ETF portfolios as a result of index rebalancing. Given our results from Section 5.4, we expect such bonds to be disproportionately included in RD baskets and excluded from CR baskets. We expect $CRInstr_{it}$ to better capture exclusion from CR baskets because the set J_t^{CR} over which $CRInstr_{it}$ is computed contains ETFs that have CR baskets but need not have RD baskets. Similarly, we expect $RDInstr_{it}$ to better capture inclusion in RD baskets because the set J_t^{RD} contains ETFs that have RD baskets. We include both instruments in each first-stage specification to ensure the consistency of our estimates.

As in the ordinary least squares (OLS) setting, our sample for this analysis consists of observations at the bond-day level. The first-stage specifications in our 2SLS estimation are

$$Basket_{it}^{RD} = \beta_1^{RD}\, RDInstr_{it} + \beta_2^{RD}\, CRInstr_{it} + Controls_{it} + \omega_{Ft} + \omega_{mt} + \epsilon_{it} \tag{15}$$

$$Basket_{it}^{CR} = \beta_1^{CR}\, RDInstr_{it} + \beta_2^{CR}\, CRInstr_{it} + Controls_{it} + \omega_{Ft} + \omega_{mt} + \epsilon_{it}\,. \tag{16}$$

[25]Note that both sets J_t^{CR} and J_t^{RD} can include only ETFs for which we have index composition data, because such data are needed to compute $\Delta Deviation_{ijh}$. Also note that an ETF does not have to hold bond i in order to be included in the set J_t^{CR} or J_t^{RD}; it only has to have a CR or RD basket on day t.

We expect to find $\beta_1^{RD} > 0$ and $\beta_2^{CR} < 0$, as explained in the previous paragraph.

The controls in Equations (15) and (16) include not only those from Equation (12) but also two additional controls designed to alleviate potential concerns about the exclusion restriction. In particular, one might worry that our instruments could be picking up a direct effect of index inclusion, whereby a bond becomes more liquid after being added to an index for reasons unrelated to ETFs, such as correlated trading of non-ETF investors. This concern should be alleviated by the fact that unlike other popular index-inclusion instruments, our instruments are ETF-specific: they use rebalancing-induced changes in bond overweighting for ETFs that have baskets on a given day. Thus, our instruments for basket inclusion operate on post-rebalancing days on which ETFs have baskets. In contrast, any broader effects of changes in index weights should not be limited to ETF-basket days.

To further isolate the effect of ETF basket inclusion, we add a direct control for general effects of bond index inclusion that do not operate through ETF baskets. Let $\mathbb{I}_{it}$ denote the set of indexes that include bond i and rebalance on day t. We define

$$\textit{Change in Index Weights}_{it} = \sum_{I \in \mathbb{I}_{it}} (\textit{Index Weight}_{i,I,t} - \textit{Index Weight}_{i,I,t-1}). \tag{17}$$

For each bond i on day t, we control for its most recent change in index weights.[26] Note that this control embeds information related to all indexes I that include the bond on a given day, not only indexes for which the tracking ETF has a CR or RD basket.

Our second additional control aims to capture changes in institutional investors' bond holdings that might be induced by bond index rebalancing. For each bond i on day t, we control for the most recent quarterly change in the number of shares of bond i held by institutional investors. Our institutional investor bond ownership data are quarterly and come from eMAXX.[27] Importantly, the institutions include mutual funds and insurance companies, which together own two-thirds of all U.S. investment-grade corporate bonds (Coppola 2021; Bretscher et al. 2021; Li and Yu 2023).

The first-stage results are reported in Table 7. As before, we consider two measures of bond i's basket inclusion on day t, $\textit{Basket}_{it}^{CR}$. First, we measure it by a dummy variable equal to one if the bond is included in at least one ETF's basket on that day. Second, we use the number of ETFs that include this bond in their CR baskets on that day. The results

[26] That is, for a given bond i on each day t following an index rebalancing that occurred at date $\tau_1 < t$ among at least one index in our sample that tracks the bond, we control for $\textit{Change in Index Weights}_{i\tau_1}$, until date τ_2, at which the subsequent index rebalancing occurs.

[27] We include all institutional investors reported in eMAXX. Our access to the ownership data ends in the third quarter of 2020. We impute fourth-quarter data by reusing the last available holdings.

for the first measure are in panel A of Table 7; those for the second measure are in panel B. We measure $Basket_{it}^{RD}$ analogously based on RD baskets. The results for $Basket_{it}^{RD}$ are in columns 1 to 3; those for $Basket_{it}^{CR}$ are in columns 4 to 6.

Table 7 shows that $\beta_1^{RD} > 0$ and $\beta_2^{CR} < 0$, as predicted. Both results hold for both basket inclusion measures and all three measures of liquidity, which is included among the controls. All of these estimates are significant at the 1% level.[28] We thus see that bonds overweighted after index-rebalancing shocks tend to be included in RD baskets but excluded from CR baskets. These bond-level results are not surprising given the ETF-bond-level results reported in Table 5, which we used to motivate our instruments. The cross-effects of $CRInstr_{it}$ on RD basket inclusion (β_2^{RD}) and $RDInstr_{it}$ on CR basket inclusion (β_1^{CR}) are weaker, as expected. As explained earlier, it is primarily overweighting at ETFs with CR (RD) baskets that affects a bond's CR (RD) basket inclusion.

Table 8 shows the second-stage results from the 2SLS specification of the effect of instrumented basket inclusion on bond illiquidity, which is the IV analog of the OLS specification in Equation (12). These results estimate the causal effect of basket inclusion on illiquidity. The slope estimates on instrumented basket inclusion are always negative. This is true in all 12 specifications: two measures of basket inclusion, two types of baskets (CR and RD), and three liquidity measures. Among the 12 negative estimates, four are significant at the 1% level, three at the 5% level, and three at the 10% level. We conclude that basket inclusion causes an improvement in bond liquidity.

This effect is significant not only statistically but also economically. For example, when a bond is included in one additional RD basket, its illiquidity decreases by 11.1% to 15.0% of a standard deviation. These magnitudes are substantial despite the presence of firm-time and maturity-time fixed effects, which absorb some of the variation of interest.

6.3 The role of basket imbalance

In Sections 6.1 and 6.2, we show that basket inclusion improves bond liquidity in our full sample. On most days in our sample, the shocks that give rise to buying or selling pressure in ETF shares are idiosyncratic, leading APs to alternate between CR and RD activity. Bonds used in CR or RD baskets thus tend to move back and forth between being bought and sold by the APs. Their frequent presence on both sides of AP inventory makes basket bonds more liquid, because the APs also tend to act as market makers in these bonds. However,

[28] In addition, all of the clustering-robust Kleibergen-Paap F-statistics from our first-stage regressions exceed 40, easily passing common tests for weak instruments. We report these statistics in Table 8.

when the shocks in the ETF market are asymmetric, the resulting CR-RD imbalances could potentially make basket bonds less liquid. In particular, systematic CR-RD imbalances generate imbalances in the market maker's inventory, with basket bonds concentrated on one side of the inventory, which could harm the bonds' liquidity. In this section, we examine the role of CR-RD imbalances in the basket-liquidity relation.

Panel A of Figure 4 plots the time series of the number of ETFs in our sample that create or redeem shares between January 2017 and December 2020. We see that ETFs engaged in creations generally outnumber those engaged in redemptions. On a typical day, about 20 ETFs create new shares, but only five or six ETFs redeem. Panel B shows that creations tend to outweigh redemptions also in dollar terms. However, net redemption volume exhibits substantial week-to-week variation, and there are about 40 weeks in which redemptions outweigh creations in dollar terms. Our sample period is thus characterized mostly by net creation, but with a fair amount of redemption occurring as well.

The only time when bond ETFs experienced large and persistent redemptions was during the COVID-19 crisis. In March 2020, we saw sharp increases in the number of ETFs engaged in redemption as well as in the volume of net redemptions. In the week of March 11 to 18 alone, total net redemptions by the ETFs in our sample reached $3.2 billion. This spike in redemptions is clearly visible in both panels of Figure 4.

To examine how COVID-induced redemptions affect the basket-liquidity relation, we estimate regression (12) over the "COVID subperiod" of March 2 to April 15, 2020. Starting the week of March 2, corporate bond spreads began to surge (Haddad, Moreira, and Muir 2021). March 2 also marks the beginning of systematic redemptions from corporate bond ETFs (Figure 4). Redemptions began to decline following the Federal Reserve's market interventions, and they largely returned to their pre-COVID levels by mid-April.[29] Our choice of April 15 as the end of the COVID subperiod is close to the dates chosen by others for similar purposes. For example, Haddad, Moreira, and Muir (2021) mark April 16 as the end of the recovery period in the corporate bond basis, while He, Nagel, and Song (2022) end their sample period on April 13.

Table 9 shows that the slopes from regression (12) estimated in the COVID subperiod are very different from those obtained in the full period. Recall that, in the full period, all 12 estimates of the slopes on basket inclusion are negative and significant (Table 6). In contrast,

[29] The key interventions in the corporate bond market took place on March 23, 2020, when the Federal Reserve announced that it would buy investment-grade corporate bonds, both through ETFs and directly, and on April 9, 2020, when it announced purchases of non-investment-grade bond ETFs. These interventions appear to have stemmed the ongoing decline in corporate bond liquidity (Kargar et al. 2021; O'Hara and Zhou 2021; Boyarchenko, Kovner, and Shachar 2022; Hempel, Kim, and Wermers 2022).

nine of the 12 estimates are positive in the COVID subperiod. The results for RD baskets are especially different. The magnitudes of the estimated slopes on RD basket inclusion are similar to those in Table 6, but the signs are opposite. All six of these RD basket slopes are positive, and five of them are statistically significant (four at the 1% level, one at the 10% level). We thus see that inclusion in RD baskets decreases rather than increases bond-level liquidity during the COVID subperiod.

In this exercise, we cannot follow our instrumental variables approach due to the short length of the COVID subperiod, which includes only one index rebalancing date. Moreover, several index providers postponed their March 2020 rebalancing due to extreme market conditions. However, the results in Table 9 are based on specifications that include firm-day and maturity-day fixed effects and also control for recent bond-level illiquidity, all of which alleviate endogeneity concerns. Moreover, the conclusions from Table 9 are robust to the inclusion of additional controls and fixed effects (see the Internet Appendix).

The COVID subperiod provides preliminary evidence that inclusion in RD baskets can hurt a bond's liquidity when redemptions are systematic and persistent. In this subperiod, many investors experienced liquidity shocks that led them to sell ETF shares. This selling pressure was met by APs who purchased many ETF shares from investors, redeemed them, and then tried to sell the bonds acquired through RD baskets. Bonds heavily represented in RD baskets thus became heavily represented in APs' inventory. Given their balance sheet constraints, APs became reluctant to purchase even more of the same bonds in their role as market makers. Bonds present in RD baskets thus lost their most natural buyers. When its own market makers do not want to buy it, a security can become quite illiquid.

We now explore the role of basket imbalance more generally. To measure basket imbalance for a given bond, we first let N_{it}^{CR} and N_{it}^{RD} denote the numbers of CR baskets and RD baskets, respectively, in which bond i appears any time during the week immediately preceding day t. We then define two basket imbalance variables:

$$Imbal_{it}^{CR} \equiv \max\left(N_{it}^{CR} - N_{it}^{RD}, 0\right) \tag{18}$$

$$Imbal_{it}^{RD} \equiv \max\left(N_{it}^{RD} - N_{it}^{CR}, 0\right) . \tag{19}$$

That is, when a bond is included in more CR baskets than RD baskets (i.e., $N_{it}^{CR} > N_{it}^{RD}$), we set $Imbal_{it}^{CR}$ equal to the positive difference and $Imbal_{it}^{RD}$ to zero. And vice versa. We also let $Basket_{it}$ denote the total number of baskets, CR or RD, that contain bond i in the week preceding day t. We then estimate a modified version of regression (12):

$$Illiquidity_{it+1} = \beta_1\, Basket_{it} + \beta_2^{CR}\, Basket_{it} \times Imbal_{it}^{CR} + \beta_2^{RD}\, Basket_{it} \times Imbal_{it}^{RD} + e_{it}\,, \tag{20}$$

where $e_{it} \equiv Controls_{it} + \omega_{Ft} + \omega_{mt} + \epsilon_{it}$. As before, the controls include the bond's average illiquidity over the prior week and average basket size, and we also include a firm-day fixed effect, ω_{Ft}, and a maturity-day fixed effect, ω_{mt}.

Table 10 reports the regression estimates. We find that $\beta_1 < 0$, indicating that a bond's liquidity improves after the bond is included in a larger number of baskets. This result is similar to that observed in Table 6 for CR and RD baskets separately. However, we also find that $\beta_2^{CR} > 0$ and $\beta_2^{RD} > 0$, both highly significant, indicating that basket imbalance of either kind, CR or RD, weakens the favorable effect of basket inclusion on bond liquidity. This effect can become unfavorable if basket imbalance is sufficiently large.

To get a sense of the magnitudes, consider a bond included in five CR baskets and five RD baskets (i.e., $Imbal_{it}^{RD} = 0$). If this bond were instead included in 10 CR baskets and no RD baskets (i.e., $Imbal_{it}^{CR} = 10$), the bond's liquidity would be lower by 15.5% to 18.0% of a standard deviation. Alternatively, if this bond were included in 10 RD baskets and no CR baskets (i.e., $Imbal_{it}^{RD} = 10$), its liquidity would be lower by 14.3% to 30.5% of a standard deviation. To put these numbers into context, at the height of the COVID-19 crisis, 32 ETFs experienced redemptions, while only five ETFs had creations. The effect of basket imbalance on the basket-liquidity relation can therefore be substantial.

Overall, we find that basket inclusion need not always improve a bond's liquidity. When there are systematic creations or redemptions across many ETFs, basket bonds can become less liquid as a result of being held one-sidedly by their market makers. This scenario, in the form of systematic redemptions, materialized during the COVID-19 crisis.

In similar spirit, Saglam, Tuzun, and Wermers (2021) find ETFs to be liquidity-enhancing in normal times but liquidity-reducing in a period of market stress. Their work differs from ours in three major ways. First, they examine the effects of ETF ownership, whereas we study the effects of basket inclusion. Baskets include only a subset of ETF holdings, as we show earlier. Second, they consider the 2011 U.S. debt-ceiling crisis as the period of market stress, whereas we analyze the COVID-19 crisis and basket imbalance more generally. Finally, they examine stock ETFs while we study bond ETFs. For bond ETFs, the APs are more likely to act as market makers in the underlying securities. This market-making role is also emphasized by Hempel, Kim, and Wermers (2022), who argue that the Federal Reserve's purchases of corporate bond ETFs in 2020 improved bond liquidity by alleviating the APs' balance sheet constraints.

6.4 Robustness

In this subsection, we show that our results in Section 6 are robust to a number of alternative specifications. We summarize our findings here and show the details in the Internet Appendix.

Sections 6.1 through 6.3 show that our results on the effect of basket inclusion on bond liquidity hold for two measures of basket inclusion and three measures of liquidity, and they survive the inclusion of multiple fixed effects and controls. The results also hold for two measures of basket inclusion based on the number of bond shares included in ETF baskets: the log of one plus the number of shares of this bond that are included in ETFs' baskets on this day, and the number of shares of this bond included across all ETF baskets divided by the overall number of bond shares included in all ETF baskets on this day. The results also survive controls for credit rating, amount outstanding, and the level of institutional holdings. In fact, the results survive the addition of a bond fixed effect, which isolates the time variation in a given bond's basket inclusion. The results also survive a control for ETF ownership, which is explored in prior studies and correlated with basket inclusion. We measure a bond's ETF ownership by the fraction of the bond's shares outstanding that are held by ETFs. We find that ETF ownership is positively related to future bond liquidity in normal times and negatively in the COVID subperiod, but adding this control does not affect any of our conclusions regarding the effects of basket inclusion.

Our measures of bond liquidity are sometimes missing because not all bonds trade every day. However, we find very similar results when we repeat our tests using bond liquidity averaged over the first three trading days with observed bond trades after basket inclusion. We further control for the number of zero trading days—that is, the number of days that a bond has gone without trading over the past month. Again, all of our results remain, confirming that our findings are not driven by infrequent trading of a subset of bonds.

Finally, we consider a simpler measure of basket imbalance:

$$Imbal_{it} \equiv |N_{it}^{CR} - N_{it}^{RD}|, \tag{21}$$

that is, the absolute value of the difference between the numbers of CR and RD baskets in which the bond appears. We then run the regression

$$Illiquidity_{it+1} = \beta_1\, Basket_{it} + \beta_2\, Basket_{it} \times Imbal_{it} + Controls_{it} + \omega_{Ft} + \omega_{mt} + \epsilon_{it}\,. \tag{22}$$

The results mirror those in Table 10: $\beta_1 < 0$ and $\beta_2 > 0$, both significant at the 1% level. Overall, we find robust evidence that a bond's inclusion in ETF baskets improves the bond's liquidity, and that this effect is attenuated in times of high CR/RD imbalance.

7 Conclusion

We show that passive ETFs actively manage their portfolios, balancing index tracking against liquidity transformation. ETFs update their baskets frequently to steer their portfolios toward the index. For example, when an ETF's portfolio underweights a security relative to the index, the ETF tends to add this security to the CR basket and remove it from the RD basket. These basket adjustments are attenuated for ETFs holding less liquid securities because ETFs also aim to incentivize arbitrage activity. Arbitrageurs boost the liquidity of ETF shares by absorbing other investors' trades, mitigating their price impact. To help reduce the arbitrageurs' transaction costs, ETFs adjust their baskets by including cash and excluding some of the index securities. The use of cash and basket concentration is more common when the index securities are less liquid, and it comes at the expense of index tracking. We capture the trade-off between index tracking and liquidity transformation in a theoretical model. While we model ETFs in general, our evidence is based on corporate bond ETFs, for which liquidity transformation is particularly important.

To offer an analogy, imagine steering a ship powered by a team of rowers to a harbor on the other side of the lake. If the lake water were clear, you would chart a straight path and track it closely by quickly correcting all random deviations. However, the lake is full of mud. Rowing through mud is more exhausting. How do you adjust the ship's path in light of the mud? Instead of always tracking the straight path, you may want to avoid the muddiest parts. By steering clear of mud, you make it easier for the rowers to move the ship forward, which helps you get to the harbor more effectively. An astute reader already knows that the ship is a metaphor for an ETF portfolio, you are the ETF's manager, and the rowers symbolize the arbitrageurs. Clear water represents liquidity and mud illiquidity.

ETFs' efforts to improve the liquidity of their shares have consequences for the liquidity of the underlying securities. We find that a bond's inclusion in an ETF basket has a significant state-dependent effect on the bond's liquidity. This effect is positive in normal times but negative in periods of large imbalance between creations and redemptions. A salient example occurred in the spring of 2020. The COVID-19 crisis witnessed acute selling pressure in the bond market, which led to net redemptions from bond ETFs, which in turn strained the liquidity of the bonds concentrated in RD baskets. Given the growing role of ETFs in liquidity transformation, future episodes of ETF-induced liquidity strains seem likely.

Future research can examine additional consequences of ETFs' active basket management. One promising step in this direction is Reilly (2022), which studies the performance of bonds included in CR baskets. It would also be useful to analyze the security-level determinants of basket inclusion. ETF baskets remain severely understudied.

Similar to ETFs, index mutual funds also trade off index tracking against liquidity transformation. While mutual funds and ETFs face the same friction—the illiquidity of their portfolio securities—they differ in how they overcome it. To reduce trading costs, mutual funds choose which securities to trade in secondary markets, whereas ETFs negotiate the composition of their CR/RD baskets for in-kind exchanges with APs. In addition, mutual funds hold cash in their portfolios as a liquid buffer to meet investor redemptions, whereas ETFs use cash in their baskets to incentivize AP trading. Documenting how ETFs respond to illiquidity by actively managing their baskets, including their cash component, is a key contribution of our paper. ETFs are well suited for this analysis because their portfolio holdings and baskets are available on a daily basis, whereas mutual fund holdings are only quarterly. Nevertheless, it would be useful to investigate liquidity-motivated active portfolio management also for index mutual funds. We leave such analysis for future work.

Code Availability: The replication code is available in the Harvard Dataverse at https://doi.org/10.7910/DVN/RIFTUJ.

References

Anand, A., C. Jotikasthira, and K. Venkataraman. 2021. Mutual fund trading style and bond market fragility. *Review of Financial Studies* 34:2993–3044.

Agarwal, V., P. Hanouna, R. Moussawi, and C. W. Stahel. 2018. Do ETFs increase the commonality in liquidity of underlying stocks? Working Paper.

Akey, P., A. Z. Robertson, and M. Simutin. 2021. Closet active management of passive funds. Working Paper.

Ben-David, I., F. Franzoni, and R. Moussawi. 2017. Exchange-traded funds. *Annual Review of Financial Economics* 9:169–89.

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

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

Bessembinder, H., S. Jacobsen, W. Maxwell, and K. Venkataraman. 2018. Capital commitment and illiquidity in corporate bonds. *Journal of Finance* 73:1615–61.

Box, T., R. Davis, R. Evans, and A. Lynch. 2021. Intraday arbitrage between ETFs and their underlying portfolio. *Journal of Financial Economics* 141:1078–95.

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

Bretscher, L., L. Schmid, I. Sen, and V. Sharma. 2021. Institutional corporate bond pricing. Working Paper.

Brogaard, J., D. Heath, and D. Huang. 2024. ETF sampling and index arbitrage. Working Paper.

Chen, Q., I. Goldstein, and W. Jiang. 2010. Payoff complementarities and financial fragility: Evidence from mutual fund outflows. *Journal of Financial Economics* 97:239–62.

Cheng, S., M. Massa, and H. Zhang. 2019. The unexpected activeness of passive investors: a worldwide analysis of ETFs. *Review of Asset Pricing Studies* 9:296–355.

Chernenko, S., and A. Sunderam. 2017. Liquidity transformation in asset management: Evidence from the cash holdings of mutual funds. Working Paper.

Choi, J., S. Hoseinzade, S. S. Shin, and H. Tehranian. 2020. Corporate bond mutual funds and asset fire sales. *Journal of Financial Economics* 138:432–57.

Cohn, J. B., Z. Liu, and M. I. Wardlaw. 2022. Count (and count-like) data in finance. *Journal of Financial Economics* 146:529–51.

Cong, L. W., S. Huang, and D. Xu. 2024. The rise of factor investing: "Passive" security design and market implications. Working Paper.

Coppola, A. 2021. In safe hands: The financial and real impact of investor composition over the credit cycle. Working Paper.

Cremers, M., and A. Petajisto. 2009. How active is your fund manager? A new measure that predicts performance. *Review of Financial Studies* 22:3329–65.

Da, Z., and S. Shive. 2018. Exchange traded funds and asset return correlations. *European Financial Management* 24:136–68.

Dannhauser, C. 2017. The impact of innovation: Evidence from corporate bond ETFs. *Journal of Financial Economics* 125:182–99.

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

Dannhauser, C., and M. Dathan. 2023. Passive investors in primary bond markets. Working Paper.

Dannhauser, C., and E. Karmaziene. 2023. The dealer warehouse: Corporate bond ETFs. Working Paper.

Diamond, D., and P. Dybvig. 1983. Bank runs, deposit insurance, and liquidity. *Journal of Political Economy* 91:401–19.

Dick-Nielsen, J. 2014. How to clean enhanced TRACE data. Working Paper.

Du, L. 2023. Inventory in liquidity transformation: Evidence from corporate bond ETF creations. Working Paper.

Easley, David., D. Michayluk, M. O'Hara, and T. Putnins. Forthcoming. The active world of passive investing. *Review of Finance*.

Falato, A., I. Goldstein, and A. Hortacsu. 2021. Financial fragility in the COVID-19 crisis: The case of investment funds in corporate bond markets. *Journal of Monetary Economics* 123:35–52.

Feldhutter, P. 2012. The same bond at different prices: Identifying search frictions and selling pressures. *Review of Financial Studies* 25:1155–206.

Finnerty, J., N. Reisel, and X. Zhong. Forthcoming. ETFs, creation and redemption processes, and bond liquidity. *Journal of Financial and Quantitative Analysis*.

Fremault, A. 1991. Stock index futures and index arbitrage in a rational expectations model. *Journal of Business* 64:523–47.

Garleanu, N., and L. Pedersen. 2013. Dynamic trading with predictable returns and transaction costs. *Journal of Finance* 68:2309–40.

Goldstein, I., H. Jiang, and D. T. Ng. 2017. Investor flows and fragility in corporate bond funds. *Journal of Financial Economics* 126:592–613.

Gorbatikov, E., and T. Sikorskaya. 2022. Two APs are better than one: ETF mispricing and primary market participation. Working Paper.

Goyenko, R. Y., C. W. Holden, and C. A. Trzcinka. 2009. Do liquidity measures measure liquidity. *Journal of Financial Economics* 92:153–81.

Haddad, V., A. Moirera, and T. Muir. 2021. When selling becomes viral: Disruptions in debt markets in the Covid-19 crisis and the Fed's response. *Review of Financial Studies* 34:5309–51.

He, Z., P. Khorrami, and Z. Song. 2022. Commonality in credit spread changes: dealer inventory and intermediary distress. *Review of Financial Studies* 35:4630—73.

He, Z., S. Nagel, and Z. Song. 2022. Treasury inconvenience yields during the COVID-19 crisis. *Journal of Financial Economics* 143:57–79.

Helmke, A. 2023. Will ETFs drive mutual funds extinct? Working Paper.

Hempel, S., D. Kim, and R. Wermers. 2022. Financial intermediary funding constraints and segmented markets: Evidence from SMCCF ETF purchases. Working Paper.

Hendershott, T., D. Li, D. Livdan, N. Schürhoff, and K. Venkataraman. 2021. Quote competition in corporate bonds. Working Paper.

Holden, C. W. 2009. New low-frequency spread measures. *Journal of Financial Markets* 12:778–813.

Holden, C. W., and J. Nam. Forthcoming. Market accessibility, corporate bond ETFs, and liquidity. *Review of Finance*.

Israeli, D., C. Lee, and S. Sridharan. 2017. Is there a dark side to exchange traded funds? An information perspective. *Review of Accounting Studies* 22:1048–83.

Jin, D., M. Kacperczyk, B. Kahraman, and F. Suntheim. 2022. Swing pricing and fragility in open-end mutual funds. *Review of Financial Studies* 35:1–50.

Kacperczyk, M., C. Sialm, and L. Zheng. 2008. Unobserved actions of mutual funds. *Review of Financial Studies* 21:2379–416.

Kacperczyk, M., and P. Schnabl. 2013. How safe are money market funds? *Quarterly Journal of Economics* 128:1073–122.

Kargar, M., B. Lester, D. Lindsay, S. Liu, P. Weill, and D. Zuniga. 2021. Corporate bond liquidity during the COVID-19 crisis. *Review of Financial Studies* 34:5352–401.

Kumar, P., and D. Seppi. 1994. Information and index arbitrage. *Journal of Business* 67:481–509.

Lettau, M., and A. Madhavan. 2018. Exchange-traded funds 101 for economists. *Journal of Economic Perspectives* 32:135–54.

Li, S. 2021. Should passive investors actively manage their trades? Working Paper.

Li, J., and H. Yu. 2023. Investor composition and the liquidity component in the U.S. corporate bond market. Working Paper.

Ma, Y., K. Xiao, and Y. Zeng. 2023. Bank debt, mutual fund equity, and swing pricing in liquidity provision. Working Paper.

Ma, Y., Y. Zeng, and A. L. Zhang. 2024. Stablecoin runs and the centralization of arbitrage. Working Paper.

Marta, T. 2024. Corporate bond ETFs, bond liquidity, and ETF trading volume. Working Paper.

O'Hara, M., and Xing Z. 2021. Anatomy of a liquidity crisis: Corporate bonds in the COVID-19 crisis. *Journal of Financial Economics* 142:46–68.

Pan, K., and Y. Zeng. 2019. ETF arbitrage under liquidity mismatch. Working Paper.

Pastor, L., R. F. Stambaugh, and L. A. Taylor. 2020. Fund tradeoffs. *Journal of Financial Economics* 138:614–34.

Petajisto, A. 2017. Inefficiencies in the pricing of exchange-traded funds. *Financial Analysts Journal* 73:24–54.

Pu, X. 2009. Liquidity commonality across the bond and CDS markets. *Journal of Fixed Income* 19:26–39.

Reilly, C. 2022. The hidden cost of corporate bond ETFs. Working Paper.

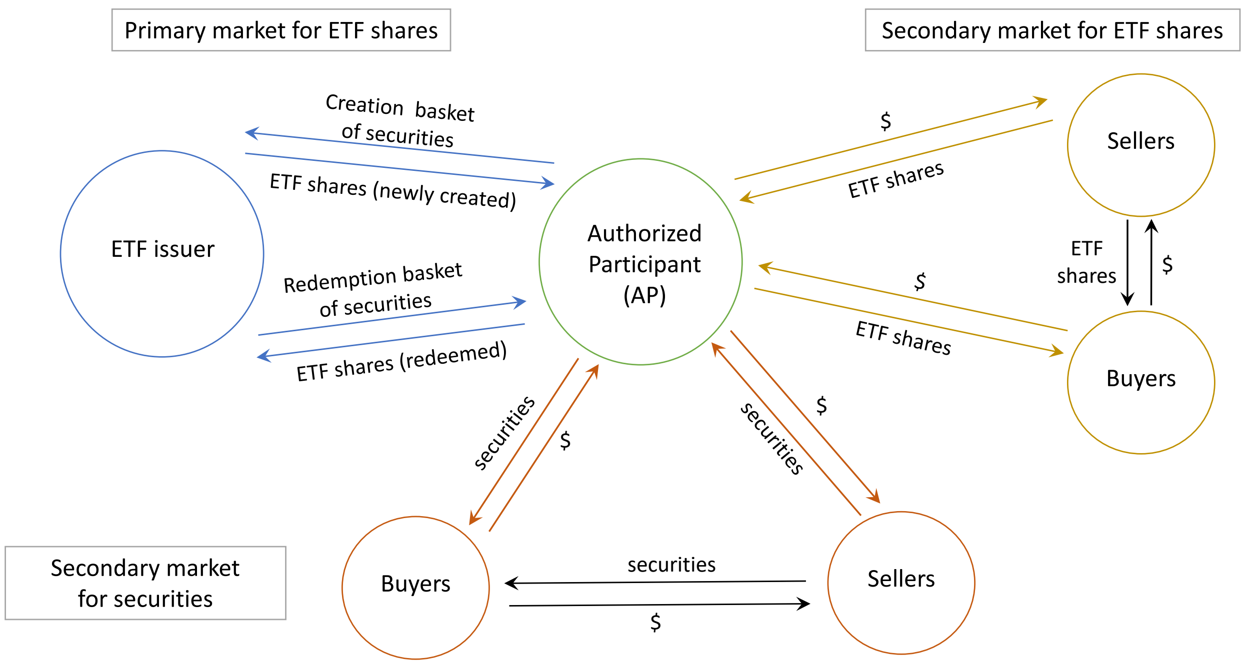
Saglam, M., T. Tuzun, and R Wermers. 2021. Do ETFs increase liquidity? Working Paper.

Shim, J. J., and K. Todorov. 2023. ETFs, illiquid assets, and fire sales. Working Paper.

Song, H., and H. Zhou. 2007. Nondefault bond spread and market trading liquidity. Working Paper.

Sunderam, A. 2015. Money creation and the shadow banking system. *Review of Financial Studies* 28:939–77.

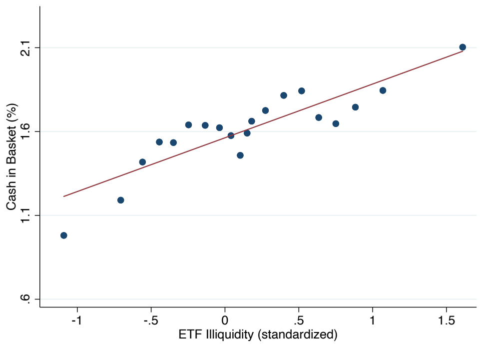
Figure 1
ETF arbitrage



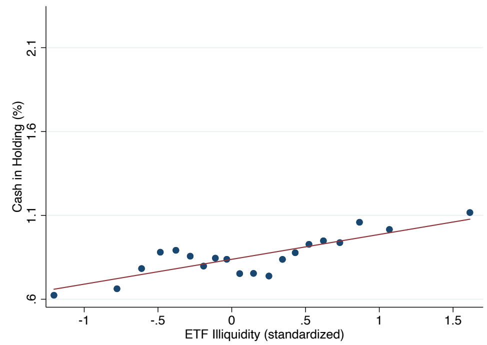
This figure provides a visual representation of the ETF arbitrage mechanism.
Alt Text: Graphical representation of how the ETF arbitrage process connects three markets: the primary and secondary markets for ETF shares and the secondary market for the underlying securities. Depicts transfers of securities, ETF shares, and cash during the arbitrage process among the ETF issuer, authorized participant, and buyers and sellers of ETF shares and securities.

Figure 2
Cash and liquidity

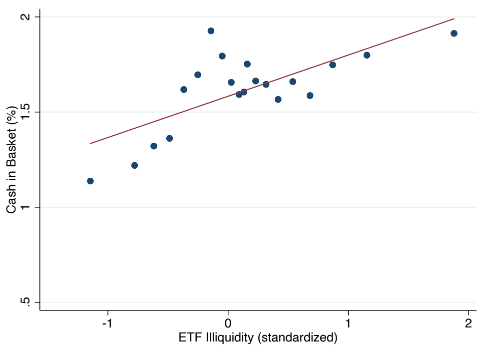
(a) Basket cash and illiquidity IL1



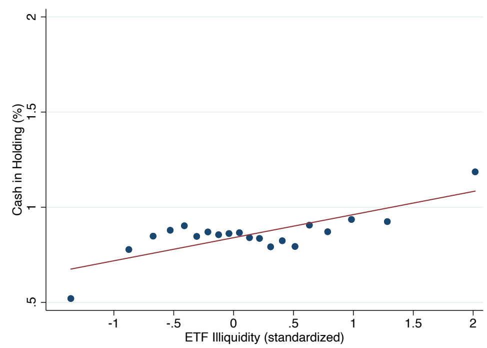
(b) Holdings cash and illiquidity IL1



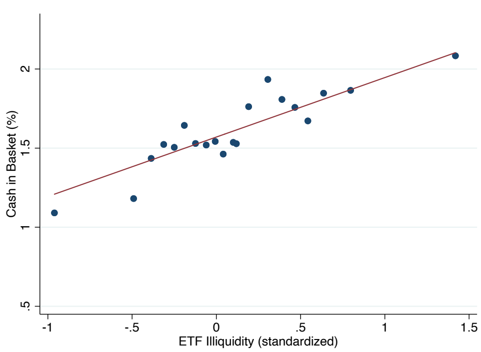
(c) Basket cash and illiquidity IL2



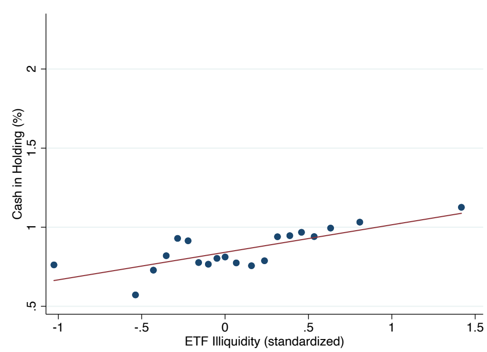
(d) Holdings cash and illiquidity IL2



(e) Basket cash and illiquidity IL3



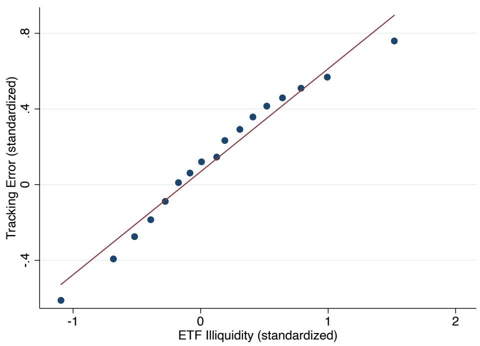
(f) Holdings cash and illiquidity IL3



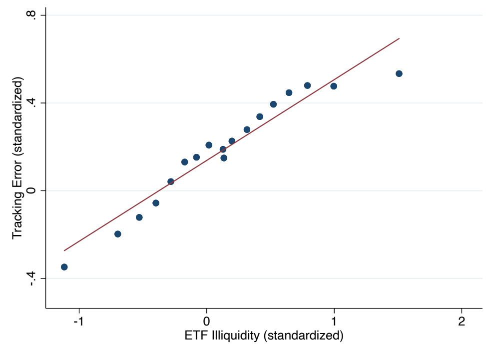
This figure shows binned scatterplots of the proportion of cash in reported ETF baskets (panels A, C, and E) and ETF portfolio holdings (panels B, D, and F) against the average illiquidity of the non-cash securities in the ETF's portfolio. The three measures of illiquidity are the effective tick size (IL1), imputed roundtrip cost (IL2), and interquartile range (IL3). Each panel also plots the line of best fit from a linear regression with issuer-time fixed effects and a control variable for the ETF's bid-ask spread in the secondary market.
Alt Text: Six binned scatterplots and regression lines depicting the relationship between the proportion of cash in ETF baskets or ETF portfolios, and ETF portfolio illiquidity, for three measures of illiquidity.

Figure 3
Tracking error and liquidity

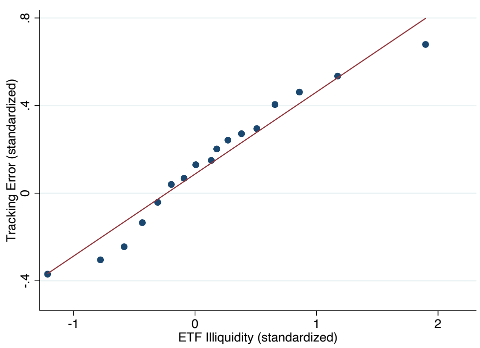
(a) Tracking error 1 and illiquidity IL1



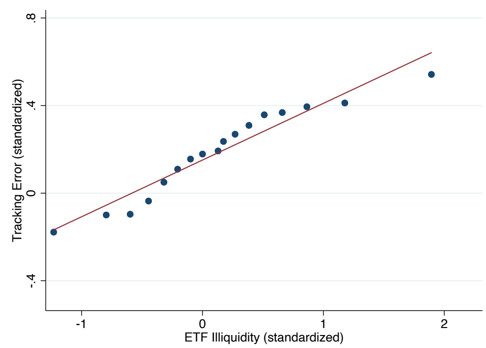
(b) Tracking error 2 and illiquidity IL1



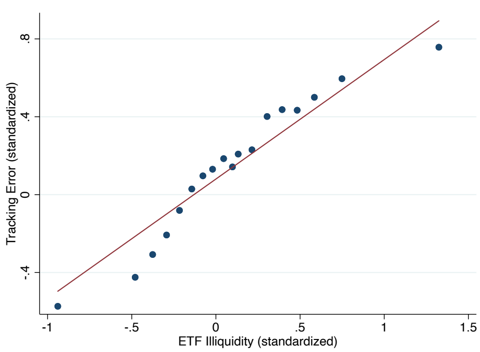
(c) Tracking error 1 and illiquidity IL2



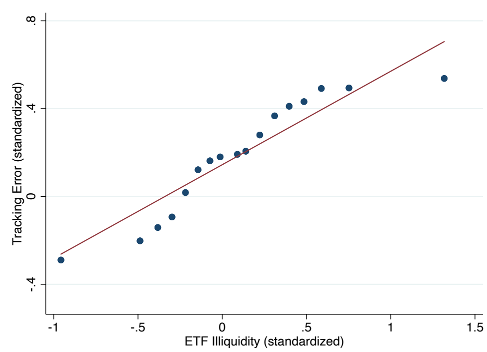
(d) Tracking error 2 and illiquidity IL2



(e) Tracking error 1 and illiquidity IL3



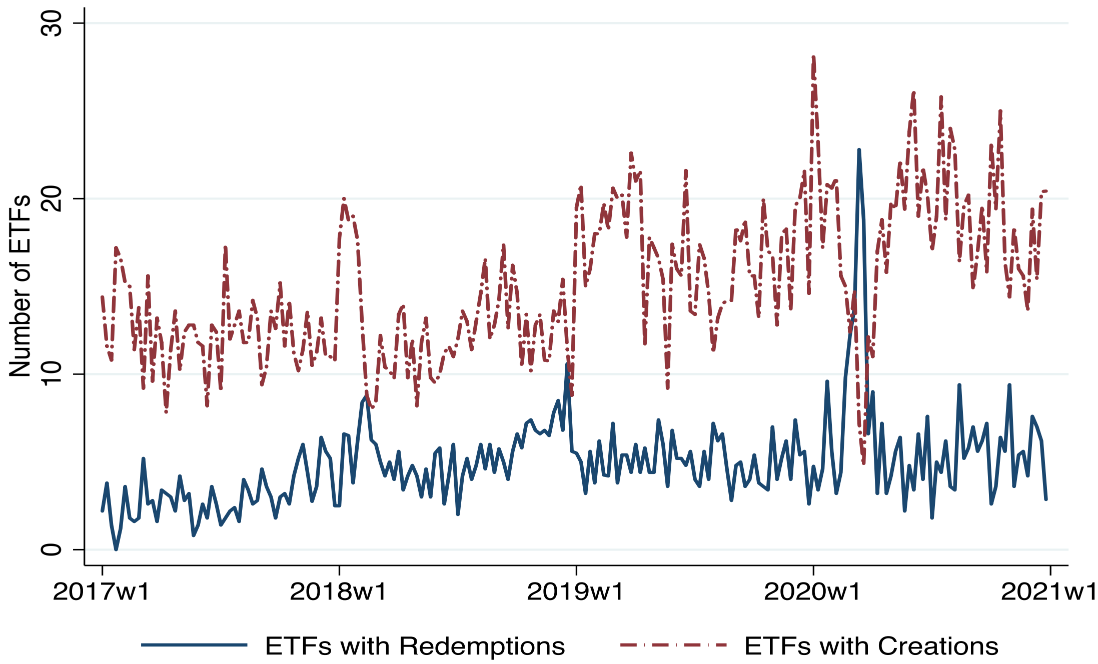
(f) Tracking error 2 and illiquidity IL3



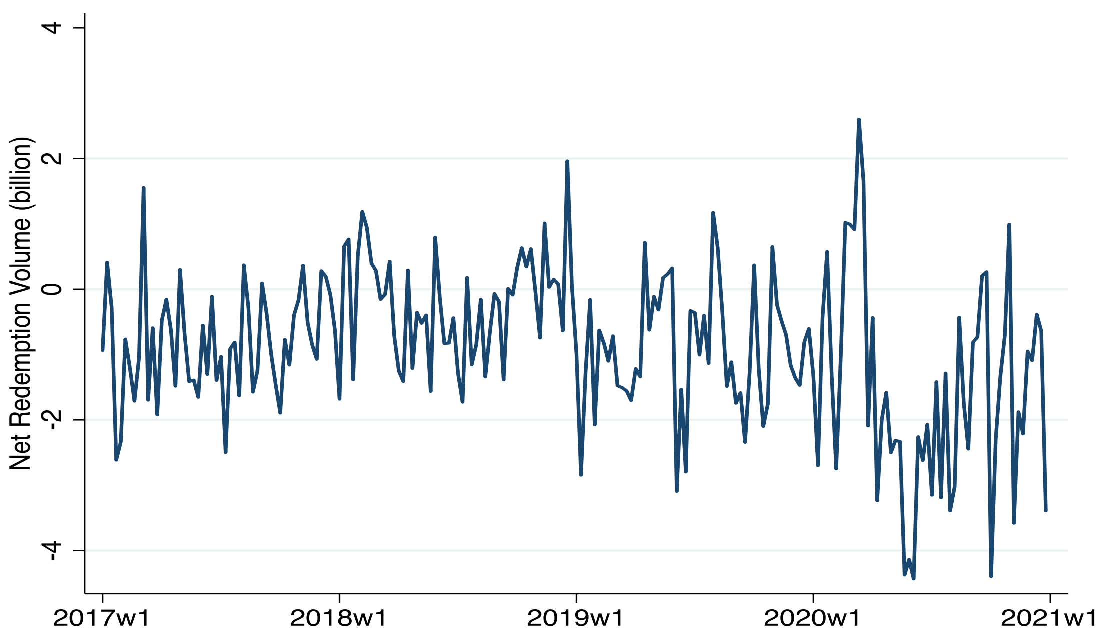
This figure shows binned scatterplots of ETF tracking error against the average illiquidity of the non-cash securities in the ETF's portfolio. The three measures of illiquidity are the effective tick size (IL1), imputed roundtrip cost (IL2), and interquartile range (IL3). Both measures of tracking error are monthly standard deviations of daily differences between ETF returns and index returns. For tracking error 1, ETF returns are from CRSP ETF price returns, including dividends, and index returns come from Bloomberg. For tracking error 2, ETF returns are from CRSP ETF NAV returns, including dividends, and index returns come from Bloomberg. Each panel also plots the line of best fit from a linear regression with issuer-time fixed effects and a control variable for the ETF's bid-ask spread in the secondary market.
Alt Text: Six binned scatterplots and regression lines depicting the relationship between tracking error and ETF portfolio illiquidity, using two constructions of tracking error and three measures of illiquidity.

Figure 4
Creations and redemptions over time

(a) Number of funds that redeem and create (weekly average)



(b) Net redemption volume (weekly sum)



Panel A plots the number of corporate bond ETFs that create or redeem shares on a given day, averaged weekly. Panel B plots the weekly volume of net redemptions (i.e., redemptions minus creations) by corporate bond ETFs, in billions of dollars.
Alt Text: Two line charts depicting how the amount of ETF creation and redemption varies over time.

Table 1
Summary statistics

This table shows summary statistics for the full ETF-level sample (panel A), the ETF-level sample for which index data is available (panel B), the number of securities adjusted on ETF-days with and without imputed baskets (panel C), the ratio of ETFs with creations and redemptions for each day in our sample (panel D), and summary statistics for the bond-level sample (panel E). In panel C, we exclude days immediately preceding and following CR and RD days to avoid mistakenly counting CR/RD activity due to potential shifts in reporting or trading. The bond-level sample includes all bonds that appear in ETF portfolios.

A. ETF level (full sample)

	Mean	Std. dev.	p25	p50	p75
AUM (million $)	2,127	5,547	77	340	1,346
ETF days	748	251	547	823	969
Bonds in portfolio	753	1,008	186	323	1,004
Bonds in reported basket	424	485	121	223	556
Bonds in realized RD basket	147	137	52	97	205
Bonds in realized CR basket	104	90	51	78	119
Proportion of RD days	0.058	0.096	0.003	0.012	0.058
Proportion of CR days	0.169	0.169	0.022	0.107	0.274
Proportion of no change days	0.773	0.240	0.679	0.847	0.966
ETF bid-ask spread	0.089	0.072	0.042	0.072	0.111
Observations	118				

B. ETF level (index-merged sample)

	Mean	Std. dev.	p25	p50	p75
AUM (million $)	2,702	5,645	118	576	1,723
ETF days	585	289	364	575	790
Bonds in index	1,153	1,217	297	536	1,847
Bonds in portfolio	835	849	256	414	1,126
Bonds in reported basket	696	683	170	523	924
Bonds in realized RD basket	138	123	54	91	242
Bonds in realized CR basket	128	96	74	97	172
Proportion of RD days	0.078	0.112	0.006	0.016	0.139
Proportion of CR days	0.208	0.190	0.044	0.155	0.342
Proportion of no change days	0.715	0.269	0.635	0.806	0.931
ETF bid-ask spread	0.073	0.052	0.035	0.057	0.102
Observations	57				

C. ETF-day level

	Mean	Std. dev.	p10	p25	p50	p75	p90
Number of securities adjusted: no basket	3	13	0	0	1	2	4
Number of securities adjusted: basket	188	282	6	41	89	221	478
Observations	19,704						

D. Day level

	Mean	Std. dev.	p25	p50	p75
Ratio of ETFs with redemption	7.65	5.12	4.26	6.98	10.34
Ratio of ETFs with creations	19.21	7.57	14.16	18.52	23.86

E. Bond level

	Mean	Std. dev.	p25	p50	p75
Days held by ETFs	615	383	281	572	987
Days in RD baskets	57	93	1	14	72
Days in CR baskets	92	121	4	36	141
RD basket share	0.208	0.130	0.114	0.188	0.282
CR basket share	0.181	0.120	0.096	0.148	0.244
Effective tick	0.003	0.001	0.002	0.002	0.003
IRC	0.193	0.148	0.095	0.149	0.247
IQR	0.003	0.002	0.002	0.003	0.004
Observations	18,746				

Table 2
Cash in ETF baskets and holdings

This table shows the cross-sectional distributions of cash ratios in ETF baskets and portfolios. Cash ratios are computed by dividing the amount of cash in the basket (portfolio) by total basket (portfolio) value on the same day. We first calculate the time-series average of a cash ratio at the ETF level and then report two averages and the 10th, 25th, 50th, 75th, and 90th percentiles of its cross-sectional distribution. The "unweighted" average is equal-weighted; the "weighted" average is weighted by assets under management. We report the distributions for realized RD baskets, realized CR baskets, reported baskets, and ETF portfolio holdings. The distributions of reported cash ratios are computed across three sets of days: days with CR activity, days with RD activity, and all days, which include days with CR or RD activity as well as no activity.

	Average		Distribution				
	Unweighted	Weighted	p10	p25	p50	p75	p90
Realized creation baskets	11.60	7.27	−0.44	0.28	6.25	18.96	34.26
Realized redemption baskets	8.18	3.68	−2.05	−0.00	0.73	6.43	35.19
Reported baskets (All days)	5.39	2.03	0.19	0.54	1.09	2.27	9.28
Reported baskets (CR days)	4.58	1.89	0.11	0.34	0.97	2.23	5.94
Reported baskets (RD days)	7.78	2.45	0.01	0.24	0.80	2.69	10.89
Portfolio holdings	1.70	0.85	0.00	0.44	0.83	1.57	2.47

Table 3
Concentration of ETF baskets and holdings

This table shows the cross-sectional distributions of concentration ratios for ETF baskets and portfolios. The concentration ratios for ETF baskets are computed by dividing the number of bonds in the basket by the number of bonds in the underlying index on the same day. The concentration ratios for ETF portfolios are computed by dividing the number of bonds in the ETF portfolio by the number of bonds in the index. We first calculate the time-series average of the given ratio at the ETF level and then report two averages and the 10th, 25th, 50th, 75th, and 90th percentiles of its cross-sectional distribution. The "unweighted" average is equal-weighted; the "weighted" average is weighted by assets under management. We report the distributions for realized RD baskets, realized CR baskets, reported baskets, and ETF portfolio holdings. The distributions of reported basket ratios are computed across three sets of days: days with CR activity, days with RD activity, and all days, which include days with CR or RD activity as well as no activity.

	Average		Distribution				
	Unweighted	Weighted	p10	p25	p50	p75	p90
Realized creation baskets	24.56	18.13	4.22	10.21	19.44	37.44	52.75
Realized redemption baskets	29.80	18.00	3.77	9.64	22.01	38.72	87.12
Reported baskets (All days)	76.06	57.51	44.65	64.28	84.09	94.88	97.35
Reported baskets (CR days)	77.56	59.53	45.15	63.98	85.42	95.13	97.56
Reported baskets (RD days)	78.40	56.22	45.73	69.80	86.75	96.72	100.00
Portfolio holdings	81.28	91.39	45.64	64.76	88.03	96.70	102.39

Table 4
Active cash management

This table reports the slope estimates from the regressions of the proportion of cash in a CR or RD basket on one-day-lagged values of Δ *Cash* and their interactions with the illiquidity of the ETF's index. Δ *Cash* is the difference between the fraction of cash in the ETF's portfolio and the average of those fractions over the prior month. The three measures of index illiquidity are the effective tick size (IL1), imputed roundtrip cost (IL2), and interquartile range (IL3). Observations are at the ETF-day level. An ETF fixed effect and a control variable for the ETF's bid-ask spread in the secondary market is included in all specifications. Standard errors, which are clustered at the fund and day level, are reported in parentheses. $^{*}p < .1;^{**}p < .0.5;^{***}p < .01$.

	RD Basket			CR Basket		
	IL1	IL2	IL3	IL1	IL2	IL3
Δ Cash	3.08**	3.17**	3.16**	−3.39***	−3.58***	−3.34***
	(1.46)	(1.50)	(1.50)	(1.03)	(1.04)	(1.09)
Δ Cash × IL	−1.43**	−1.18**	−1.12**	1.68**	1.59***	1.29*
	(0.58)	(0.51)	(0.48)	(0.63)	(0.54)	(0.67)
Observations	2,272	2,272	2,272	5,108	5,108	5,108
Adjusted R^2	0.13	0.13	0.13	0.09	0.09	0.09

Table 5
Active basket management

This table reports the slope estimates from the regressions of a bond's basket inclusion on $\Delta Deviation$ and its interaction with the illiquidity of the ETF's index. Basket inclusion is measured either by an indicator variable, which is equal to one if the bond is included in the basket and zero otherwise (panel A), or by the number of bond shares in the basket using a Poisson regression (panel B). $\Delta Deviation$ for a given bond held by a given ETF is an index-rebalancing-induced shock to the bond's overweighting in the ETF's portfolio relative to the index. The three measures of index illiquidity are the effective tick size (IL1), imputed roundtrip cost (IL2), and interquartile range (IL3). Observations are at the ETF-bond-day level. An ETF-day fixed effect is included in all specifications. The dependent variable in panel A by 100 to improve readability of the coefficients. Standard errors, which are clustered at the bond, day, and rebalancing-group level, are reported in parentheses. $^{*}p < .1;^{**}p < .0.5;^{***}p < .01$.

A. Basket indicator

	RD Basket			CR Basket		
	IL1	IL2	IL3	IL1	IL2	IL3
Δ Deviation	1.98***	1.74***	1.99***	−2.96***	−2.88***	−2.97***
	(0.19)	(0.18)	(0.19)	(0.14)	(0.14)	(0.14)
Δ Deviation × IL	−0.59***	−0.32**	−0.57***	0.58***	0.54***	0.53***
	(0.14)	(0.14)	(0.14)	(0.11)	(0.11)	(0.11)
Observations	2,726,592	2,726,592	2,726,592	7,803,001	7,803,001	7,803,001
Adjusted R^2	0.411	0.411	0.411	0.303	0.303	0.303

B. Basket shares

	RD Basket			CR Basket		
	IL1	IL2	IL3	IL1	IL2	IL3
Δ Deviation	0.15***	0.11***	0.16***	−0.16***	−0.13***	−0.17***
	(0.04)	(0.03)	(0.04)	(0.03)	(0.02)	(0.03)
Δ Deviation × IL	−0.06**	−0.03	−0.06**	0.05**	0.01	0.05***
	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)	(0.02)
Amount Outstanding	0.80***	0.80***	0.80***	0.79***	0.79***	0.79***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Observations	2,082,410	2,082,410	2,082,410	5,574,849	5,574,849	5,574,849
Pseudo R^2	0.414	0.413	0.414	0.392	0.391	0.392

Table 6
The effect of basket inclusion on bond liquidity

This table reports the slope estimates from the regressions of an individual bond's illiquidity on the bond's basket inclusion on the previous day. There are two measures of basket inclusion: an indicator variable tracking if the bond is included in the basket of an ETF (columns 1 to 3), and the number of ETFs that include this bond in their baskets (columns 4 to 6). There are three measures of bond illiquidity: the effective tick size (IL1), imputed roundtrip cost (IL2), and interquartile range (IL3). We standardize illiquidity by its mean and standard deviation and multiply it by 100 to improve readability of the coefficients. We control for the average bond illiquidity over the prior week, the average basket size of ETFs holding the bond. Observations are at the bond-day level. All specifications include firm-day and maturity-day fixed effects. Standard errors, which are clustered at the bond and day level, are reported in parentheses. $^{*}p < .1;^{**}p < .0.5;^{***}p < .01$.

	Basket Indicator			Number of Baskets		
	IL1	IL2	IL3	IL1	IL2	IL3
RD	−5.21***	−2.57***	−2.81***	−4.22***	−1.78***	−2.25***
	(0.30)	(0.32)	(0.28)	(0.26)	(0.27)	(0.26)
CR	−4.98***	−1.66***	−2.81***	−3.66***	−1.00***	−1.95***
	(0.21)	(0.21)	(0.21)	(0.14)	(0.14)	(0.14)
Bond IL	12.70***	11.81***	19.68***	12.69***	11.82***	19.69***
	(0.19)	(0.19)	(0.35)	(0.19)	(0.19)	(0.35)
Avg Basket Size	−0.80***	−1.48***	−2.12***	−0.86***	−1.58***	−2.17***
	(0.28)	(0.28)	(0.27)	(0.28)	(0.28)	(0.27)
Observations	3,251,815	2,829,012	2,898,616	3,251,815	2,829,012	2,898,616
Adjusted R^2	0.23	0.13	0.44	0.23	0.13	0.44

Table 7
First stage: The effect of index rebalancing on basket inclusion

This table reports the slope estimates from the first-stage regression of basket inclusion on our instruments, $CRInstr_{it}$ ("CR Instrument") and $RDInstr_{it}$ ("RD Instrument"). Basket inclusion is measured either by an indicator variable tracking if the bond is included in the basket of an ETF (panel A) or by the number of ETFs that include this bond in their baskets (panel B). We control for the average bond illiquidity over the prior week, the average basket size of ETFs holding the bond, the log amount of the bond outstanding, changes in index weights, and changes in institutional holdings. The three measures of bond illiquidity are the effective tick size (IL1), imputed roundtrip cost (IL2), and interquartile range (IL3). Observations are at the bond-day level. All specifications include firm-day and maturity-day fixed effects. We multiply both dependent variables by 100 to improve readability of the coefficients. Standard errors, which are clustered at the bond and day level, are reported in parentheses. $^{*}p < .1$;$^{**} p < .0.5$;$^{***} p < .01$.

A. Basket indicator

	RD Basket			CR Basket		
	IL1	IL2	IL3	IL1	IL2	IL3
RD Instrument	0.82***	0.83***	0.80***	−0.12***	−0.13***	−0.13***
	(0.09)	(0.09)	(0.09)	(0.04)	(0.04)	(0.04)
CR Instrument	−0.48***	−0.47***	−0.46***	−1.45***	−1.43***	−1.42***
	(0.05)	(0.05)	(0.05)	(0.07)	(0.07)	(0.07)
Bond IL	−0.54***	−0.34***	−0.48***	−0.68***	−0.25***	−0.71***
	(0.06)	(0.05)	(0.08)	(0.07)	(0.05)	(0.11)
Avg Basket Size	6.49***	6.44***	6.47***	7.29***	7.28***	7.29***
	(0.64)	(0.64)	(0.64)	(0.60)	(0.61)	(0.61)
Log Amount Outstanding	9.35***	9.52***	9.39***	14.67***	15.07***	14.61***
	(0.39)	(0.39)	(0.39)	(0.35)	(0.36)	(0.35)
Change in Index Weights	−0.05	−0.03	−0.05	0.06	0.05	0.04
	(0.06)	(0.06)	(0.06)	(0.07)	(0.08)	(0.07)
Change in Institutional Holdings	−0.15**	−0.14**	−0.17***	−0.30***	−0.31***	−0.33***
	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)	(0.06)
Observations	2,020,680	1,753,717	1,803,584	2,020,680	1,753,717	1,803,584
Adjusted R^2	0.46	0.46	0.46	0.39	0.39	0.39

B. Number of baskets

	RD Basket			CR Basket		
	IL1	IL2	IL3	IL1	IL2	IL3
RD Instrument	1.02***	1.04***	1.01***	−0.20***	−0.21***	−0.22***
	(0.11)	(0.11)	(0.11)	(0.06)	(0.06)	(0.06)
CR Instrument	−0.54***	−0.52***	−0.52***	−2.62***	−2.60***	−2.61***
	(0.05)	(0.05)	(0.05)	(0.14)	(0.14)	(0.14)
Bond IL	−0.56***	−0.34***	−0.49***	−1.06***	−0.19**	−0.89***
	(0.07)	(0.06)	(0.10)	(0.10)	(0.08)	(0.15)
Avg Basket Size	8.95***	8.90***	8.92***	11.13***	11.12***	11.27***
	(0.81)	(0.83)	(0.82)	(0.91)	(0.92)	(0.93)
Log Amount Outstanding	11.13***	11.34***	11.19***	22.97***	23.68***	23.10***
	(0.49)	(0.49)	(0.49)	(0.67)	(0.69)	(0.68)
Change in Index Weights	−0.05	−0.02	−0.06	0.29**	0.29**	0.25*
	(0.08)	(0.08)	(0.08)	(0.13)	(0.14)	(0.13)
Change in Institutional Holdings	−0.21***	−0.20***	−0.23***	−0.55***	−0.57***	−0.57***
	(0.07)	(0.07)	(0.07)	(0.10)	(0.10)	(0.10)
Observations	2,020,680	1,753,717	1,803,584	2,020,680	1,753,717	1,803,584
Adjusted R^2	0.51	0.51	0.51	0.43	0.43	0.44

Table 8
Second stage: The effect of instrumented basket inclusion on bond liquidity

This table reports the slope estimates from the second-stage regression of next-day bond illiquidity on instrumented basket inclusion. There are two measures of basket inclusion: an indicator variable tracking if the bond is included in the basket of an ETF (columns 1 to 3), and the number of ETFs that include this bond in their baskets (columns 4 to 6). There are three measures of bond illiquidity: the effective tick size (IL1), imputed roundtrip cost (IL2), and interquartile range (IL3). We standardize illiquidity by its mean and standard deviation and multiply it by 100 to improve readability of the coefficients. We control for the average bond illiquidity over the prior week, the average basket size of ETFs holding the bond, the log amount of the bond outstanding, changes in index weights, and changes in institutional holdings. Observations are at the bond-day level. All specifications include firm-day and maturity-day fixed effects. Standard errors, which are clustered at the bond and day level, are reported in parentheses. These standard errors are properly adjusted for the fact that the fitted values from the first stage are estimated variables. We report the Kleibergen-Paap F-statistic, which jointly tests for weak identification of the instruments included in each specification and is robust to clustering. $^{*}p < .1;^{**} p < .0.5;^{***} p < .01$.

	Basket Indicator			Number of Baskets		
	IL1	IL2	IL3	IL1	IL2	IL3
RD	−13.82*	−18.81**	−5.54	−11.07*	−14.96**	−4.26
	(7.20)	(8.40)	(6.26)	(5.78)	(6.71)	(4.94)
CR	−14.90***	−9.41*	−17.62***	−8.50***	−5.56**	−9.73***
	(4.54)	(5.00)	(3.94)	(2.45)	(2.69)	(2.07)
Bond IL	10.40***	11.03***	17.60***	10.43***	11.05***	17.65***
	(0.20)	(0.22)	(0.36)	(0.20)	(0.22)	(0.37)
Avg Basket Size	0.88	1.39**	0.03	0.83	1.44**	−0.14
	(0.54)	(0.66)	(0.52)	(0.57)	(0.70)	(0.54)
Log Amount Outstanding	−13.73***	−3.22***	−4.89***	−14.02***	−3.41***	−5.26***
	(0.81)	(0.98)	(0.75)	(0.77)	(0.94)	(0.71)
Change in Index Weights	0.18**	0.17*	0.04	0.20**	0.18**	0.06
	(0.08)	(0.09)	(0.07)	(0.08)	(0.09)	(0.07)
Change in Institutional Holdings	−0.82***	−0.92***	−1.14***	−0.82***	−0.93***	−1.14***
	(0.09)	(0.11)	(0.09)	(0.09)	(0.11)	(0.09)
Observations	2,020,680	1,753,717	1,803,584	2,020,680	1,753,717	1,803,584
First-Stage F-Statistic	52.21	51.65	47.94	52.65	52.42	50.28
Adjusted R^2	0.01	0.01	0.02	0.01	0.01	0.02

Table 9
The effect of basket inclusion on bond liquidity during COVID-19

This table is the counterpart of Table 6 estimated over the COVID subperiod. Instead of using the full sample of 2017 to 2020, this table uses the subperiod of March 2 to April 15, 2020.

	Basket Indicator			Number of Baskets		
	IL1	IL2	IL3	IL1	IL2	IL3
RD	2.54*	4.89***	7.49***	0.92	3.99***	4.69***
	(1.43)	(1.38)	(2.52)	(0.94)	(1.01)	(1.50)
CR	1.01	−0.54	6.70**	−1.18	−0.30	3.04
	(1.80)	(1.09)	(2.70)	(1.30)	(0.81)	(1.88)
Bond IL	9.40***	7.96***	15.37***	9.42***	7.95***	15.42***
	(0.66)	(0.55)	(0.74)	(0.65)	(0.55)	(0.74)
Avg Basket Size	−0.71	−1.61	−1.57	−0.22	−1.70	−1.00
	(1.16)	(1.07)	(2.09)	(1.20)	(1.06)	(2.16)
Observations	111,703	100,251	101,252	111,703	100,251	101,252
Adjusted R^2	0.26	0.08	0.36	0.26	0.08	0.36

Table 10
Interactions with basket imbalance

This table reports the slope estimates from the regressions of an individual bond's next-day illiquidity on the number of baskets that contain the bond in the prior week and its interactions with two measures of the bond's basket imbalance. These imbalance measures are defined in Equations (18) and (19). There are three measures of bond illiquidity: the effective tick size (IL1), imputed roundtrip cost (IL2), and interquartile range (IL3). We standardize illiquidity by its mean and standard deviation and multiply it by 100 to improve readability of the coefficients. We control for the average bond illiquidity over the prior week and the average basket size of ETFs holding the bond. Observations are at the bond-day level. All specifications include firm-day and maturity-day fixed effects. Standard errors, which are clustered at the bond and day level, are reported in parentheses. $^{*}p < .1;^{**}p < .0.5;^{***}p < .01$.

	(1) IL1	(2) IL2	(3) IL3
Num Baskets	-10.83^{***}	-4.14^{***}	-6.50^{***}
	(0.40)	(0.40)	(0.40)
Num Baskets × CR Imbal	1.79^{***}	1.80^{***}	1.55^{***}
	(0.23)	(0.23)	(0.21)
Num Baskets × RD Imbal	3.05^{***}	1.43^{**}	2.52^{***}
	(0.50)	(0.69)	(0.64)
Bond IL	12.45^{***}	11.79^{***}	19.56^{***}
	(0.18)	(0.19)	(0.35)
Avg Basket Size	-0.48^{*}	-1.44^{***}	-1.91^{***}
	(0.26)	(0.27)	(0.26)
Observations	3,251,815	2,829,012	2,898,616
Adjusted R^2	0.23	0.13	0.44