Sparse Portfolios and Benchmarking in Corporate Bond Markets[§]

Zheng Li,[*] Anna Pavlova,[†] and Taisiya Sikorskaya[‡]

August 27, 2026

Abstract

We use comprehensive fixed-income benchmark-index data in Canada and the U.S. to show how benchmarking shapes corporate bond fund ownership and prices. Funds hold sparse portfolios, and index weights strongly influence which bonds active and passive funds select. A model of benchmarked managers facing portfolio management costs rationalizes these patterns and predicts which bonds managers exclude. A bond's price rises with its benchmarking intensity (BMI)—capital benchmarked against the bond. Sparsity attenuates this effect for excluded bonds. Exploiting discontinuities at maturity cutoffs, we find that increases in BMI indeed lower yield spreads and raise fund ownership, primarily for included bonds.

JEL Classification: G11, G12, G23

Keywords: Benchmark, corporate bonds, sampling, index rebalancing, mutual funds, ETFs

[§]We thank Manav Chaudhary (discussant), Linda Du (discussant), Peter Feldhütter, Paul Huebner (discussant), Yiming Ma, Yann Neto, Yoshio Nozawa (discussant), Lubos Pastor, Stephen Schaefer, and Yao Zeng for helpful comments, along with seminar participants at the 6th MBS–Southampton–Surrey Virtual Research Seminar; Bocconi University; Cambridge University; City University of Hong Kong; CUHK; the DAM Working Group; Georgia State University; HKU; HKUST; Johns Hopkins University; Michigan State University; Vanderbilt University; Warwick Business School; and Università della Svizzera italiana, and conference participants at the 2025 CAMR Conference at Bayes Business School, the 2026 EFA Annual Meeting, the 2026 UC Davis–FMA Napa Finance Conference, the 2026 Duke-UNC Asset Pricing Conference, the 2026 SFS Cavalcade North America, the 2025 WAPFIN (NYU Stern), and the 2026 World Symposium on Investment Research. We are grateful to Yuan Gao, Nilesh Hegde, Austin Liu, Daniel Yoon, and especially Jinrui Wang for excellent research assistance. This research was supported by the Fama-Miller Center for Research in Finance at the University of Chicago Booth School of Business.

[*]London Business School, zli@london.edu

[†]London Business School and CEPR, apavlova@london.edu

[‡]University of Chicago Booth School of Business, taisiya.sikorskaya@chicagobooth.edu

1 Introduction

An important trend in the corporate bond market is the increasing presence of mutual funds and exchange-traded funds (ETFs). For example, in the U.S., mutual funds and ETFs have increasingly replaced households as significant domestic holders of corporate bonds (see Appendix Figure A1). Academics and policymakers have raised concerns that this may contribute to increased financial fragility in bond markets (see, e.g., Cai, Helmke, Mosk, and Suntheim, 2025), underscoring the importance of understanding the behavior of these investment vehicles. Yet, standard portfolio-choice models often abstract from two features that are central to fixed-income asset management. First, fund performance is typically evaluated relative to a benchmark index, directly influencing portfolio choices made by fund managers. Second, due to the large size of the fixed-income investment universe—generally much larger than that of equities—it is impractical for funds to hold every available bond. As a result, funds typically construct sparse portfolios comprising only a limited subset of bonds, a practice known as sampling.

We ask how benchmarking affects corporate bond ownership and prices. Using data covering most fixed-income benchmarks in Canada and the U.S., we document that fund managers hold sparse portfolios and that benchmark weights strongly predict bond ownership of both active and passive funds. We rationalize these findings in a model in which benchmarked managers face portfolio management costs proportional to portfolio size. While adding an asset is costly, excluding it increases the tracking error of the manager's portfolio. The model predicts that fund managers are more likely to hold assets with larger size, higher benchmark weight, and higher idiosyncratic risk. Moreover, the price of an asset increases with its benchmarking intensity (BMI), which is a measure of the amount of capital benchmarked against the asset. However, because managers optimally hold sparse portfolios, benchmark-induced demand effects are concentrated in the assets they include and muted for those they exclude. Exploiting discontinuities in benchmarked assets around bond maturity cutoffs, we provide empirical evidence consistent with the model's predictions: increases in bonds' BMIs lead to reductions in yields and increases in fund ownership, but primarily for

bonds predicted to enter sparse portfolios.

We first present stylized facts for a comprehensive sample of fixed-income benchmark indexes used by mutual funds and ETFs in the U.S. and Canada. We find that, in both countries, the set of benchmarks is heterogeneous, catering to different investment mandates and styles. In contrast to equity benchmarks, bond benchmarks have received limited attention in the literature so far. We document that they generally contain far more assets than equity benchmarks and exhibit significantly higher turnover. The two-way turnover is over 40% annually for the Bloomberg U.S. Aggregate Bond Index compared to approximately 10% for the S&P 500. This high turnover imposes considerable rebalancing costs, making it impractical for fund managers to hold every bond in their benchmark index. Nevertheless, fund managers keep their funds' performance close to their benchmark indexes, as evidenced by the tracking error levels we document.

Our next set of stylized facts concerns fund portfolios of investment-grade corporate bonds. We document that these portfolios are highly sparse: U.S. fixed-income funds hold only 7% of the bonds included in their benchmarks, whereas Canadian funds hold approximately 15%. Even passive mutual funds and ETFs hold only a fraction of the assets in their benchmarks—about 27% in the U.S. and 47% in Canada. The Canadian corporate bond market is much narrower, yet there is still significant sparsity. These figures imply a level of sparsity substantially greater than what is typically observed in equity portfolios.

We then show that a bond's benchmark membership and weight strongly predict the likelihood of its inclusion in portfolios, underscoring the central role of benchmarks in portfolio selection. This pattern is not confined to passive vehicles but holds across both active and passive funds. Both types of funds tend to hold more bonds with higher benchmark weights. However, since managers typically seek exposure across all relevant bond categories (e.g., maturity and credit rating), they hold at least some bonds from the entire range of benchmark weights. We document sparsity even within narrowly defined categories.

In the second part of the paper, we develop a novel theory of portfolio sparsity in asset management and derive its asset pricing implications. Our model features fund managers

who are benchmarked to an index and face portfolio management costs, which we interpret as rebalancing, monitoring, and related operational frictions. These costs rise with the number of assets held, preventing full benchmark replication and inducing sparse portfolios.

When considering the inclusion of an additional asset, managers trade off the marginal benefit—through reduced tracking error and improved expected returns—against the incremental portfolio management cost. As a result, assets with low benchmark weights are more likely to be excluded, given their limited contribution to tracking error reduction. Funds with stronger benchmarking incentives exclude fewer assets.

The model highlights asset substitutability as another determinant of which assets are excluded. Specifically, managers are more likely to exclude assets with many close substitutes. We find that, within each category of closely substitutable assets, managers are more likely to hold those with higher benchmark weights, larger size, and higher idiosyncratic risk. Furthermore, sparsity is higher for broader benchmarks (i.e., when average benchmark weights are lower), when idiosyncratic risk is lower, and when trading costs are higher.

When deriving optimal portfolios, we show that a portion of fund managers' demand is inelastic—even among active managers. This inelastic demand affects equilibrium asset prices: an asset's price depends on its BMI, which captures the aggregate inelastic demand it attracts from benchmarked managers. In equilibrium, assets with higher BMI command higher prices and offer lower expected returns. Portfolio sparsity refines this prediction. Because managers optimally choose sparse portfolios, benchmark-induced demand is concentrated in included assets and attenuated for excluded ones. This mechanism is particularly relevant in corporate bond markets. For example, within categories of substitutable assets—such as rating or maturity buckets—smaller bonds are predicted to be excluded from fund portfolios, so fund flows have little effect on their prices.

In the third part of the paper, we test these predictions for investment-grade corporate bonds in Canada and the U.S. We first construct a measure of BMI. Consistent with our theoretical framework, the measure includes both passive and active assets under management (AUM). In our sample, active AUM substantially exceeds passive AUM, especially

in Canada. A bond's total BMI is calculated as its cumulative weight in all benchmarks, weighted by assets following each benchmark, and divided by the market value of the bond.

Importantly, as time passes, bonds roll down the maturity spectrum and transition from long- to intermediate- and short-term market indexes. Because these indexes contain different amounts of benchmarked capital, this roll-down produces sizeable changes in BMI. As a result, much of the within-bond variation in BMI comes from crossing maturity-based cutoffs. Our identification strategy exploits such index reconstitutions, which create quasi-exogenous variation in BMI.

Our research design additionally distinguishes between assets predicted to be included in sparse portfolios and those predicted to be excluded. Guided by the model, we use bond size as a key determinant of inclusion. Specifically, we predict inclusion by sorting bonds by par value outstanding within sector-maturity–rating buckets and classifying the bottom deciles as excluded and the remainder as included. Consistent with the model's predictions, increases in BMI are followed by persistent declines in yield spreads for included bonds in both Canada and the U.S., and by substantially weaker and largely undetectable responses for excluded bonds. While the concentration of effects in larger bonds may seem at odds with the literature documenting stronger demand effects among smaller assets (Wurgler and Zhuravskaya, 2002), it aligns with our theoretical framework: within a bucket of bonds with similar cash flows, the bonds with the largest benchmark weights are those a manager can least afford to drop, so their prices are the most sensitive to benchmark-driven demand.

The data on bond ownership lend further support to our mechanism. First, we document that long-term funds reduce their holdings of bonds crossing relevant maturity cutoffs, while short-term funds increase theirs. Importantly, this is true both for active and passive funds. While this result is somewhat expected for passive funds, the literature to date has not provided strong evidence that active fixed-income funds buy additions and sell deletions to their benchmark indexes. Second, increases in BMI lead to higher ownership of both active and passive funds. Consistent with our model, such an ownership response is concentrated in bonds predicted to be included in sparse portfolios.

To address alternative interpretations of the importance of bond size for our results, we document sharp ownership discontinuities around benchmark size-eligibility cutoffs in both Canada and the U.S. Even for the smallest bonds, there is a discrete jump in fund ownership exactly at the threshold for benchmark inclusion. These discontinuities hold for both active and passive funds, ruling out the simple interpretation that issue size predicts fund ownership solely because it is a smooth proxy for liquidity.

Related Literature. Our paper contributes to several strands of the literature. Our model is related to the literature on benchmarking, mandates, and habitats. Like us, Brennan (1993), Basak and Pavlova (2013), He and Xiong (2013), Kashyap, Kovrijnykh, Li, and Pavlova (2021), Buffa, Vayanos, and Woolley (2022), and Pavlova and Sikorskaya (2023) study the implications of benchmarking for portfolio management and asset prices. However, none of these models produces portfolio sparsity or explores how it affects asset prices.

Previous research proposes several mechanisms that can generate portfolio sparsity, including ambiguity aversion (Dow and Werlang, 1992) and limited attention or information-processing constraints (Gabaix, 2014; Van Nieuwerburgh and Veldkamp, 2010). Our mechanism is distinct. In particular, the resulting portfolios are not concentrated in particular industries or in assets that attract managers' attention. In Merton (1987), investors hold only "familiar" stocks, leading to price differences between familiar and neglected assets. In contrast, market bifurcation in our model is endogenous, and the theory characterizes the properties of included and excluded assets.

There are several empirical papers that discuss sparsity in holdings. Koijen and Yogo (2019) introduce a demand system approach to asset pricing and discuss how it should be adjusted to handle sparse portfolios. Graves (2024) investigates the origins of portfolio sparsity of institutional investors and attributes it to binding short-sale constraints on assets they choose not to hold. The mechanism generating sparsity that is closest to ours is discussed in the ETF context by Shim and Todorov (2022), Koont, Ma, Pastor, and Zeng (2025), and Brogaard, Heath, and Huang (2026). The first two of these papers note that ETF creation/redemption baskets exclude some (illiquid) assets and attribute this to a trade-off

between transaction costs and tracking error. They argue that basket choices may contribute to financial fragility. We study fund portfolios and argue that benchmarking and portfolio management costs affect the allocations of both passive and active bond fund managers.

More generally, our work connects to the emerging literature on institutional demand in the corporate bond market. Dannhauser (2017), Dick-Nielsen and Rossi (2019), Dannhauser and Hoseinzade (2022), Ma, Xiao, and Zeng (2022), Bretscher, Schmid, Sen, and Sharma (2025), Chaudhary, Fu, and Li (2026), Gabaix, Koijen, Richmond, and Yogo (2025), and Sialm, Shin, and Zhu (2025) shed light on the nature of institutional investors' portfolios in this market and their price impact, but do not discuss the importance of fixed-income benchmarks in asset allocation. Choi, Cremers, and Riley (2025) do consider fund benchmarks, but their focus is on active shares and performance of passive versus active bond mutual funds rather than the effects of benchmarking on bond prices. The latter are discussed in Ottonello (2019), who studies high-yield bond prices and funds' liquidity management. A recent paper by Feldhütter, Poulsen, and Woschitz (2026) introduces a complementary transaction-based measure of benchmark trading intensity and, like us, show that the price and ownership effects of index exclusion are concentrated among bonds that are more frequently used to implement benchmark demand.[1] Our measure is an ex ante model-based prediction of which bonds enter sparse portfolios, constructed from both passive and active benchmarked AUM without using trading data. The paper closest to ours is Bretscher, Schmid, and Ye (2024), which proposed to use index maturity cutoffs to study the effects of passive ownership on bond prices. In contrast, we find that active funds also shape inelastic demand and prices around the cutoffs, especially if one accounts for sparsity. Finally, rather than using a maturity-cutoff dummy as an instrument for fund ownership, we employ BMI—a continuous measure that provides richer variation than a cutoff indicator.

[1]Related evidence in market microstructure literature highlights the growing importance of index-tracking trades by mutual funds and ETFs in the corporate bond market (see, e.g., Li, O'Hara, Rapp, and Zhou, 2025; Shin, Zhou, and Zhu, 2025). Theoretically, even funds with weaker index-tracking incentives may optimally trade on index-rebalancing days (Pegoraro, Sammon, and Shim, 2026).

2 Stylized Facts

In this section, we document new stylized facts about corporate bond fund portfolios and benchmark indexes in Canada and the U.S.

2.1 Corporate bond investment management

Mutual funds and ETFs have become increasingly important holders of corporate bonds, especially in the investment-grade segment. Their share of outstanding corporate debt has risen steadily over the past two decades (see Appendix Figure A1), making these funds key intermediaries in the transmission of demand shocks and in the provision of secondary market liquidity. Managers of these funds almost universally operate relative to benchmark indexes, which serve two purposes: they define the investment universe for eligible holdings and provide the tool for performance evaluation (Ma, Tang, and Gomez, 2019).

However, as we demonstrate in Section 2.3, the corporate bond benchmarks themselves are extremely broad, often spanning thousands of individual securities across maturities, sectors, and rating buckets. Fully replicating these indexes is operationally infeasible and would impose prohibitively high portfolio management costs. Consequently, managers face a trade-off between tracking error and operational complexity: they hold relatively sparse subsets of benchmark constituents while seeking to remain close to benchmark performance. This trade-off is clearly illustrated in Bloomberg's portfolio allocation tools, snapshots of which we report in Appendix A.2. These tools highlight how portfolio managers balance tracking error targets with the desire to minimize portfolio management costs, resulting in sparse but benchmark-oriented bond portfolios.

To operationalize this sparsity, many fixed-income index fund managers employ stratified sampling techniques—selecting a representative subset of benchmark securities that aligns with the portfolio's key risk factors such as duration, credit quality, sector weights, and convexity—while keeping portfolio management costs manageable. For example, Vanguard describes this method in its white papers and fund materials, emphasizing that its

bond index funds use stratified sampling strategies to approximate index exposure without the need to hold every constituent security.[2]

2.2 Data

We construct our dataset from multiple sources. For Canada, we obtain benchmark constituent weights for all Canada Domestic Fixed Income Indexes from FTSE Russell, part of London Stock Exchange Group (LSEG), from July 2003 to June 2024. The FTSE data also include daily bond characteristics and pricing data as well as daily index returns. We complement the bond-level data with LSEG variables such as Return Index and bid-ask spread. We use the Bank of Canada's bond yield curves to construct maturity-matched credit spreads.[3] We collect mutual fund and ETF holdings and static and dynamic fund characteristics from Morningstar, including a snapshot of funds' primary prospectus benchmarks and fund returns. Because we observe all data on Canadian mutual funds, ETFs, and separate accounts in Morningstar, we treat all these fund structures as separate funds as long as they have different portfolio identifiers. We use USD-CAD exchange rates from FRED to align the fund asset currency (USD) with bond market value currency (CAD). Finally, we classify funds into active and passive using Morningstar's 'index fund' and 'investment type' fields. Table A1 in the Appendix characterizes our sample of Canadian funds.

For the U.S., we obtain benchmark constituent weights for all Fixed Income Indexes from Bloomberg Aggregate index family as well as the bond characteristics and pricing data in 2012–2023. We get additional bond-level variables, including bid-ask spread and issue date, from WRDS bond return database. The mutual fund and ETF holdings as well as their static and dynamic characteristics come from CRSP Mutual Fund database. We augment the assets of funds in our sample by the size of their matched separate accounts reported in Morningstar, as those separate accounts hold virtually identical portfolios and contribute significantly to AUM (Huang, Lu, Song, and Xiang, 2026). We use CRSP to classify funds

[2]See Vanguard research note https://corporate.vanguard.com/content/dam/corp/research/pdf/a_bond_index_funds_balancing_act_tracking_error_and_cost.pdf.
[3]See https://www.bankofcanada.ca/rates/interest-rates/bond-yield-curves/.

into active and passive, as described in Appendix A.11.2. We further extract a dynamic panel of each fund's primary prospectus benchmark over time directly from regulatory filings on the U.S. SEC's EDGAR system; we combine this panel with a Morningstar snapshot and the U.S. SEC Mutual Fund Risk/Return database. Details and validation exercises are provided in Appendix A.11. Importantly, by tracking fund benchmark histories and benchmark composition histories, we are able to measure shifts in benchmarking intensity over time. Table A2 in the Appendix characterizes our sample of U.S. funds.

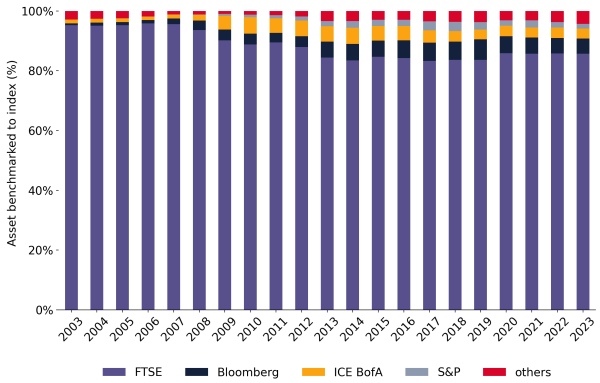
In Canada, we only consider corporate bonds by selecting 'corporate' FTSE industry sector. Within that sector, there are 7 industry groups (communication, energy, financial, industrial, infrastructure, real estate, and securitization) and 21 finer industry subgroups. In the U.S., we filter for corporate bonds by setting Bloomberg's bclass1 to 'corporates.' Within this class, there are 3 industry groups (financial institutions, industrial, and utility) that include 18 finer industry subgroups and 58 sectors. When constructing a bucket of bond substitutes in Section 4, we use the most granular available classification, that is, the bond's industry subgroup in Canada and its sector in the U.S.

2.3 Stylized facts

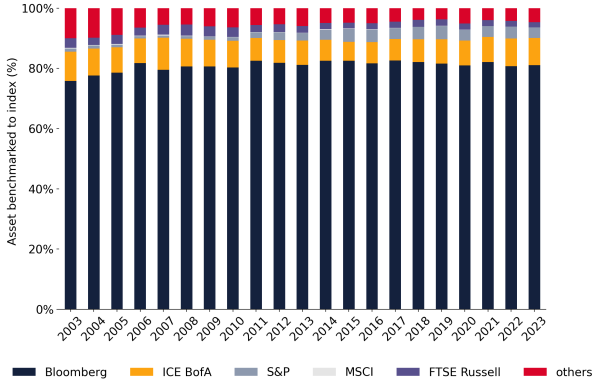
In this section, we document new stylized facts on corporate bond benchmarks and fund portfolios in Canada and the U.S., with a specific focus on the investment-grade segment of those portfolios.

Among the universe of fixed-income benchmarks adopted by mutual funds and ETFs in Canada, over 90% of total assets are benchmarked to FTSE indexes (Figure 1a). Therefore, our detailed benchmark data from FTSE cover most of the assets of those institutional investors. Importantly, within FTSE Canada bond index family, there is an entire suite of indexes—such as short-term, mid-term, long-term, and aggregate—catering to different investment objectives. The assets under management benchmarked to each of them are shown in Figure 1c. In the U.S., on the other hand, FTSE Russell is very small, and the market is dominated by Bloomberg indexes. Despite slightly more heterogeneity than

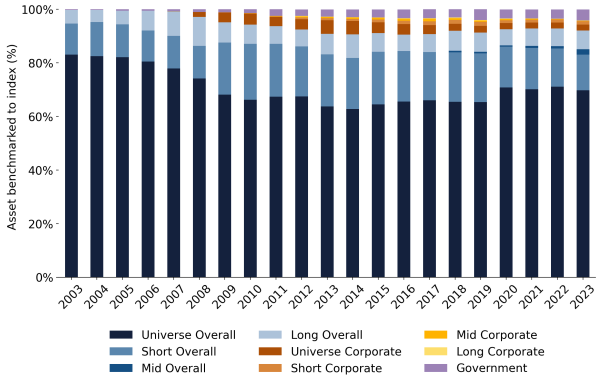
Figure 1: Fund assets benchmarked to bond indexes in Canada and the U.S.



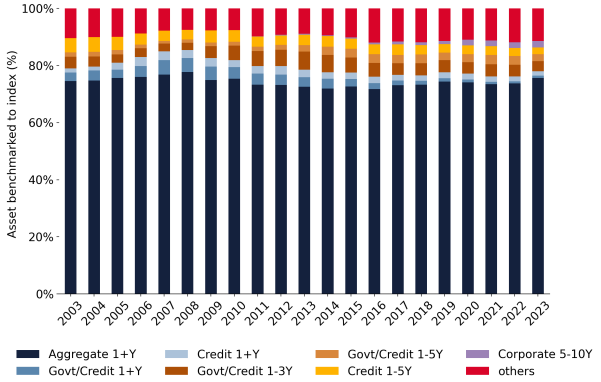
(a) Assets benchmarked to index providers (Canada)



(b) Assets benchmarked to index providers (U.S.)



(c) Assets benchmarked to FTSE indexes (Canada)



(d) Assets benchmarked to Bloomberg indexes (U.S.)

This figure shows the share of assets under management benchmarked to each index group over time for Canadian and U.S. domestic fixed-income mutual funds and ETFs.

demonstrated in Canada, Bloomberg still accounts for the dominant share of more than 80% of the total assets (Figure 1b). Within Bloomberg indexes, there are also indexes covering narrower market segments, such as short-term credit and 5-10 year corporate bonds.

Our first stylized fact concerns index turnover: fixed-income indexes typically have significantly higher turnover than equity indexes. Bond issuance and maturity create mechanical changes in bond market indexes, especially for the shorter segments of the market. This results in an average two-way turnover of 39% (71%) for the Canadian (U.S.) bond indexes in our sample, as shown in Table 1. Broad bond indexes turn over roughly three to five times as fast as the S&P 500, whose two-way turnover is around 10%. The short-maturity

indexes turn over more than 100% each year—134% for the Aggregate 1–3 year index and 150% for the Aggregate 3–5 year index—because of maturity alone. We argue that this high turnover significantly increases portfolio management costs for corporate bond portfolios.

Second, corporate bond fund portfolios are extremely sparse. Canadian funds on average hold only 15% of bonds in their benchmark indexes. Passive bond funds tend to hold more bonds, with an average of 47%, but still fewer than passive equity funds would. Active bond funds, on the other hand, hold very concentrated portfolios with only 11% of the benchmark bonds. As Table 1 illustrates, the corporate bond universe in the U.S. is roughly seven times larger than Canada's, so we see even more sparsity. U.S. funds on average hold only 7% of bonds in their benchmark indexes. Passive funds hold 27%, while active funds hold only around 4% of benchmark bonds.

This sparsity is also very persistent. As shown in the transition matrix in Appendix A.5, an asset that is not held by a fund on average has over a 96% (90%) chance of not being held over the next 3 years by U.S. (Canadian) funds. On the other hand, if an asset is held, it is likely to be held in the next quarter. However, such holding probabilities drop much more quickly, with 55–65% of bonds still being held at the 3-year horizon, which likely reflects high turnover documented above.

Despite the high turnover and sparsity levels, fund managers in our sample manage to keep their funds' performance close to their benchmark indexes. The average annualized tracking error of passive funds in our Canadian (U.S.) sample is well below 1% at 83 bps (91 bps). Even for active funds, tracking errors are low in absolute terms, at 137 bps (203 bps) for Canada (the U.S.). Relative to benchmark volatility, however, they are not small: bond benchmarks are 4–5 times less volatile than equity benchmarks, so measured as a fraction of the volatility of the index being tracked, these tracking errors are comparable to those of equity funds and are even larger for passive funds. Within all types of funds, there is considerable cross-sectional variation, with standard deviations of tracking error generally exceeding the averages we report here. Details of our tracking error calculation as well as additional statistics are reported in Appendix A.6.

Table 1: Annual Turnover of FTSE Canada and Bloomberg U.S. Fixed-Income Indexes

Benchmark	Number of bond members	Additions	Deletions	Turnover rate (%)	
				Total	Maturity-driven
Panel A: Canada FTSE Indexes					
Universe Overall Bond Index	1272	168	129	24	9
Short Overall Bond Index	489	139	126	54	40
Mid Overall Bond Index	306	89	81	56	37
Long Overall Bond Index	477	53	36	19	7
Universe Corporate Bond Index	810	124	95	28	10
Short Corporate Bond Index	349	102	91	56	38
Mid Corporate Bond Index	184	57	52	59	35
Long Corporate Bond Index	277	30	17	18	6
Universe Corporate BBB Bond Index	319	63	44	36	9
Short Mid Corporate BBB Bond Index	229	54	40	44	16
Average	471	88	71	39	21
Panel B: U.S. Bloomberg indexes					
Aggregate	9430	2292	1893	46	11
Aggregate 1-3 year	2459	1694	1611	134	86
Aggregate 1-5 year	4468	1908	1766	83	41
Aggregate 3-5 year	2010	1534	1475	150	97
Aggregate 1-10 year	7513	2067	1804	52	16
Aggregate 1-10 year Ex-BAA	5706	1757	1581	60	16
Corporate	4828	960	686	34	8
Corporate 1-3 year	893	523	474	111	92
Corporate 1-5 year	1758	619	528	65	42
Corporate 1-10 year	3283	763	607	42	16
Corporate 5-10 year	1526	476	412	58	30
Corporate 10+ year	1544	320	202	36	9
Credit	5457	1084	782	35	9
Credit 1-3 year	1035	602	549	111	92
Credit 1-5 year	2009	705	606	65	42
Credit 1-10 year	3667	859	690	42	17
Credit 5-10 year	1658	521	451	59	30
Credit 10+ year	1791	362	228	35	8
Govt/Credit	6298	1534	1232	46	11
Govt/Credit 1-3 year	1411	943	887	132	89
Govt/Credit 1-3 year Ex-BAA	996	722	692	143	87
Govt/Credit 1-5 year	2591	1117	1016	84	42
Govt/Credit 1-10 year	4405	1299	1130	57	19
Govt/Credit 1-10 year A+	2624	949	873	70	21
Govt/Credit 5-10 year	1814	606	539	64	32
Govt/Credit 10+ year	1893	383	250	36	9
Average	3195	1023	883	71	37

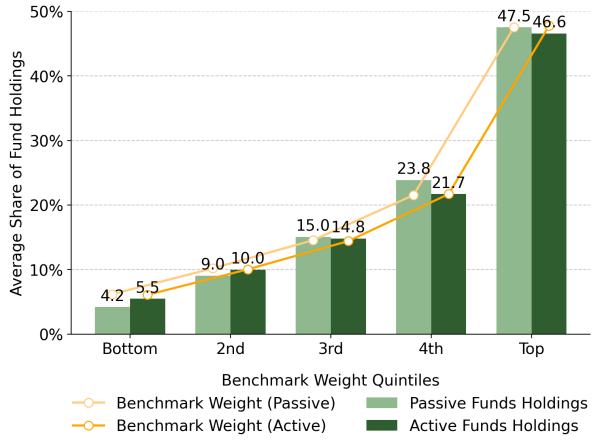
This table reports annual bond additions, deletions, and turnover rates for FTSE Canada and Bloomberg U.S. fixed-income indexes. Bloomberg U.S. Aggregate indexes are rebalanced at monthly frequency, and FTSE Canada indexes are rebalanced at daily frequency. The annual figures reported are calculated by aggregating all rebalances over a year. Total turnover is measured as two-way turnover, including both additions and deletions. Maturity-driven turnover comprises index additions and deletions resulting from bonds crossing the time-to-maturity thresholds of the benchmark. For indexes with two-sided maturity bounds (e.g., Corporate 5-10 year), bonds exiting upon falling below the lower bound are classified as maturity-driven deletions, and bonds entering upon falling below the upper bound are classified as maturity-driven additions. For indexes with a one-sided maturity constraint (such as U.S. Aggregate Index, which requires time to maturity above one year, or the Credit 10+ Year Index), maturity-driven turnover arises exclusively from deletions as bonds fall below the minimum maturity requirement. All statistics are calculated over the full sample period of benchmark index data, which spans 2005–2023 for U.S. indexes and 2004–2023 for Canadian indexes.

Exploiting the granularity of our dataset, we aggregate assets of all funds with the same benchmark and of the same type (active or passive) into one portfolio and characterize how close that portfolio is to the benchmark. Even at this aggregate level, portfolios are sparse, with active funds holding, on average, 66%–72% of benchmark bonds and passive funds holding 75%–90%. Appendix A.7 reports these statistics as well as active shares and tracking errors of the aggregate portfolios.

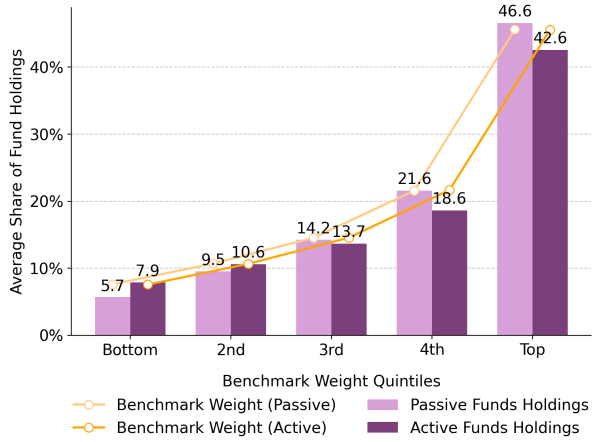
Next, we describe how benchmark index weights affect bond portfolio holdings and their sparsity within the investment-grade corporate bond segment of fund portfolios and their benchmarks. Figure 2 characterizes asset allocation across benchmark weight quintiles. We see that funds typically allocate larger shares of assets to bonds with larger index weights, with over 40% of fund assets allocated to the bonds in the top quintile of benchmark weight across the two geographies. Asset distribution across benchmark weight quintiles is similar to the distribution of benchmark weights for passive funds. Active funds deviate from benchmark weights (and hold more cash and bonds outside of the benchmark) but the distribution across benchmark weight quintiles is roughly similar. Importantly, we see that sparsity also follows this monotonic pattern. As shown in Figure 2c, both active and passive funds hold smaller fractions of benchmark bonds in the lower benchmark weight quintiles. Moreover, benchmark weights are also important within credit-rating buckets. As shown in Appendix Figure A3, funds are more likely to hold bonds in the top quintiles, particularly for bonds with higher ratings.

To shed more light on the relationship between benchmark index weights and fund portfolio holdings, we study their conditional correlations. Specifically, Table 2 reports the results of panel regressions of fund holdings on a benchmark-membership dummy variable and benchmark index weights. In all the regressions, we include both fund-by-date and bond-by-date fixed effects, which remove all the variation in fund or bond characteristics over time, such as, e.g., liquidity, duration, and rating. Thus, we only use the variation across funds in holdings of the same bond at the same time. We study both the intensive margin of holdings (portfolio weights) and the extensive margin (propensity to hold the

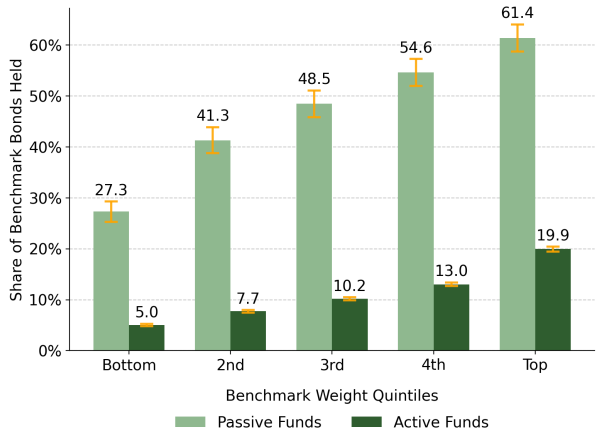
Figure 2: Fund holdings across benchmark weight quintiles



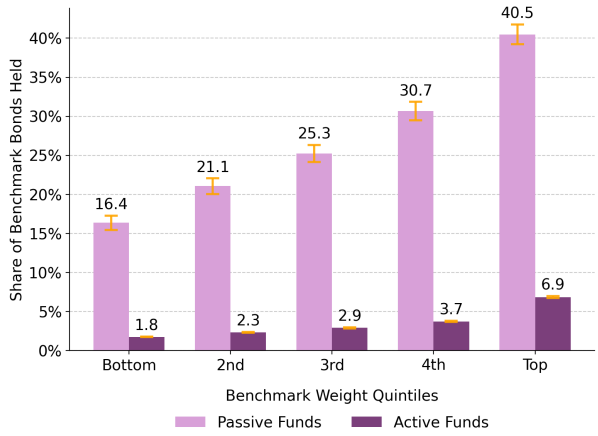
(a) Share of fund holdings across benchmark weight quintiles (Canada)



(b) Share of fund holdings across benchmark weight quintiles (U.S.)



(c) Share of benchmark bonds held across benchmark weight quintiles (Canada)



(d) Share of benchmark bonds held across benchmark weight quintiles (U.S.)

This figure illustrates allocation of active and passive fund assets across bonds in different quintiles of benchmark index weight. Passive funds include index mutual funds and ETFs. We only consider the investment-grade corporate bond segment of fund and benchmark portfolios, reweighting holdings to 100% within this segment.

bond). We take the bonds in the broadest benchmark index as the reference universe for the extensive margin analysis, that is, all corporate bonds in the Bloomberg Aggregate Bond Index in the U.S. and all corporate bonds in FTSE Canada Overall Bond Index in Canada.

Columns (1) to (3) of Table 2 show that both Canadian and U.S. funds put higher weights on bonds that are within their benchmarks and have a higher benchmark weight. Similarly, according to columns (4) to (6), funds are more likely to hold a bond (weight above

Table 2: Regressions of fund holdings on benchmark index membership and weights for Canada and the U.S.

	Fund weight | fund weight $\neq 0$			**Dummy(fund weight > 0)**		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Canada						
Dummy(within benchmark)		0.002***	0.0001		0.249***	0.163***
		(7.33)	(0.25)		(28.92)	(20.15)
Benchmark weight			0.659***			32.46***
			(9.74)			(13.68)
Observations	559,559	559,559	559,559	4,653,674	4,653,674	4,653,674
Adj. R^2	0.722	0.722	0.723	0.229	0.262	0.266
Panel B: U.S.						
Dummy(within benchmark)		0.001***	0.0003***		0.154***	0.129***
		(16.76)	(4.45)		(59.35)	(32.13)
Benchmark weight			0.563***			45.64***
			(11.75)			(17.06)
Observations	5,086,092	5,086,092	5,086,092	117,046,528	117,046,528	117,046,528
Adj. R^2	0.792	0.792	0.792	0.146	0.176	0.176

This table reports estimates of the conditional correlation between fund weight or the propensity to hold a bond and a benchmark-membership dummy variable and benchmark index weight. The specification is the following:

$$y_{i,j,t} = \alpha_{i,t} + \alpha_{j,t} + \beta_1 \, \mathbf{1}(j \in \mathrm{BM}_{i,t}) + \beta_2 \, w^{\mathrm{BM}}_{i,j,t} + \varepsilon_{i,j,t},$$

where $y_{i,j,t}$ is either the portfolio weight of bond j in fund i at time t (conditional on holding, columns 1–3) or an indicator for whether fund i holds bond j (columns 4–6); $\alpha_{i,t}$ and $\alpha_{j,t}$ are fund-by-date and bond-by-date fixed effects; $\mathbf{1}(j \in \mathrm{BM}_{i,t})$ is an indicator equal to one if bond j is a constituent of fund i's benchmark at time t; and $w^{\mathrm{BM}}_{i,j,t}$ is the benchmark index weight of bond j. We only consider the investment-grade corporate bond segment of fund and benchmark portfolios, reweighting holdings to 100% within this segment. Panel A reports the results for Canada, and Panel B for the U.S. The reference universe is the broadest benchmark available in each geography, i.e., the Bloomberg Aggregate Index in the U.S. and the Universe Overall Bond Index in Canada. All regressions are saturated by including fund-by-date and bond-by-date fixed effects. t-statistics clustered at the bond and year-month levels are presented in parentheses. $^*p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

zero) if it is in their benchmark index, especially if it has a high weight in the benchmark. Moreover, these variables contribute significantly to the R^2, explaining a considerable share of variation in fund holdings especially on the extensive margin (columns (4)–(6)). These findings are overall very similar between the two geographies. We have also verified that these results are unchanged if, instead of including both fund-by-date and bond-by-date fixed effects, we include a comprehensive set of control variables, such as bond size, rating, liquidity, duration, convexity, price, yield, as well as fund-by-date fixed effects and bond fixed effects.

Importantly, we see that these correlations are not driven by passive funds only.

Appendix Table A9 reports the estimates in the subsamples of active and passive funds. Even though benchmark weights and especially benchmark membership explain a larger share of variation in passive fund portfolios, these results highlight the importance of benchmarks in the asset allocation of active funds as well.

Overall, the stylized facts we documented above suggest that benchmark membership and benchmark weight are important contributors to funds' portfolio choices, both at the extensive margin (whether to include a bond in the portfolio) and at the intensive margin (how much of it to hold). This underscores an interaction between benchmarking and portfolio management costs in corporate bond markets, which we will explore theoretically in the following section.

3 Model

The model features two periods, $t = 0, 1$. The financial market consists of a riskless asset with an exogenous interest rate normalized to zero (e.g., a storage technology) and N risky assets that pay cash flows D_i in period 1, for $i = 1, \ldots, N$. The per-share cash flows of the risky assets are given by

$$D_i = \mu_i + \beta_{1,i} Z_1 + \beta_{2,i} Z_2 + \epsilon_i, \quad \beta_{m,i} \geq 0, \ m = 1, 2, \ i = 1, \ldots, N, \tag{1}$$

where $Z_m \sim \mathcal{N}(0, \sigma^2_{Z_m})$, $m = 1, 2$, are common shocks (factors) and $\epsilon_i \sim \mathcal{N}(0, \sigma^2_\epsilon)$ are idiosyncratic shocks. All shocks are mutually orthogonal. Let $D = (D_1, \ldots, D_N)'$ denote the vector of period-1 cash flows. We denote the variance-covariance matrix of the cash flows by $\Sigma = \text{Var}(D)$, the vector of expected cash flows by $\mu = (\mu_1, \ldots, \mu_N)'$, the factor-loading vectors by $\boldsymbol{\beta}_m = (\beta_{m,1}, \ldots, \beta_{m,N})'$, and period-0 prices by $p = (p_1, \ldots, p_N)'$. The risky assets are in fixed supply of $\overline{\theta} \equiv (\overline{\theta}_1, \ldots, \overline{\theta}_N)'$ shares, which we interpret as the total supply of each asset net of the holdings of buy-and-hold investors (e.g., pension funds and insurance companies in the case of corporate bonds).

There are two types of investors: direct investors (mass λ_D) and fund managers (mass

λ_F). Direct investors, who can also be interpreted as arbitrageurs, manage their own portfolios. In contrast, fund managers manage portfolios on behalf of fund investors. Fund investors can invest directly in the riskless asset but do not trade risky assets themselves; instead, they delegate portfolio selection to fund managers, who are compensated for this service. A manager's performance is evaluated relative to the benchmark index. The benchmark index is a portfolio of $\omega \equiv (\omega_1, \ldots, \omega_N)'$ shares of the assets described above. For realism, we constrain the components of ω to be nonnegative; some may be zero. The assets are ordered by their benchmark weight ω_i, with the first asset having the highest weight. All investors have constant absolute risk aversion (CARA) utility over terminal wealth (or compensation), given by $U(W) = -\exp(-\gamma W)$, where γ is the absolute risk aversion coefficient.

The terminal wealth of a direct investor is given by $W = W_0 + \theta_D'(D - p)$, where the $N \times 1$ vector θ_D denotes the number of shares held by the direct investor, and W_0 is the investor's initial wealth. The direct investor chooses the portfolio θ_D to maximize their utility $U(W)$. The assets under management delegated to the fund manager are denoted by W_F. The fund manager chooses both which assets to include in the portfolio and how many shares of each included asset to hold; the resulting vector of holdings is denoted by θ. She maximizes her utility $U(w)$, where w is her compensation. The fund manager's compensation w consists of three components: the first is a linear payout based on the fund's absolute performance, the second depends on the fund's performance relative to a benchmark portfolio, and the third component (c) is independent of performance. Specifically,

$$w = aR + b(R - B) + c - Cn, \quad a \geq 0, \quad b > 0, \quad n \leq N, \tag{2}$$

where $R \equiv \theta'(D - p)$ represents the fund portfolio's absolute performance, and $B \equiv \omega'(D - p)$ denotes the benchmark portfolio's performance. Let n denote the number of assets included in the fund's portfolio. We assume that managing each included asset entails a fixed cost C, so total portfolio management costs are Cn. We interpret these costs as encompassing transaction and rebalancing expenses, monitoring costs, and other portfolio management

overhead. Larger portfolios naturally entail higher costs. These costs need not arise immediately; they may reflect, for instance, expected future trading or rebalancing expenditures. For tractability, we model them as fixed per included asset and independent of the number of shares traded. Finally, the parameters a and b capture the manager's compensation sensitivity to absolute and relative portfolio performance, respectively.

The direct investor's optimization problem is standard. The optimization problem of a fund manager can be represented as $\max_\theta E[-\exp\{-\gamma((a+b)\theta'(D-p)-b\omega'(D-p)-Cn+c)\}]$, which is equivalent to the following mean-tracking-error-variance optimization problem:

$$\max_{\theta} \ (a+b)\theta'(\mu-p) - \frac{1}{2}\gamma \underbrace{((a+b)\theta - b\omega)'\Sigma((a+b)\theta - b\omega)}_{\text{Tracking-error variance}} - Cn. \tag{3}$$

Formally, the manager's choice of θ also determines the set of assets with nonzero positions, and hence the value of n. We present details of the optimization problems in Appendix B.1; the proof of Proposition 1 writes this support choice explicitly.

3.1 Equilibrium with Uncorrelated Cash Flows

We begin by solving a simplified version of the model without systematic risk, i.e., $\beta_{m,i} = 0$. To sharpen our results in this section, we assume that the benchmark weights ω are proportional to the asset supply $\overline{\theta}$—a realistic assumption given that bond indexes are typically value-weighted. To account for fund size, we normalize the weights in the fund's benchmark as $\omega = W_F^{-1}\overline{\theta}$.

Equilibrium prices in this economy are derived from the market-clearing condition $\lambda_D\theta_D + \lambda_F\theta = \overline{\theta}$. The following proposition reports the portfolios of direct investors and fund managers and equilibrium asset prices.

Proposition 1. *The portfolios of a direct investor and of a fund manager are given by*

$$\theta_{D,i} = \frac{1}{\gamma\sigma_\epsilon^2}(\mu_i - p_i), \tag{4}$$

$$\theta_i = \begin{cases} \frac{1}{\gamma\sigma_\epsilon^2(a+b)}\left(\mu_i - p_i\right) + \frac{b}{a+b}\omega_i, & \text{if } i \leq n^* \quad \textit{(included by managers)} \\ 0, & \text{if } n^* < i \leq N \quad \textit{(excluded by managers)}, \end{cases} \tag{5}$$

and the equilibrium asset prices are

$$p_i = \begin{cases} \mu_i - \gamma\frac{1}{\lambda_D + \frac{\lambda_F}{a+b}}\sigma_\epsilon^2\left(\overline{\theta}_i - \frac{b}{a+b}\underbrace{\lambda_F\omega_i}_{BMI_i}\right), & i \leq n^* \quad \textit{(included by managers)} \\ \mu_i - \gamma\frac{1}{\lambda_D}\sigma_\epsilon^2\overline{\theta}_i, & n^* < i \leq N \quad \textit{(excluded by managers)}, \end{cases} \tag{6}$$

where $BMI_i \equiv \lambda_F\omega_i$ denotes the benchmarking intensity of asset i.

The optimal number of assets in the fund manager's portfolio in a pure cutoff equilibrium, n^*, satisfies the following inequalities:

$$\frac{\gamma}{2}\sigma_\epsilon^2\left(\frac{\overline{\theta}_{n^*} + b\lambda_D\omega_{n^*}}{\lambda_D + \frac{\lambda_F}{a+b}}\right)^2 - C \geq 0, \tag{7}$$

$$\frac{\gamma}{2}\sigma_\epsilon^2\left(\frac{\overline{\theta}_{n^*+1}}{\lambda_D} + b\,\omega_{n^*+1}\right)^2 - C < 0. \tag{8}$$

At the boundary, the first inequality is omitted if $n^* = 0$, and the second inequality is omitted if $n^* = N$.

To develop intuition for the structure of equilibrium reported in Proposition 1, it is useful to consider a special case without portfolio management costs. In that special case, no assets are excluded by fund managers, i.e., $n^* = N$. Direct investors hold standard mean-variance portfolios. Fund managers' portfolios consist of a combination of the mean-variance portfolio and an additional portfolio that hedges the managers against their underperformance relative to the benchmark. That hedging portfolio is the benchmark index, and, as evident from (5), the demand for this portfolio is inelastic, i.e., it does not depend on asset

prices and depends only on the strength of the absolute and relative performance sensitivities of the managers' pay.[4] Equilibrium asset prices reflect the presence of both direct investors and fund managers. The presence of the latter gives rise to a term $b/(a+b)\lambda_F\omega_i$ in asset prices, which reflects the aggregate inelastic demand from fund managers. It implies that the higher the mass of asset managers (λ_F) and the higher the weight of an asset in the benchmark ω_i, the higher the price of the asset. Pavlova and Sikorskaya (2023) generalize these insights to a setting with heterogeneous benchmarks, in which case the aggregate inelastic demand of fund managers is proportional to $\sum_j \lambda_{F,j}\omega_{i,j}$, where j indexes benchmarks and $\lambda_{F,j}$ represents the mass of managers evaluated relative to benchmark j. They term this quantity Benchmarking Intensity (BMI), and we adopt the same terminology in this paper.

We now turn to the case with portfolio management costs, $C > 0$. In this case, it may no longer be optimal for fund managers to include all available assets in their portfolios, as each additional asset contributes to the portfolio management costs. Proposition 1 reveals that the managers hold *sparse* portfolios that exclude all assets with a rank above n^*. Recall from our assumption earlier in this section that these assets are the ones with the lowest benchmark weight. While excluding any assets from the portfolio increases its tracking error, assets with the lowest benchmark weight have the smallest effect on it.

The cutoff conditions have a simple interpretation. At equilibrium prices, the last asset held by fund managers must be worth keeping, while the next asset must not be worth adding. This is exactly what (7)–(8) require. The two inequalities differ because asset n^* is priced by both direct investors and fund managers, whereas asset n^*+1 is priced only by direct investors. In both inequalities, the first term is the marginal benefit of including the asset and the last term is the per-asset management cost C.

As (7)–(8) reveal, the marginal benefit depends positively on three asset characteristics: the asset's (i) weight in the benchmark ω_i, (ii) size $\overline{\theta}_i$, and (iii) idiosyncratic risk σ_ϵ.[5] We have already discussed that assets with higher benchmark weights contribute more

[4] This observation was first made in Kashyap, Kovrijnykh, Li, and Pavlova (2021).

[5] Note that these results are comparative statics: the likelihood of inclusion depends on the interaction of idiosyncratic risk, σ_ϵ, with benchmark weight, ω_i, and asset size, $\overline{\theta}_i$. In the model, idiosyncratic risk is independent of both benchmark weight and asset size. In the data, however, measures of idiosyncratic risk

to reducing the fund's tracking error. Idiosyncratic risk acts in a similar way: if an asset has high idiosyncratic risk, including it in the portfolio reduces the fund's tracking error.[6] Furthermore, assets of larger size are more likely to be included in the portfolio because they have higher expected returns. In equilibrium, the additional supply of a larger asset requires a price discount to clear the market, implying a higher expected return; see equation (6).

Our model also contains predictions regarding the environments or asset classes in which one would expect sparser portfolios. Equations (7)–(8) imply that the number of assets included in a sparse portfolio increases with the strength of relative performance incentives, b. (From (5), a higher b implies that such managers track the benchmark more closely.) Stronger benchmarking concerns raise the marginal benefit of including additional assets by increasing the importance of tracking error considerations. A testable implication is therefore that portfolio sparsity should be lower for funds with stronger relative performance incentives—such as passive funds—than for active funds. This prediction is strongly supported by the evidence presented in Section 2.

In addition, equations (7)–(8) predict greater portfolio sparsity (i) in asset classes with lower idiosyncratic risk and (ii) for funds benchmarked to broader indexes, where individual assets carry smaller benchmark weights. In both cases, individual assets have a limited impact on tracking error, reducing the marginal benefit of including them in the portfolio. Finally, interpreting the cost parameter C as capturing rebalancing or trading costs, the model predicts sparser portfolios in less liquid markets. Taken together, these implications suggest that corporate bond funds should hold sparser portfolios than equity funds—a prediction borne out in the data.

Another relevant comparison is between investment-grade and high-yield bond funds. High-yield bonds have higher idiosyncratic risk (issuer-specific risk) and are tracked by narrower benchmark indexes. Both forces raise the marginal benefit of adding another bond to

are typically negatively related to these variables. Moreover, the former are positively related to measures of illiquidity, which in our model may be captured by C. These relationships may therefore obscure the comparative-static effect of idiosyncratic risk in the data.

[6] Although we assume for parsimony that all assets share the same idiosyncratic risk, equations (7)–(8) remain valid even without this assumption. One simply needs to replace σ_ϵ with σ_{ϵ_i}, where the latter denotes asset-specific idiosyncratic risk.

the portfolio and therefore increase the optimal number of bonds held. The model therefore predicts that high-yield bond funds hold less sparse portfolios than investment-grade corporate bond funds, despite the lower liquidity of high-yield bonds.

We next discuss asset prices in the presence of portfolio sparsity. Since funds optimally assign zero weight to assets at the bottom of the benchmark, their prices are determined solely by direct investors. In contrast, the prices of the top n^* assets are determined by both direct investors and fund managers. This is the key intuition behind equation (6). The pricing difference between the top n^* assets and the remaining ones parallels that in Merton (1987), where certain groups of investors (exogenously) are unaware of and do not hold subsets of assets.

Finally, we note that the discrete choice of the number of included assets may introduce an additional mixed equilibrium with partial participation of fund managers in asset $n^* + 1$. We discuss it in Appendix B.3.

We conclude this section by presenting several testable predictions of our model for the special case of uncorrelated cash flows.

Testable Prediction 1: Fund managers hold sparse portfolios. Portfolios of funds with weaker relative performance concerns are sparser.

Testable Prediction 2: Fund managers are more likely to include assets with (i) higher benchmark weights, (ii) larger size, (iii) higher idiosyncratic risk.

Testable Prediction 3: Assets included by fund managers have higher prices and lower expected returns than excluded assets.

Testable Prediction 4: The index inclusion effect occurs only for assets that managers optimally include in their sparse portfolios.

3.2 Equilibrium with Correlated Cash Flows

While the uncorrelated cash flows specification offers analytical tractability, it rules out asset substitutability—a key channel for our application. We therefore introduce common shocks (factors) to the cash flows in (1) by assuming $\beta_{m,i} > 0$.

Solving the N-asset version of the model is a combinatorial problem. To highlight the three main insights from our model related to asset substitutability, it is not essential to present the full solution to the problem. Instead, we focus on characterizing the fund manager's decision to include an additional asset in a sparse portfolio.

We begin with a two-asset version of our model with one common shock (risk factor). For convenience, we adopt the following notation for the variance–covariance matrix Σ:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}, \tag{9}$$

where $\sigma_i^2 \equiv \mathrm{Var}(D_i)$ is the (unconditional) cash-flow variance of asset i, and $\rho_{12} \equiv \mathrm{Corr}(D_1, D_2)$.

A manager facing no portfolio management costs optimally holds both assets. In an economy with portfolio management costs, the cost of including both assets in the portfolio may be prohibitively high. We focus on the interesting and realistic scenario in which the cost is sufficiently high and portfolios are sparse. For illustration, we consider the case in which the manager opts for Asset 1 rather than Asset 2. Her (sparse) optimal portfolio is presented in the following lemma.

Lemma 1. *If portfolio management costs are sufficiently high so that the manager includes only one asset, Asset 1, in her portfolio, her optimal portfolio is*

$$\theta_i = \begin{cases} \frac{\mu_1 - p_1}{\gamma\sigma_1^2(a+b)} + \frac{b}{a+b}\omega_1 + \underbrace{\frac{b}{a+b}\frac{\rho_{12}\sigma_1\sigma_2}{\sigma_1^2}\omega_2}_{\textit{Additional demand}}, & \textit{if } i = 1, \\ 0, & \textit{if } i = 2. \end{cases} \tag{10}$$

The optimal portfolio of a fund manager who includes both assets is given by

$$\theta = \frac{1}{\gamma(a+b)}\Sigma^{-1}(\mu - p) + \frac{b}{a+b}\omega \tag{11}$$

Lemma 1 shows that if—due to portfolio management costs—the manager opts to include only Asset 1, she holds *more* of it than she would in the economy without portfolio management costs. With correlated assets, Asset 1 acts as a substitute for Asset 2, so by overweighting Asset 1, the manager gains exposure to Asset 2's common risk factor. This effect disappears when cash flows are uncorrelated: setting $\rho_{12} = 0$ eliminates it.

As in the uncorrelated cash flows case considered in Section 3.1, the prices of assets not included in managers' (sparse) portfolios are lower than in the economy without portfolio management costs. The following lemma reports asset prices in both economies.

Lemma 2. *If portfolio management costs are sufficiently high so that the manager includes only one asset, Asset 1, in her portfolio, asset prices are given by*

$$p_1 = \mu_1 - \gamma A \left[\sigma_1^2\,\overline{\theta}_1 + \rho_{12}\sigma_1\sigma_2\,\overline{\theta}_2 - \frac{b}{a+b}\left(\sigma_1^2\,\underbrace{\lambda_F\omega_1}_{BMI_1} + \rho_{12}\sigma_1\sigma_2\,\lambda_F\omega_2\right)\right], \tag{12}$$

$$p_2 = \mu_2 - \gamma A \left[\rho_{12}\sigma_1\sigma_2\,\overline{\theta}_1 + \rho_{12}^2\sigma_2^2\,\overline{\theta}_2 - \frac{b}{a+b}\left(\rho_{12}\sigma_1\sigma_2\,\lambda_F\omega_1 + \rho_{12}^2\sigma_2^2\,\underbrace{\lambda_F\omega_2}_{BMI_2}\right)\right] - \frac{\gamma}{\lambda_D}\,\sigma_2^2(1-\rho_{12}^2)\,\overline{\theta}_2, \tag{13}$$

where $A \equiv \left[\lambda_D + \frac{\lambda_F}{a+b}\right]^{-1}$ *and* $BMI_i \equiv \lambda_F\omega_i$ *denotes the benchmarking intensity of asset* i.

In the economy without portfolio management costs, where asset managers include both assets, the price of Asset 1 is the same as in (12)*. The price of Asset 2 is given by*

$$p_2 = \mu_2 - \gamma A\left[\rho_{12}\sigma_1\sigma_2\,\overline{\theta}_1 + \sigma_2^2\,\overline{\theta}_2 - \frac{b}{a+b}\left(\rho_{12}\sigma_1\sigma_2\,\lambda_F\omega_1 + \sigma_2^2\,\lambda_F\omega_2\right)\right].$$

By examining the expressions for p_2 in both economies, it is easy to see that Asset 2's price is lower in the economy with portfolio management costs, unless $\rho_{12} = 1$. Since the

managers exclude Asset 2, the demand for it is lower and so is its price.

In contrast to the economy of Section 3.1, in the economy with correlated cash flows, the price of Asset 2 depends on its BMI. The intuition is that the manager can reduce the tracking-error variance of her portfolio by substituting Asset 2 with Asset 1. This channel is reflected in the last term of (10), which depends on Asset 2's benchmark weight ω_2.

However, assets included by managers and excluded assets differ in their price sensitivity to BMI. Lemma 2 shows that Asset 2's price is less sensitive to BMI than Asset 1's price, with the sensitivity scaled by a factor of ρ_{12}^2. Thus, asset substitutability attenuates the transmission of benchmark-induced demand: the lower the correlation between Assets 1 and 2, the weaker the price response of the excluded asset. In the limit, when cash flows are uncorrelated ($\rho_{12} = 0$), this sensitivity vanishes.

We now examine the manager's trade-off involved in adding Asset 2 to her portfolio. For any subset of assets $\mathsf{I} \subset \{1, 2\}$, let $V(\mathsf{I})$ denote the fund manager's value function: the maximized value of the objective in (3) when she is restricted to include only assets in I, i.e., when $\theta_i = 0$ for all $i \notin \mathsf{I}$. In particular, $V(\{1\})$ is the value when the manager can trade only Asset 1, and $V(\{1, 2\})$ is the value when she can trade both assets. The marginal benefits and costs of adding Asset 2 can be assessed by comparing the two value functions. The following lemma reports the outcome of this comparison.

Lemma 3. *Let $V(\mathsf{I})$ denote the manager's value function when the admissible set of assets is $\mathsf{I} \subset \{1, 2\}$. The incremental value of adding Asset 2 to a portfolio that already contains Asset 1 is*

$$V(\{1, 2\}) - V(\{1\}) = (1 - \rho_{12}^2)\, \frac{\gamma}{2}\, \sigma_2^2 \left(\frac{\overline{\theta}_2}{\lambda_D} + b\omega_2 \right)^2 \; - \; C.$$

The expression for the costs and benefits of adding Asset 2 to the manager's portfolio presented in Lemma 3 parallels that in equation (8) for the uncorrelated cash-flow case. Specifically, holding fixed its correlation with Asset 1, the marginal benefit of including Asset 2 is higher if the asset has (i) a higher benchmark weight, (ii) larger size, and (iii) higher idiosyncratic risk. The marginal cost of including Asset 2 is C. The same logic

extends naturally to asset-specific portfolio-management costs. If the manager pays C_i for including asset i, so that the cost term is $\sum_{i \in \mathsf{I}} C_i$, then the net gain from adding Asset 2 is obtained by replacing C with C_2 in Lemma 3 (Appendix B.5). This formulation is useful if C_i is interpreted as an asset-specific trading or rebalancing cost that captures liquidity differences across assets: more liquid assets have lower C_i and, therefore, a lower marginal cost of inclusion in a portfolio.

A distinct implication of the correlated cash-flow case is the role of asset substitutability, as measured by the correlation ρ_{12}. The multiplicative term $1 - \rho_{12}^2$ in Lemma 3 shows that the more substitutable Assets 1 and 2 are, the lower the marginal benefit of adding Asset 2 to the manager's portfolio. This is intuitive because the manager can substitute for holding Asset 2 by holding more of Asset 1, thereby saving on portfolio management costs while still obtaining similar exposure to expected returns and tracking-error variance reduction. In the limiting case as $\rho_{12} \to 1$, Asset 1 becomes a perfect substitute for Asset 2 and the marginal benefit of adding Asset 2 vanishes.

Note that substitutability and the asset's own characteristics enter the marginal benefit multiplicatively. Outside the knife-edge case of perfect substitutes ($\rho_{12}^2 = 1$), in which the benefit is zero and Asset 2 is never added whatever its other characteristics, benchmark weight, size, and idiosyncratic risk govern the inclusion decision at every degree of substitutability. Substitutability scales the benefit of adding an asset; it does not displace the other determinants of inclusion.

For illustrative purposes, we have assumed in this section that fund managers first include Asset 1 in their portfolios and then Asset 2, provided that the benefits of doing so exceed the portfolio management costs. The case of managers first including Asset 2 and then Asset 1 is analyzed analogously. Solving the problem fully requires the comparison of the value functions in each of these cases and establishing which combination of assets yields the highest value function. Since our goal here is to illustrate the trade-off facing the manager contemplating the inclusion of an additional asset, we do not present the full solution.

In practice, fund managers sample bonds from distinct categories, for example, credit

rating or maturity buckets. To model this choice, we consider a 4-asset version of our model with two common shocks. For expositional simplicity, we assume that the two common shocks are orthogonal to each other (this can be achieved by appropriately redefining the shocks). Under this assumption, the variance-covariance matrix of the cash flows has a block-diagonal form, representing two distinct buckets. (We use the term "bucket" to refer to a set of assets whose cash flows load on the same common shock, so that correlations are high within the set and negligible across sets.)

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & 0 & 0 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & \rho_{34}\sigma_3\sigma_4 \\ 0 & 0 & \rho_{34}\sigma_3\sigma_4 & \sigma_4^2 \end{pmatrix}, \tag{14}$$

where $\rho_{12} \equiv \text{Corr}(D_1, D_2)$ and $\rho_{34} \equiv \text{Corr}(D_3, D_4)$. Of specific interest to us is the symmetric case in which the two blocks are the same. For notational convenience, we order assets so that indices 1–2 and 3–4 correspond to two distinct blocks.

The intuition from the two-asset case extends directly. Fund managers prefer to add assets that are less correlated with their existing holdings. For example, suppose the portfolio currently contains Asset 1. With a block-diagonal covariance structure—payoffs correlated within a block but uncorrelated across blocks—the marginal benefit of adding Asset 2 (same block) is smaller, *ceteris paribus*, than the benefit of adding Asset 3 or Asset 4 (a different block).[7] This structure mirrors fixed-income practice, where managers first allocate across broad risk buckets (e.g., maturity or rating) and then select a small number of representative securities within each bucket.

One may think of ρ as a property of the bucket rather than of the individual asset, and so it cannot rank assets within a bucket. Substitutability therefore governs how sparsely a manager samples a given bucket, while benchmark weight, size, and idiosyncratic risk

[7] A full characterization of optimal portfolios would require comparing all feasible subsets across blocks, which is combinatorial in the number of assets and adds little economic insight for our application.

govern which assets within it she holds. Our empirical classification in Section 4 will follow this idea.

The analysis in this section generates several additional testable predictions of our model.

Testable Prediction 5: Managers select within "buckets" of substitutable assets.

Testable Prediction 6: Within each bucket, managers are more likely to include assets with (i) higher benchmark weights, (ii) larger size, (iii) higher idiosyncratic risk.

Testable Prediction 7: Prices are more sensitive to BMI for included assets than for excluded ones.

4 BMI in Bond Markets: Empirical Analysis

In this section, we construct a measure of benchmarking intensity using data on Canadian and U.S. fixed-income mutual funds and ETFs and their benchmark indexes. We then exploit discontinuous changes in this measure around maturity-based index eligibility cutoffs to test the model's predictions in the U.S. and Canadian corporate bond markets.

4.1 Empirical measure of benchmarking intensity

Guided by the model, we calculate the benchmarking intensity (BMI) for bond i in month t as

$$BMI_{i,t} = \frac{\sum_{j=1}^{J} \lambda_{j,t} \times \omega_{i,j,t}}{MV_{i,t}}, \tag{15}$$

where $\lambda_{j,t}$ is the total AUM of mutual funds and ETFs benchmarked to index j in month t, $\omega_{i,j,t}$ is the weight of bond i in index j in month t, and $MV_{i,t}$ is the market value of bond i in month t. The definition therefore matches the one first introduced for equities (Pavlova and Sikorskaya, 2023).

This measure captures the amount of benchmarked capital effectively allocated to a bond and reflects heterogeneity in benchmark index coverage. Economically, BMI mea-

sures the fraction of a bond's outstanding market value that would be held by benchmark-constrained capital under full index replication. Tables A3 and A4 report descriptive statistics in the full sample of bonds for the U.S. and Canada. The average total BMI in the U.S. is over 17%, while in Canada it is over 9%. These statistics mask considerable heterogeneity by bond maturity, as we show below.

Figure C1 in the Appendix illustrates the evolution of BMI over a bond's lifetime using two example bonds, IBM in the U.S. and ENBWE (Westcoast Energy) in Canada. In both geographies, the broadest indexes contribute the most to the BMIs of these bonds. Yet, a significant fraction of BMI stems from capital benchmarked to maturity-based indexes. In particular, when the IBM bond enters the short-term investment universe, around 10% of its value is benchmarked by Aggregate index investors while almost 15% of it is benchmarked by mid and short-term index investors.

Because indexes in our sample are value-weighted and do not use float factors, equation (15) simplifies mechanically. Substituting the definition of index weights and canceling the bond's own market value yields

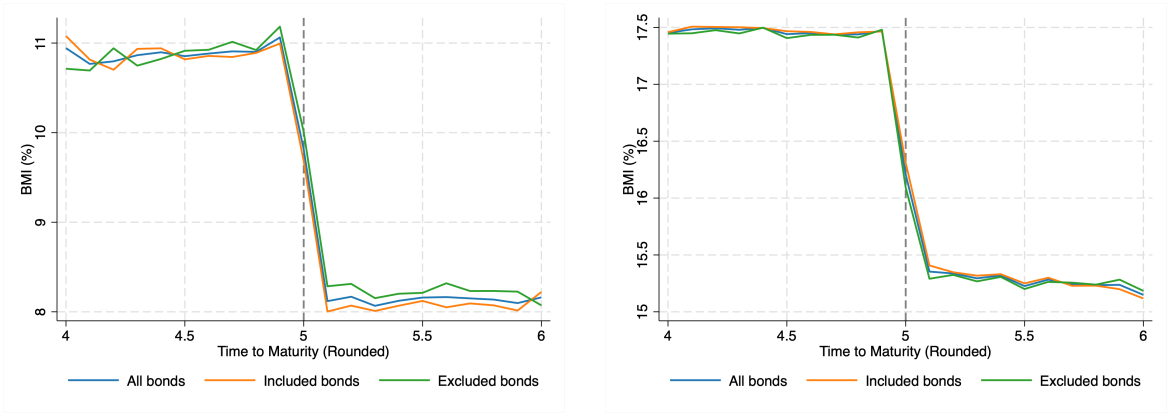
$$BMI_{i,t} = \sum_{j} \frac{\lambda_{j,t}}{IndexMV_{j,t}} M_{i,j,t},$$

where $IndexMV_{j,t}$ is the market value of index j, $\lambda_{j,t}$ is benchmarked AUM as defined around (15), and $M_{i,j,t} = 1$ if bond i belongs to index j in month t. Thus, for a given bond, variation in BMI over time arises from two sources: (i) changes in the ratio $\lambda/IndexMV$, which capture how much fund capital is benchmarked to each index relative to its market value, and (ii) changes in the bond's index membership, such as maturity-driven transitions between indexes. We explore such transitions in the following section.

4.2 Maturity-based cutoffs in fixed-income benchmarks

A key empirical challenge is that BMI is endogenous. Bonds with higher BMI tend to be larger, more liquid, or higher quality, and these characteristics may independently affect

Figure 3: Discontinuity in BMI around the 5-year maturity cutoff



(a) Canada

(b) U.S.

This figure illustrates the average benchmarking intensity by a bond's time to maturity around the 5-year mark in (a) Canada and (b) the U.S. We report the average BMIs across all bonds in our baseline sample as well as within bonds predicted to be included or excluded from fund portfolios, as defined later in the section.

ownership and prices. As a result, cross-sectional regressions cannot isolate the causal effect of benchmarking intensity. To address this concern, we exploit the mechanical changes in benchmark index membership that occur at the common bond maturity cutoff—5 years to maturity—in both Canada and the U.S. When a bond crosses this maturity threshold, its index eligibility changes mechanically, altering its inclusion in maturity-segmented benchmark indexes. Because benchmarked AUM differs substantially across these indexes, index membership changes generate discontinuous shifts in BMI that are unrelated to contemporaneous changes in bond fundamentals. This variation allows us to exploit plausibly exogenous shifts in BMI to test the model predictions.

In both Canada and the U.S., the benchmarked AUM in the mutual fund and ETF sector varies strongly by bond maturity. Figure 3 shows that this variation in assets generates discontinuities in BMI around the 5-year maturity cutoff: an average bond sees a 1.9–2.1 p.p. increase in BMI as it crosses the cutoff. These discontinuities are sizeable and highly statistically significant. In Appendix C.2, we formally estimate their size by regressing the changes in BMI on a 5-year switch dummy variable.

Our design develops upon two papers. Within corporate bonds, an experiment ex-

ploiting the maturity-based cutoffs for identification was first proposed by Bretscher, Schmid, and Ye (2024), who study the effects of passive ownership on bond prices in the U.S. market. Methodologically, our use of within-asset variation in BMI as the treatment intensity follows Pavlova and Sikorskaya (2023), who introduce BMI for equities and validate this identification approach using index reconstitutions and the discontinuity at the Russell 1000/2000 cutoff. We combine the two: the maturity cutoffs supply the quasi-exogenous shift in index membership, and BMI translates it into a theory-implied measure of the demand shock intensity. We use this design to test the prediction of our model that an increase in BMI lowers bond yields and increases benchmarked fund ownership. In particular, in the monthly panel of corporate bonds around a maturity cutoff, we estimate

$$\Delta Y_{i,t+h} = \boldsymbol{\beta} \Delta BMI_{i,t} + \zeta' \bar{X}_{i,t} + \nu_i + \mu_t + \epsilon_{i,t+h}. \tag{16}$$

The dependent variable $\Delta Y_{i,t+h}$ denotes either (i) the change in the bond's yield spread or total return between t and $t+h$, or (ii) the change in fund ownership over the same horizon. We consider horizons from one month to one year. $\Delta BMI_{i,t}$ is the change in BMI from month t to $t+1$.[8] Bond-level controls in $\bar{X}_{i,t}$ include the logarithm of the bond's par value, numeric credit rating, bid-ask spread quartile, time to maturity, and duration, all measured at t. For pricing variables, we primarily rely on index provider data. Accordingly, in the U.S., we use total return and option-adjusted duration and spread from Bloomberg, while in Canada, we use modified duration and yield from FTSE and construct bond returns from the LSEG (Refinitiv) total return index. ν_i and μ_t are bond and year-month fixed effects. We estimate this specification only in the neighborhood of the 5-year cutoff, namely, for bonds with between 4 and 6 years to maturity.[9] Finally, we only consider bonds with liquidity data

[8] Bloomberg bond indexes are rebalanced at the start of the month, while FTSE indexes are rebalanced daily. We cumulate these daily changes into a monthly measure. In both geographies, the timing of BMI change in our tests is aligned with when index trackers would have traded.

[9] We do not study the 10-year cutoff because insurance companies and other long-horizon investors often reduce holdings of long-term bonds for portfolio management or regulatory reasons (see Chaudhary, 2024; Li, 2023), which could confound the interpretation of our findings. There is also a 3-year cutoff available for the U.S. indexes, at which many bonds are already close to maturity and often held to maturity by institutional investors, making this cutoff less suitable for testing our theory.

in a given month.[10]

We report wild cluster bootstrap inference with clustering by issuer and year-month. This choice is motivated by the finite-sample performance of asymptotic clustered standard errors in our setting, where the number of effective clusters can be limited in the narrow maturity window and included/excluded subsamples, so the cluster sizes are highly unbalanced. The resulting confidence intervals are virtually the same as with asymptotic clustered standard errors in cases where those errors can be obtained.

Our identifying assumption is that, conditional on this set of controls and fixed effects, changes in BMI are plausibly exogenous to bond returns and changes in fund ownership. Since changes in bond index membership around maturity cutoffs are predetermined, our design effectively exploits treatment intensity around a sharp treatment cutoff.[11] Such a design requires that no other variables related to bond performance discontinuously change at maturity cutoffs. For this reason, we control for duration and only consider non-callable and make-whole callable bonds in our analysis.[12] Other types of bonds experience sharp drops in their option-adjusted duration at the 5-year maturity cutoff. One potential reason is that the 5-year mark is the common end of the call protection period. Focusing on non-callable and make-whole callable bonds also makes our analysis in the U.S. more comparable with that in Canada, as the majority of callable bonds in Canada are make-whole (Afik, Jacoby, Stangeland, and Wu, 2019).

Our model predicts that fund portfolios are sparse, and hence estimating the effect of benchmarking on bond prices and portfolios is challenging because the average effect of

[10]Specifically, we require a non-missing bid-ask spread, which in the U.S. means the bond traded in both directions during the month and in Canada means a composite dealer quote is available. The requirement binds far more in Canada, where quote coverage is sparse: adding back the omitted bonds raises the sample by 9% in the U.S. but by more than a third in Canada.

[11]Two features of BMI change decomposition support our interpretation of the design. First, within the bond and year–month fixed effects of specification (16), index membership changes account for 96% of the variance of ΔBMI in both the U.S. and Canada, so the identifying variation is driven by membership changes at the cutoffs rather than by contemporaneous fund flows. Second, our estimates are unchanged if we rebuild ΔBMI holding $\lambda_j / IndexMV_j$ fixed at its pre-window level, so that only index membership varies; if anything, the estimates become somewhat larger.

[12]To implement this filter, we use the call type provided by Bloomberg in combination with the first call date provided by LSEG FISD. The filter removes around 10% of bond-month observations.

crossing the cutoff across all bonds may be attenuated toward zero. We address this challenge by estimating specification (16) in subsamples of included and excluded bonds.

Following Testable Predictions 2 and 6, we proxy for a bond's likelihood of inclusion in fund portfolios using its relative size within a bucket of close substitutes. Specifically, we classify a bond as *excluded* if it is in the bottom three deciles by size (par value) within its sector-rating-maturity bucket in all indexes to which it belongs and is considered *included* otherwise.[13] The cutoff at the third decile of size is motivated by the aggregate benchmark-level sparsity in both geographies, with around 70% of bonds included in fund portfolios in value-weighted terms (see Section 2). Following Testable Prediction 6, we approximate a bucket of closely substitutable bonds with bonds within the same sector, rating, and maturity group as defined by the index provider. We also present and discuss the results for alternative cutoffs and bucket definitions below.

4.3 BMI and bond pricing

In this section, we report and discuss the estimation results for pricing variables.

Panel A of Table 3 documents how corporate bond yield spreads and total returns react to BMI changes. Consistent with the model's predictions, increases in BMI are associated with persistent declines in bond yield spreads in both Canada and the U.S.

Importantly, these full-sample effects are attenuated because fund portfolios are sparse. In line with our model, the response of yield spreads to changes in BMI is significantly larger and more persistent in the set of included bonds, as documented in Panels B and C of Table 3. The estimates are plotted in Figure 4 and demonstrate that the spread response of included bonds is around twice as strong at all horizons.

The estimated effects are economically sizeable and statistically detectable, particularly for the included bonds. Upon crossing the 5-year cutoff, BMI on average increases

[13] Although relative issue size is correlated with liquidity, we use it here as a theory-guided predictor of portfolio inclusion, not as a liquidity classification. Appendix C.11 separately shows that fund ownership changes discontinuously at benchmark issue-size eligibility thresholds, which is inconsistent with the interpretation that size matters only through liquidity.

Table 3: Changes in BMI and bond pricing

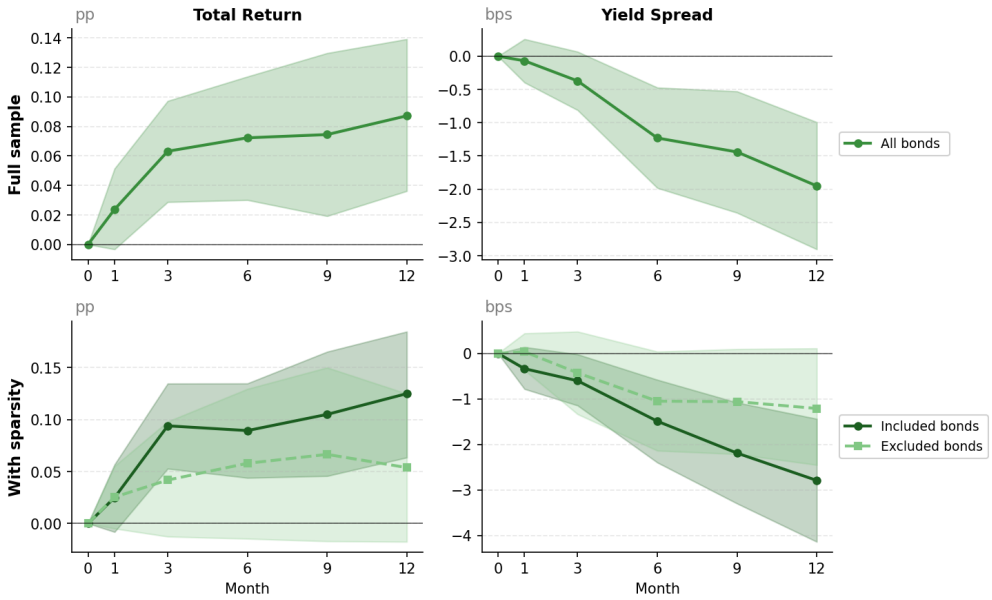
	Change in the pricing variable over horizon									
	U.S.					Canada				
	t+1	t+3	t+6	t+9	t+12	t+1	t+3	t+6	t+9	t+12
Panel A: All bonds										
Total return (%)	0.040***	0.046***	0.035	0.052**	0.047*	0.024	0.063***	0.072***	0.075**	0.087***
	(3.29)	(2.86)	(1.61)	(2.09)	(1.74)	(1.56)	(3.11)	(2.81)	(2.29)	(2.83)
Observations	95,772	93,960	91,069	88,201	85,197	21,786	21,414	20,878	20,328	19,766
Adj. R^2	0.758	0.752	0.814	0.860	0.891	0.760	0.812	0.864	0.887	0.903
Yield spread (bps)	−0.578***	−0.875***	−0.529*	−0.506	−0.355	−0.069	−0.372	−1.226***	−1.442**	−1.947***
	(−3.73)	(−3.66)	(−1.66)	(−1.40)	(−0.90)	(−0.37)	(−1.40)	(−2.68)	(−2.61)	(−3.37)
Observations	95,772	93,961	91,079	88,218	85,219	21,851	21,479	20,943	20,393	19,831
Adj. R^2	0.619	0.660	0.664	0.693	0.710	0.663	0.767	0.780	0.800	0.809
Panel B: Included bonds										
Total return (%)	0.049***	0.063***	0.056**	0.071**	0.074**	0.025	0.094***	0.089***	0.105***	0.125***
	(3.82)	(3.64)	(2.40)	(2.51)	(2.54)	(1.25)	(3.78)	(3.25)	(2.94)	(3.42)
Observations	53,869	52,974	51,507	50,018	48,446	13,276	13,033	12,703	12,365	12,023
Adj. R^2	0.762	0.760	0.814	0.859	0.890	0.764	0.827	0.886	0.907	0.924
Yield spread (bps)	−0.789***	−1.188***	−0.976***	−0.784*	−0.862**	−0.337	−0.602*	−1.489***	−2.189***	−2.788***
	(−3.98)	(−4.30)	(−2.67)	(−1.85)	(−1.83)	(−1.17)	(−1.72)	(−2.72)	(−3.27)	(−3.42)
Observations	53,869	52,975	51,514	50,029	48,457	13,307	13,064	12,734	12,395	12,053
Adj. R^2	0.640	0.673	0.680	0.703	0.716	0.703	0.796	0.807	0.826	0.826
Panel C: Excluded bonds										
Total return (%)	0.031**	0.038**	0.026	0.041	0.017	0.025	0.042	0.058	0.066	0.054
	(2.15)	(2.08)	(1.09)	(1.51)	(0.59)	(1.35)	(1.27)	(1.40)	(1.32)	(1.29)
Observations	41,654	40,737	39,319	37,944	36,523	8,463	8,334	8,130	7,915	7,697
Adj. R^2	0.770	0.765	0.834	0.880	0.909	0.772	0.812	0.848	0.875	0.886
Yield spread (bps)	−0.370**	−0.709**	−0.318	−0.432	0.208	0.036	−0.429	−1.049	−1.059	−1.208
	(−2.10)	(−2.35)	(−0.79)	(−0.98)	(0.44)	(0.15)	(−0.79)	(−1.61)	(−1.53)	(−1.62)
Observations	41,654	40,737	39,322	37,950	36,534	8,498	8,369	8,165	7,951	7,733
Adj. R^2	0.618	0.674	0.684	0.723	0.747	0.626	0.754	0.774	0.798	0.817

This table reports the ΔBMI coefficient estimates from regression (16) in the U.S. and Canadian samples of all bonds (Panel A), included bonds (Panel B), and excluded bonds (Panel C) with time to maturity between 4 and 6 years. The dependent variable is either the total return or the change in yield spread from month t to the stated horizon. All regressions control for the logarithm of the bond's par value, numeric credit rating, bid-ask spread quartile, time to maturity, duration, and include bond and year–month fixed effects. t-statistics based on wild bootstrap with clustering by issuer and month are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$ based on wild-bootstrap p-values.

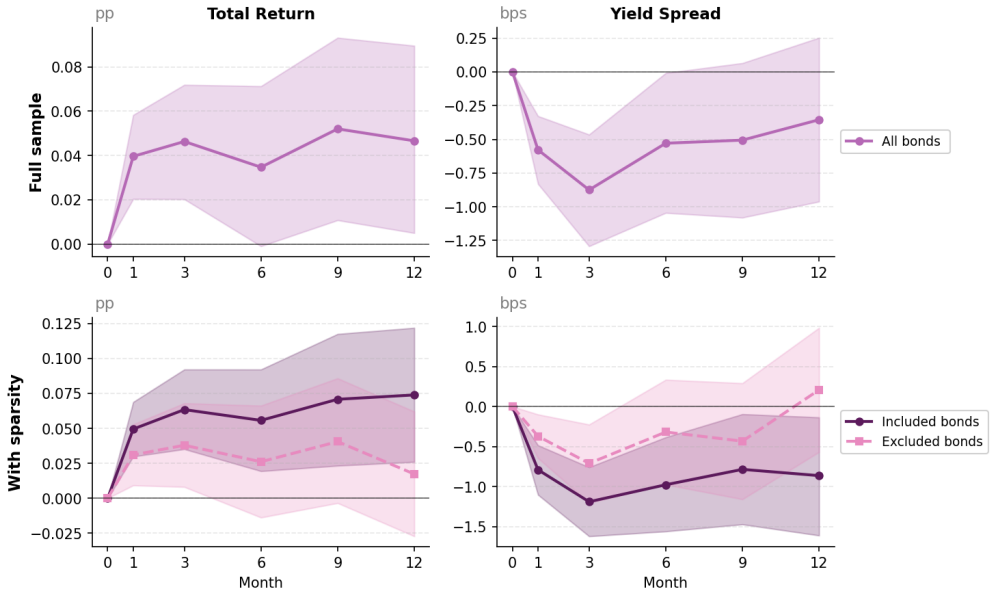
by 2.1 p.p. in the U.S. and 1.9 p.p. in Canada, for both included and excluded bonds. Our estimates imply that this increase in BMI persistently reduces yield spreads of included bonds by approximately 2 bps in the U.S. and 5 bps in Canada.

Given the importance of the sparsity adjustment, we consider alternative size thresholds for inclusion, such as at the first decile or the median of the size distribution, as well as alternative buckets of close substitutes, including bonds with the same rating or same segment-maturity-rating. Across specifications, the change in BMI has smaller predictive power in the full sample, while the included bonds consistently experience larger and more

Figure 4: The estimated effect of BMI on corporate bond prices



(a) Canada



(b) U.S.

This figure plots the estimates and 90% confidence intervals based on Table 3 in (a) Canada and (b) the U.S. A bond is considered *excluded* if it is in the bottom three deciles by size (par value) within its sector-rating-maturity bucket in all indexes to which it belongs and is considered *included* otherwise.

persistent effects. These results are reported in Appendix C.4.

Because we estimate the two subsamples separately, a natural question is whether the

estimates for included and excluded bonds are statistically different. Appendix C.3 reports pooled specifications that interact ΔBMI with an indicator for included bonds and allow group-specific controls and fixed effects, so that the interaction is tested directly. In the U.S., the difference always has the right economic sign and is significant for total returns at $t+1$, $t+3$, and $t+12$ and for yield spreads at $t+1$ and $t+12$. We emphasize that our model does not predict a zero response for excluded bonds. Lemma 2 implies that an excluded bond's price sensitivity to BMI increases with ρ^2, the squared correlation of its payoff with that of its included substitutes, and vanishes only in the limiting case $\rho = 0$. So the model predicts attenuation in ΔBMI sensitivity rather than a null for excluded bonds.

Several features of our setting affect the measured difference between the two groups of bonds. First, our size-based classification is a theory-guided prediction for portfolio inclusion, and the associated classification error likely biases the estimated difference toward zero. Second, the pricing variables are index-provider evaluated prices, which for thinly traded bonds may be inferred in part from comparable bonds in the same bucket. Because those comparables are disproportionately the larger, included bonds, any such interpolation transmits part of the included-bond price response mechanically into excluded-bond prices and shrinks the estimated gap.[14] Third, the Canadian sample is an order of magnitude smaller than the U.S. sample—roughly 8,500 excluded bond-months around the cutoff against about 41,000 in the U.S.—so the Canadian tests have limited power, yet the estimates are directionally consistent.

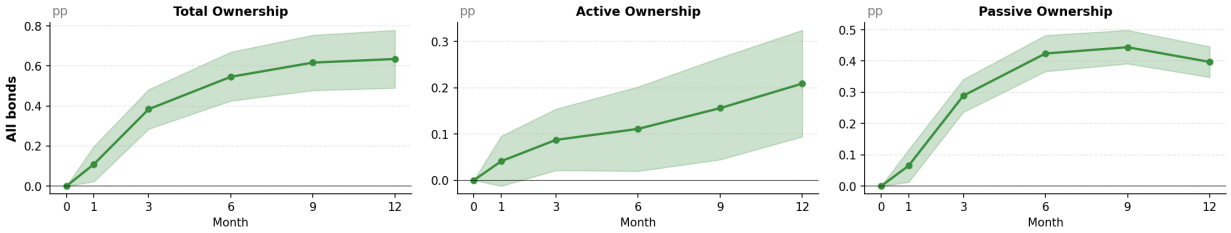
4.4 BMI and fund holdings

According to our model, equilibrium yields are lower for bonds with larger BMIs because funds hedge against underperforming their benchmark indexes. The model predicts that funds, both active and passive, rebalance in line with their mandates and, therefore, their assets contribute to BMI. In this section, to complement the stylized facts in Section 2, we document that active and passive fund ownership changes around maturity cutoffs as

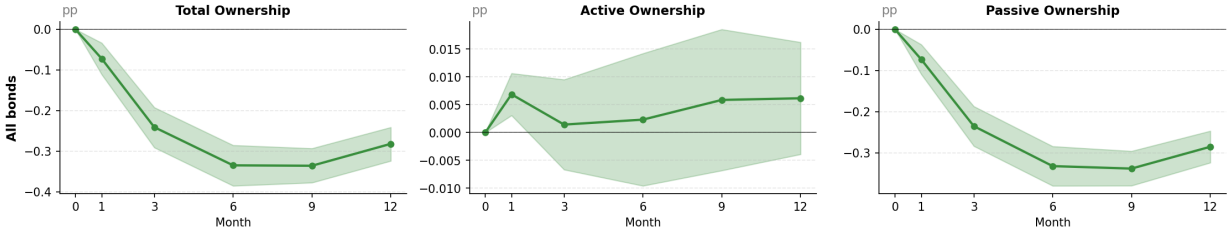
[14] See the discussion in Section 4.6.

predicted by BMI and show that an increase in BMI leads to an increase in both types of ownership in our sample, especially for bonds predicted to enter managers' sparse portfolios.

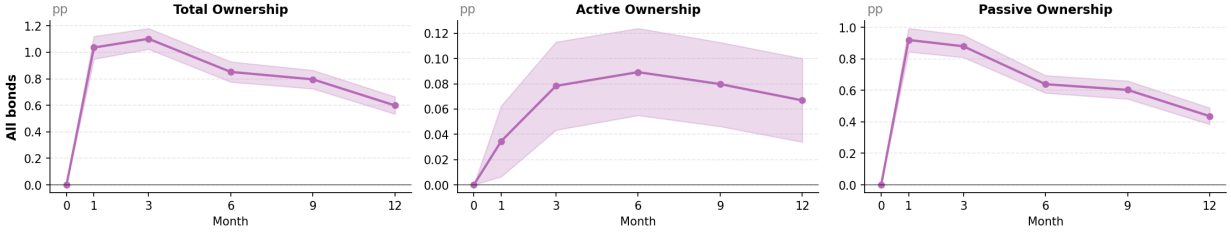
Figure 5: Changes in benchmarked ownership around the 5-year maturity cutoff



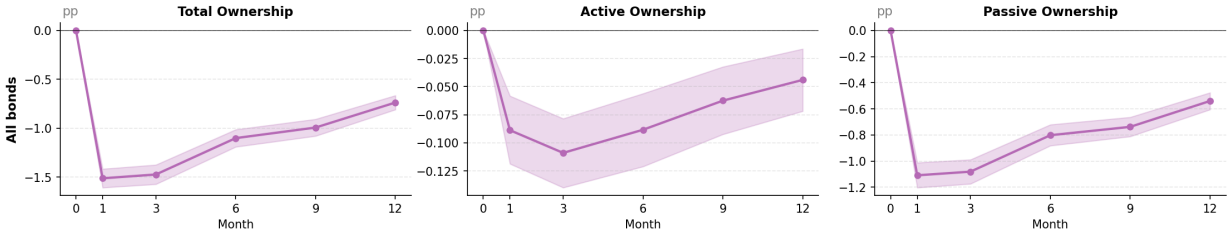
(a) Short-term fund ownership (Canada)



(b) Mid-term fund ownership (Canada)



(c) Short-term fund ownership (U.S.)



(d) Mid-term fund ownership (U.S.)

This figure illustrates how benchmarked fund ownership of corporate bonds changes around the 5-year maturity cutoffs in Canada (panels a and b) and the U.S. (c and d). We report changes in total fund ownership by funds with short-term (1-5 year, panels a and c) and mid-term (5-10 year, panels b and d) benchmark indexes, as well as their active and passive components. All estimates and confidence intervals are from estimating specification (16) around the 5-year cutoff using the 5-year switch dummy as the key independent variable. All y-axes correspond to p.p. changes in ownership relative to one quarter before the switch. We report 90% wild cluster bootstrap confidence intervals with clustering by issuer and month.

First, our granular data allow us to characterize changes in ownership of funds with maturity-based benchmarks around the bond maturity cutoffs. To do so, we estimate a counterpart of regression (16) around the 5-year cutoff in both the U.S. and Canada using benchmarked fund ownership as the dependent variable. We present the estimates in Figure 5. The figure highlights that, when a bond crosses the 5-year cutoff, it is sold by funds benchmarked to the mid-term or 5-10 year indexes, and it is purchased by funds benchmarked to short-term or 1-5 year indexes. Importantly, we see a very similar picture for both active and passive funds in both geographies.[15]

Second, we study how aggregate fund ownership changes with BMI, both for all bonds and in the subsamples of included and excluded bonds. Table 4 presents the results of estimating the regression (16) for total, active and passive fund ownership. Consistent with the model, all funds' ownership significantly rises with an increase in BMI, as reported in Panel A.

We further study ownership changes in the samples of bonds predicted to be included in fund portfolios and excluded from them. The estimates are reported in Panels B and C of Table 4 and, in line with our theoretical predictions, suggest that sparsity has a strong effect on the response of ownership to BMI. Almost all of the responses of ownership are concentrated in the included bonds. We complement the tables by plotting the estimates and confidence intervals in Figure 6.

For active funds in the U.S., the response is driven by the set of included bonds, as rebalancing within excluded bonds is close to zero. For passive funds in the U.S., the magnitude of rebalancing is significantly larger for the included bonds, yet excluded bonds are also traded in the direction predicted by the model. These patterns are consistent with relatively lower aggregate sparsity in passive portfolios in the U.S. discussed in Section 2.

The response of Canadian funds is slightly different. Passive portfolios have a higher

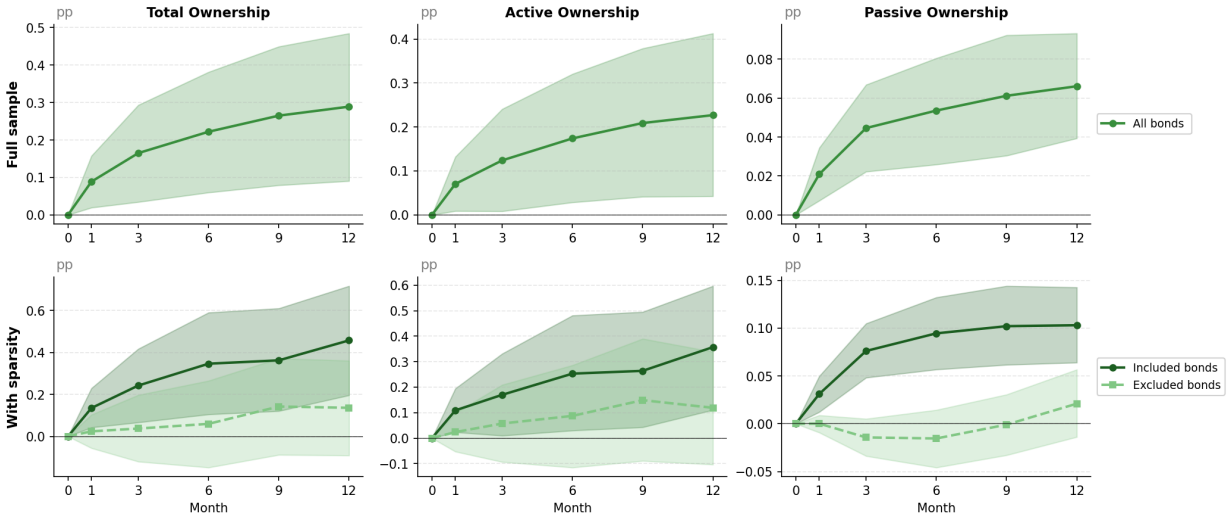
[15] An exception is that mid-term active funds in Canada seem not to necessarily sell bonds crossing the 5-year cutoff, which suggests that, in light of larger transaction costs in Canada, these funds may hold such bonds to maturity. Laipply, Madhavan, Mauro, and Udevbulu (2025) document for U.S. fixed-income funds that managers can mitigate trading costs by allowing index deletions to mature rather than selling immediately.

Table 4: Changes in BMI and bond ownership

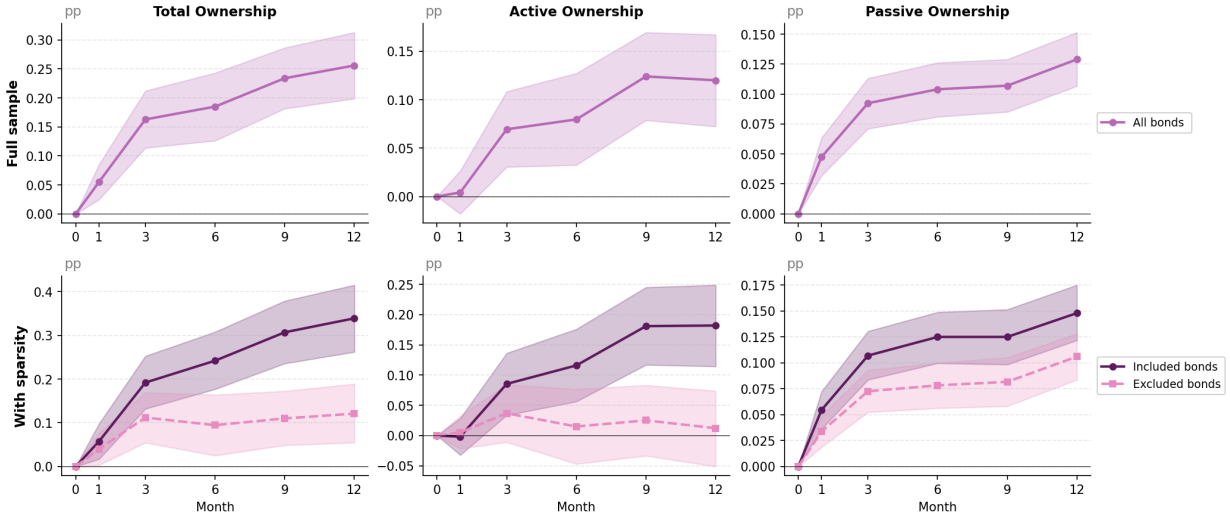
	Change in the ownership variable over horizon									
	U.S.					Canada				
	$t+1$	$t+3$	$t+6$	$t+9$	$t+12$	$t+1$	$t+3$	$t+6$	$t+9$	$t+12$
Panel A: All bonds										
Total ownership	0.055***	0.163***	0.185***	0.234***	0.256***	0.089**	0.165**	0.222**	0.265**	0.289**
	(2.82)	(5.56)	(5.44)	(7.17)	(7.21)	(2.13)	(2.15)	(2.20)	(2.38)	(2.46)
Observations	95,772	93,961	91,079	88,218	85,219	21,851	21,479	20,943	20,393	19,831
Adj. R^2	0.123	0.176	0.280	0.376	0.462	0.225	0.312	0.422	0.500	0.574
Active ownership	0.004	0.069***	0.080***	0.124***	0.120***	0.070*	0.124*	0.174*	0.209**	0.227**
	(0.30)	(2.97)	(2.87)	(4.45)	(4.07)	(1.86)	(1.76)	(1.85)	(2.06)	(2.04)
Observations	95,772	93,961	91,079	88,218	85,219	21,851	21,479	20,943	20,393	19,831
Adj. R^2	0.110	0.171	0.279	0.375	0.462	0.205	0.300	0.413	0.491	0.567
Passive ownership	0.048***	0.092***	0.104***	0.107***	0.129***	0.021***	0.045***	0.054***	0.061***	0.066***
	(4.66)	(6.85)	(7.16)	(7.46)	(9.33)	(2.70)	(3.45)	(3.28)	(3.37)	(4.01)
Observations	95,772	93,961	91,079	88,218	85,219	21,851	21,479	20,943	20,393	19,831
Adj. R^2	0.144	0.208	0.323	0.441	0.546	0.254	0.292	0.382	0.465	0.534
Panel B: Included bonds										
Total ownership	0.057**	0.192***	0.242***	0.307***	0.339***	0.135**	0.242**	0.346**	0.362**	0.457***
	(2.24)	(5.09)	(5.79)	(6.80)	(7.26)	(2.39)	(2.24)	(2.38)	(2.42)	(2.99)
Observations	53,869	52,975	51,514	50,029	48,457	13,307	13,064	12,734	12,395	12,053
Adj. R^2	0.140	0.190	0.297	0.396	0.479	0.268	0.360	0.463	0.534	0.602
Active ownership	−0.002	0.085**	0.116***	0.181***	0.182***	0.109**	0.170*	0.253*	0.264*	0.357**
	(−0.09)	(2.66)	(3.12)	(4.48)	(4.27)	(2.10)	(1.70)	(1.87)	(1.94)	(2.46)
Observations	53,869	52,975	51,514	50,029	48,457	13,307	13,064	12,734	12,395	12,053
Adj. R^2	0.126	0.184	0.293	0.391	0.477	0.249	0.349	0.455	0.525	0.595
Passive ownership	0.054***	0.107***	0.125***	0.125***	0.148***	0.031***	0.076***	0.095***	0.102***	0.103***
	(4.83)	(7.10)	(7.84)	(7.85)	(9.38)	(2.89)	(4.40)	(4.27)	(4.10)	(4.27)
Observations	53,869	52,975	51,514	50,029	48,457	13,307	13,064	12,734	12,395	12,053
Adj. R^2	0.165	0.229	0.347	0.456	0.552	0.291	0.322	0.414	0.487	0.543
Panel C: Excluded bonds										
Total ownership	0.040*	0.112***	0.095**	0.110***	0.121***	0.024	0.037	0.060	0.142	0.136
	(1.83)	(3.12)	(2.15)	(2.68)	(2.84)	(0.50)	(0.40)	(0.50)	(0.99)	(0.99)
Observations	41,654	40,737	39,322	37,950	36,534	8,498	8,369	8,165	7,951	7,733
Adj. R^2	0.136	0.211	0.322	0.417	0.501	0.199	0.285	0.414	0.505	0.592
Active ownership	0.005	0.037	0.015	0.025	0.012	0.024	0.057	0.087	0.149	0.119
	(0.31)	(1.20)	(0.38)	(0.64)	(0.31)	(0.51)	(0.64)	(0.75)	(1.04)	(0.89)
Observations	41,654	40,737	39,322	37,950	36,534	8,498	8,369	8,165	7,951	7,733
Adj. R^2	0.124	0.209	0.326	0.419	0.501	0.180	0.274	0.404	0.498	0.586
Passive ownership	0.034***	0.073***	0.078***	0.082***	0.106***	0.000	−0.014	−0.015	−0.001	0.021
	(3.58)	(5.73)	(5.32)	(5.62)	(7.48)	(0.02)	(−1.19)	(−0.87)	(−0.05)	(0.98)
Observations	41,654	40,737	39,322	37,950	36,534	8,498	8,369	8,165	7,951	7,733
Adj. R^2	0.150	0.232	0.362	0.493	0.600	0.240	0.312	0.411	0.510	0.598

This table reports the ΔBMI coefficient estimates from regression (16) in the U.S. and Canadian samples of all bonds (Panel A), included bonds (Panel B), and excluded bonds (Panel C) with time to maturity between 4 and 6 years. The dependent variable is the change in fund ownership (total, active, or passive) in percentage points from month t to the stated horizon. All regressions control for the logarithm of the bond's par value, numeric credit rating, bid-ask spread quartile, time to maturity, duration, and include bond and year–month fixed effects. t-statistics based on wild bootstrap with clustering by issuer and month are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$ based on wild-bootstrap p-values.

Figure 6: The estimated effect of BMI on corporate bond fund ownership



(a) Canada



(b) U.S.

This figure plots the estimates and 90% confidence intervals from Table 4 in (a) Canada and (b) the U.S. A bond is considered *excluded* if it is in the bottom three deciles by size (par value) within its sector-rating-maturity bucket in all indexes to which it belongs and is considered *included* otherwise.

level of sparsity, and our estimates reflect very little rebalancing in the set of bonds predicted to be excluded. In contrast, the magnitude of the increase in active ownership is similar at short horizons for both included and excluded bonds, although it is significantly different from zero only for included bonds.

The effects of BMI changes on ownership are economically non-trivial. In the U.S.,

an increase in BMI upon crossing the 5-year cutoff (around 2.1 p.p.) is associated with a 50 bps persistent increase in total fund ownership. In contrast to the previous literature, we find that active and passive funds contribute equally. Furthermore, the increase is almost entirely driven by included bonds, whose fund ownership increases by 70 bps, while excluded bonds' ownership increases by around 20 bps solely due to passive funds rebalancing. The magnitude of the effect is very similar in Canada, except it is almost entirely driven by active fund rebalancing, given their share of the market.

As with prices, Appendix C.8 reports pooled specifications testing directly whether the ownership response differs between included and excluded bonds. The difference is positive and significant in the U.S. for total ownership from $t+3$ onward, for active ownership from $t+6$ onward, and for passive ownership at every horizon; in Canada it is significant throughout for passive ownership. As discussed in Section 4.3, the Canadian excluded bond sample is much smaller, yet all estimates are directionally consistent with our mechanism.

4.5 Price impact estimates

Our results establish that changes in BMI translate into shifts in fund ownership, particularly among bonds predicted to be included in sparse portfolios. A natural next question is how large these ownership changes are relative to the observed changes in bond prices. In this section, we estimate the price impact of benchmarked investors' trades by examining how changes in their ownership of a bond affect its price. As our model highlights, bond ownership and prices are jointly determined in equilibrium, which presents an identification challenge.

To address the identification challenge, we instrument changes in fund ownership with changes in BMI, following Pavlova and Sikorskaya (2023). In particular, we estimate the following two-stage least squares specification:

$$\Delta Own_{i,t+h} = \boldsymbol{\beta} \Delta BMI_{i,t} + \zeta_1' \bar{X}_{i,t} + \nu_i + \mu_t + \epsilon_{i,t+h}, \tag{17}$$

$$\Delta Y_{i,t+h} = \alpha \widehat{\Delta Own}_{i,t+h} + \zeta_2' \bar{X}_{i,t} + \eta_i + \xi_t + \varepsilon_{i,t+h}, \tag{18}$$

where $\Delta Own_{i,t+h}$ denotes the change in total fund ownership between t and $t+h$, $\widehat{\Delta Own}_{i,t+h}$ is its fitted value, and $\Delta Y_{i,t+h}$ is either the bond's total return or the change in its yield spread over the same horizon. All other variables are defined under equation (16) above. The regressions include bond and year–month fixed effects and the same set of controls as in the baseline specifications. We estimate the model at horizons of 3, 6, and 12 months for bonds around the 5-year maturity cutoff. Following our baseline, we report results for the full sample as well as separately for bonds predicted to be *included* in and *excluded* from fund portfolios.

Table 5 reports the results. We focus the discussion on the U.S., where the instrument is strong. The first-stage estimates confirm that changes in BMI are strong predictors of ownership changes, which is primarily driven by included bonds. The Kleibergen–Paap rk Wald F-statistics are above conventional thresholds in the full and included samples. In contrast, the first stage is weaker for excluded bonds, as the model predicts, and we therefore do not interpret the corresponding second-stage estimates.

Turning to the second stage, we find an economically meaningful price impact of fund ownership changes. For total returns, the IV estimates imply that increases in fund ownership lead to positive subsequent returns. For example, at the 12-month horizon in the U.S., the price impact estimate $\hat{\alpha}$ for included bonds is approximately 0.30. Similarly, yield spreads decline following increases in fund ownership. These effects are again concentrated among included bonds. In the U.S., a one percentage point increase in ownership reduces yield spreads by roughly 11 bps at the three-month horizon and 6.5 bps at six months for included bonds.

Our estimates are of a similar order of magnitude as—and in some cases somewhat larger than—those reported in the literature. Bretscher, Schmid, and Ye (2024) exploit the same maturity-based cutoff and report a reduction in yield spreads of about 4.3 bps within three months following passive demand shocks. Dannhauser (2017) finds that a one percentage point increase in ETF ownership of investment-grade bonds is associated with an approximately 7 bps decline in spreads in the cross section. Our estimates for included bonds

Table 5: Instrumented price impact across horizons

	Price impact for								
	All bonds			Included bonds			Excluded bonds		
	t+3	t+6	t+12	t+3	t+6	t+12	t+3	t+6	t+12
Panel A: U.S.									
Total return (%)	0.56**	0.36	0.28	0.59**	0.36**	0.30**	0.74	1.05	0.29
	(2.24)	(1.43)	(1.58)	(2.48)	(2.09)	(2.17)	(1.18)	(0.49)	(0.54)
Yield spread (bps)	−10.77**	−5.62	−2.09	−11.21**	−6.50**	−3.43	−13.93	−10.84	2.77
	(−2.56)	(−1.51)	(−0.84)	(−2.59)	(−2.30)	(−1.57)	(−1.24)	(−0.42)	(0.32)
$\widehat{\Delta Own}$	0.09***	0.10***	0.17***	0.11***	0.16***	0.24***	0.05	0.02	0.06
	(3.44)	(3.06)	(4.85)	(3.35)	(3.97)	(5.17)	(1.55)	(0.59)	(1.59)
KP rk Wald F-stat.	11.81	9.68	23.17	11.50	16.09	26.23	2.30	0.35	2.49
Observations	86,370	83,694	78,198	49,516	48,144	45,217	36,592	35,302	32,743
Panel B: Canada									
Total return (%)	0.41*	0.34*	0.27*	0.44*	0.29*	0.28**	1.24	0.93	0.24
	(1.66)	(1.79)	(1.84)	(1.86)	(1.90)	(2.23)	(0.30)	(0.42)	(0.72)
Yield spread (bps)	−3.18	−6.15*	−6.61**	−2.94	−4.90*	−6.44**	−23.70	−21.17	−7.32
	(−1.39)	(−1.90)	(−2.00)	(−1.41)	(−1.88)	(−2.25)	(−0.30)	(−0.43)	(−0.91)
$\widehat{\Delta Own}$	0.15**	0.21**	0.30***	0.22**	0.32**	0.45***	0.03	0.05	0.15
	(2.05)	(2.12)	(2.61)	(2.22)	(2.29)	(3.06)	(0.36)	(0.46)	(1.17)
KP rk Wald F-stat.	4.05	4.42	6.87	4.86	5.28	9.37	0.09	0.18	1.34
Observations	20,645	20,118	19,031	12,567	12,243	11,577	8,035	7,831	7,410

This table reports 2SLS estimates of specification (18) across 3-, 6-, and 12-month horizons. It also reports the first-stage estimates of (17) and the corresponding Kleibergen–Paap rk Wald F-statistics. We report estimates for all bonds with time to maturity between 4 and 6 years as well as estimates for included and excluded bonds separately. A bond is considered *excluded* if it is in the bottom three deciles by size (par value) within its sector-rating-maturity bucket in all indexes to which it belongs and is considered *included* otherwise. The dependent variable is either the total return or the change in yield spread from month t to the stated horizon. All regressions control for the logarithm of the bond's par value, numeric credit rating, bid-ask spread quartile, time to maturity, duration, and include bond and year–month fixed effects. t-statistics, based on standard errors clustered by issuer and year–month, are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

are around 10–11 bps at comparable horizons. This difference likely reflects the fact that we isolate the price impact of benchmark-induced ownership changes using an instrumental variable approach and focus on bonds that are effectively included in fund portfolios, where demand shocks are strongest and substitution is limited.

The Canadian estimates deserve separate treatment, because the instrument is weak there. The Kleibergen–Paap rk Wald F-statistics for Canadian included bonds are below conventional thresholds at every horizon, and they fall further under the more demanding

specifications in Appendix C.9. The second-stage point estimates for Canadian included bonds are of the same order of magnitude as the U.S. estimates and are marginally significant at the longer horizons, but with a first stage this weak, conventional two-stage least squares standard errors overstate their precision and we cannot rule out weak-instrument bias. The Canadian estimates are therefore meant to support the U.S. results rather than provide independent evidence, and we base our quantitative comparisons with the literature on the U.S. sample. We note that the Canadian results in Sections 4.3 and 4.4 are unaffected by this concern.

These estimates capture equilibrium price responses to changes in BMI and therefore incorporate general-equilibrium reallocation across substitute bonds. Existing price-impact studies typically define substitutability through similarity in cash-flow characteristics (e.g., Chaudhary, Fu, and Li, 2026). Our model adds a second dimension. Cash-flow correlation determines how well one bond can replicate another's payoff; the benchmark weight determines how much tracking error reduction a manager forgoes by not holding that particular bond. Two bonds can thus be equally good substitutes in cash-flow terms and still not be interchangeable for a benchmarked manager: by Lemma 3, the value of holding a bond rises with its benchmark weight, so the manager is least willing to drop the bonds with the largest weights in each bucket. Consistent with that, included bonds are where we expect to see the strongest price effects.

In line with this interpretation, our results imply detectable multipliers at the individual-bond level. We note that our estimates should be interpreted as local multipliers relative to close substitutes in terms of cash-flow characteristics (see Haddad, He, Huebner, Kondor, and Loualiche, 2025; Koijen and Yogo, 2019). To support this interpretation, Appendix C.9 shows that the U.S. estimates are robust to specifications designed to address potential differential spillovers across nearby bonds: rating-maturity-month fixed effects, asymmetric narrowing of the window around the 5-year cutoff, and time-varying slopes on bond characteristics. Although the magnitudes somewhat decrease, these specifications continue to imply statistically significant multipliers for included bonds.

4.6 Discussion

Our findings contribute to the growing literature on institutional demand in fixed-income markets, while offering several novel angles relative to prior work. In this section, we discuss these differences and also address several potential caveats in our research design.

Our approach builds on the maturity-based discontinuities in corporate bond markets first documented by Bretscher, Schmid, and Ye (2024), but it differs in several important respects, both relative to that paper and more broadly to the literature on institutional ownership in corporate bond markets.

First, rather than focusing exclusively on passive ownership, we emphasize a broader benchmarking channel in which both active and passive funds rebalance in response to benchmark changes. Second, unlike much of the prior literature, we include recently issued bonds, which materially affects the estimated response of active ownership. Because new issues are typically offered at a discount, our model predicts that fund managers are particularly likely to include them in their portfolios. Consistent with this mechanism, Laipply, Madhavan, Mauro, and Udevbulu (2025) show that newly issued bonds account for a substantial share of index fund rebalancing activity. Excluding recent issues therefore mechanically understates rebalancing, especially in the presence of sparse portfolios. Accordingly, in contrast to previous studies, we document significant changes in active ownership around BMI shifts at maturity cutoffs.[16]

Third, we show that the effects around maturity cutoffs are concentrated among bonds predicted to enter managers' sparse portfolios, whereas the existing literature typically considers all bonds without accounting for portfolio sparsity. Finally, rather than relying on a binary indicator for crossing the maturity cutoff, we exploit variation in the theoretically motivated intensity of the shock. This allows us to compare magnitudes to the demand shift predicted by benchmarking and to account for heterogeneity across maturities, fund types,

[16] We report estimates for both prices and ownership in a sample without recently issued bonds in Appendix C.7. The estimates for prices are virtually the same, while the magnitudes of ownership responses are somewhat smaller, primarily for active funds. Although the confidence bands are slightly wider, reflecting the smaller sample size, the picture is qualitatively the same even with recently issued bonds excluded.

and time. For comparison, estimates based on the cutoff switch dummy are reported in Appendix C.5 and are quantitatively similar to our baseline results.

Our research design detects effects on prices and ownership by leveraging a large number of maturity-based shocks across bonds and time. Each time a bond crosses a maturity cutoff, the amount of capital benchmarked to it—and thus its BMI—often adjusts, creating thousands of incremental shocks across the panel. This rich cross-sectional and time-series variation, observed over more than a decade of monthly data, allows us to exploit the intensity of maturity-driven shocks even though sparsity is measured with noise.

One striking feature of the demand shocks in our setting is that the timing of maturity crossings is perfectly predictable. Market participants observe a bond's remaining maturity and can anticipate when it will enter or exit a maturity bucket. This contrasts with the index inclusion effects in equity markets, where index additions and deletions must be predicted. However, the magnitude of the demand shock is harder to evaluate in advance, as it depends not only on the evolution of index AUM and weights but also on which bonds are included in the sparse portfolios of individual funds. Consistent with this, we do not observe significant pre-trends in bond prices before bonds cross the 5-year maturity cutoff, despite some anticipatory rebalancing of passive funds in the U.S. We report those estimates in Appendix C.6.

Corporate bond markets are characterized by limited liquidity and infrequent trading, which raises concerns about stale prices and delayed price discovery.[17] To mitigate these issues, we rely on index-provider evaluated prices from Bloomberg and FTSE, which are widely used for portfolio valuation by institutional investors and have been adopted in academic studies.[18] Additionally, we (i) use wide event windows to capture gradual adjustments, and

[17] Related evidence from a more liquid market is provided by Bonfanti and Marcucci (2025), who show that European Union bonds—despite being safe and liquid assets—trade at yields around 50 bps higher than comparable sovereigns. These bonds are excluded from major bond indexes, resulting in lower BMIs and reduced benchmark-induced demand from funds. This evidence reinforces the idea that benchmarking compresses yields by expanding the inelastic investor base.

[18] Using evaluated rather than transaction-based pricing raises the concern that the prices of thinly traded bonds are partly inferred from comparable bonds. Because comparables within a sector–rating–maturity bucket are disproportionately the larger, included bonds, such interpolation would transmit part of the included-bond response into excluded-bond prices and therefore work against our finding of a difference

(ii) control for liquidity proxies such as bid-ask spread quartile in all specifications. Using the level of spreads or omitting these controls produces nearly identical results. Recent changes in corporate bond market structure, including the rise of portfolio trading (Li, O'Hara, Rapp, and Zhou, 2025) and changing investor composition (Li and Yu, 2026), likely interact with benchmarking and portfolio sparsity. Studying the implications of such interactions for bond pricing is an interesting direction for future work.

A related concern is that BMI changes move measured liquidity rather than expected returns. To assess this directly, we re-estimate regression (16) using the change in the bid-ask spread as the dependent variable. The estimates are reported in Appendix C.10. Spreads do narrow after an increase in BMI, with the response peaking at three months in the U.S. and at six months in Canada. This is consistent with the coordination mechanism emphasized by Sialm, Shin, and Zhu (2025) and Feldhütter, Poulsen, and Woschitz (2026), in which benchmark-driven demand concentrates trading in a subset of bonds and makes them easier to trade. In our setting, however, the improvement is neither persistent—by twelve months the point estimates are indistinguishable from zero in both markets—nor aligned with the sparsity split, since in Canada the response is at least as large for excluded bonds as for included ones. Controlling for contemporaneous spread changes also does not affect our main estimates. It is therefore unlikely that this transitory improvement in liquidity is the source of the pricing effects in Section 4.3.[19]

One alternative interpretation of our sparsity results is that the relation between bond size, pricing, and ownership reflects liquidity differences, as larger issues tend to be more liquid and easier to trade in over-the-counter markets. If the excluded-bond coefficients were attenuated because their prices adjust with delay, the difference between the two groups of bonds would shrink as the horizon lengthens and prices catch up. It does not.[20] This pattern is consistent with portfolio sparsity and not delayed price discovery.

between the two groups. Additionally, in unreported analyses, we re-estimate the U.S. specifications using transaction-based prices and yields from the WRDS Bond Return Database and obtain a qualitatively unchanged picture.

[19] We thank Manav Chaudhary for suggesting this test.

[20] For example, in our main tests, the difference in the return response is largest at $t + 12$.

Several unreported analyses also indicate that a liquidity classification does not identify the same set of bonds, or the same effect. First, sorting bonds by bid-ask spread within the same buckets, rather than by size, produces an inclusion classification that agrees with ours in only about half of bond-months and the two indicators are only weakly associated. Second, large bonds are usually also liquid, but the converse fails. Third, when we re-estimate our specifications splitting on liquidity instead of size, the ownership response of the included bonds becomes weaker, and the response among bonds that are liquid but small is indistinguishable from zero. Fourth, the differences in liquidity between our two groups are also modest.

Finally, to further distinguish between size and liquidity, we document sharp ownership discontinuities around benchmark issue-size eligibility cutoffs in both Canada and the U.S. In Appendix C.11, we show that even among the smallest bonds there is a discrete jump in fund ownership exactly at the benchmark inclusion threshold. This pattern holds for both active and passive funds, extending earlier evidence documented for passive funds in fixed-income markets (e.g., Dathan and Davydenko, 2026). The presence of similar discontinuities for active funds provides additional support for our mechanism: benchmarking, rather than bond size mechanically proxying for liquidity or other bond characteristics, plays an important role in shaping fund ownership.

5 Concluding Remarks

In this paper, we present evidence from Canada and the U.S. that corporate bond ownership by both active and passive funds is strongly influenced by their mandates, and in particular their benchmark indexes. In our comprehensive sample of funds and their benchmarks, tracking errors are relatively small for both active and passive funds, as both types of funds steer close to their benchmarks. Accordingly, we find that benchmark weights strongly predict active and passive fund ownership.

We document that corporate bond funds' portfolios are sparse; the average fund holds

only 15% of benchmark bonds in Canada and only 7% in the U.S. To rationalize this sparsity and the related stylized facts, we develop a model of sparsity in asset management whose two key ingredients are benchmarking and portfolio management costs. The model implies that fund managers hold sparse portfolios that include assets with higher benchmark weights, larger size, and higher idiosyncratic risk within each bond category. The model also predicts that bond prices depend on benchmarking intensity, a measure of the inelastic benchmark-induced demand from fund managers. Exploiting plausibly exogenous within-bond changes in BMI driven by reallocations of benchmarked capital around maturity cutoffs, we document that increases in bonds' BMIs lead to reductions in bond yields and increases in active and passive fund ownership—but primarily for bonds predicted to be included in sparse portfolios.

A growing literature highlights the potential financial fragility arising from bond mutual funds, which engage in liquidity transformation—offering daily redemptions while holding illiquid assets—and may be forced to sell illiquid corporate bonds to meet investor outflows (see, e.g., Goldstein, Jiang, and Ng, 2017; Jiang, Li, Sun, and Wang, 2022). Our findings add a new cross-sectional dimension to this concern. Because fund portfolios are sparse, fund ownership and flow-induced selling pressure are concentrated in the bonds that managers choose to include. These included bonds—typically larger bonds within categories of close substitutes—are therefore more exposed to fire-sale pressure from mutual-fund outflows, whereas excluded bonds are less directly exposed to this channel.

Our BMI measure, combined with the theory-implied sparsity adjustment, offers a way to quantify the inelastic demand in bond markets stemming from fund manager mandates for both active and passive funds. One method for estimating price elasticities of demand has been to use changes in *passive* bond ownership around index rebalancing events as demand shifters. Our results demonstrate that the demand of active corporate bond funds also contributes to these demand shifters. Given that active funds manage more assets—particularly in the earlier part of our sample—they represent an important component of benchmark-induced demand and thus contribute significantly to these aggregate effects.

Finally, recent work in corporate finance shows that passive ownership can lower firms'

cost of bond financing and influence real economic outcomes.[21] Our findings indicate that active bond mutual funds behave similarly to passive funds in a key respect: they closely follow their mandates, but do so through sampling. Benchmarking by active funds and portfolio sparsity therefore shape the transmission of fund demand to bond prices and may also be relevant for understanding how institutional ownership affects firms' financing conditions.

References

Afik, Zvika, Gady Jacoby, David Stangeland, and Zhenyu Wu, 2019, The Make-Whole and Canada-Call Provisions: A Case of Cross-Country Spillover of Financial Innovation, *Journal of International Financial Markets, Institutions and Money* 61, 120–127.

Basak, Suleyman, and Anna Pavlova, 2013, Asset Prices and Institutional Investors, *American Economic Review* 103, 1728–1758.

Bonfanti, Giovanni, and Juri Marcucci, 2025, A European Safe Asset? Not Without the Investors, Working Paper.

Brennan, Michael, 1993, Agency and Asset Pricing, Working Paper.

Bretscher, Lorenzo, Lukas Schmid, Ishita Sen, and Varun Sharma, 2025, Institutional corporate bond pricing, *The Review of Financial Studies* 39, 605–660.

Bretscher, Lorenzo, Lukas Schmid, and Tiange Ye, 2024, Passive Demand and Active Supply: Evidence from Maturity-mandated Corporate Bond Funds, Working Paper.

Brogaard, Jonathan, Davidson Heath, and Da Huang, 2026, ETF Sampling and Index Arbitrage, *Journal of Financial and Quantitative Analysis* 61, 547–579.

Buffa, Andrea, Dimitri Vayanos, and Paul Woolley, 2022, Asset Management Contracts and Equilibrium Prices, *Journal of Political Economy* 130, 2148–2205.

Cai, Yuhua, Anna Helmke, Benjamin Mosk, and Felix Suntheim, 2025, Fund Investor Types and Bond Market Volatility, Global Financial Stability Note, International Monetary Fund.

[21] Dathan and Davydenko (2026) argue that firms cater to passive investor demand, and Mota and Siani (2025) show that firms consider their investor clienteles in bond issuance decisions.

Chaudhary, Manav, 2024, Regulator beliefs, Working Paper.

——— , Zhiyu Fu, and Jian Li, 2026, Corporate Bond Multipliers: Substitutes Matter, *The Review of Financial Studies, forthcoming.*

Choi, Jaewon, K. J. Martijn Cremers, and Timothy Brandon Riley, 2025, Passive Bond Fund Management Is an Oxymoron (or the Case for the Active Management of Bond Funds), Working Paper.

Dannhauser, Caitlin D., 2017, The Impact of Innovation: Evidence from Corporate Bond Exchange-Traded Funds (ETFs), *Journal of Financial Economics* 125, 537–560.

——— , and Saeid Hoseinzade, 2022, The Unintended Consequences of Corporate Bond ETFs: Evidence from the Taper Tantrum, *Review of Financial Studies* 35, 51–90.

Dathan, Michele, and Sergei Davydenko, 2026, Debt Issuance in the Era of Passive Investment, *Management Science* 72, 2565–2593.

Dick-Nielsen, Jens, and Marco Rossi, 2019, The Cost of Immediacy for Corporate Bonds, *Review of Financial Studies* 32, 1–41.

Dow, James, and Sérgio R. da Costa Werlang, 1992, Uncertainty Aversion, Risk Aversion, and the Optimal Choice of Portfolio, *Econometrica* 60, 197–204.

Feldhütter, Peter, Thomas Kjær Poulsen, and Jiri Woschitz, 2026, Identifying Benchmark Demand in Corporate Bond Markets, Working Paper.

Gabaix, Xavier, 2014, A Sparsity-Based Model of Bounded Rationality, *The Quarterly Journal of Economics* 129, 1661–1710.

——— , Ralph S. J. Koijen, Robert J. Richmond, and Motohiro Yogo, 2025, Upgrading Credit Pricing and Risk Assessment through Embeddings, Working Paper.

Goldstein, Itay, Hao Jiang, and David T. Ng, 2017, Investor Flows and Fragility in Corporate Bond Funds, *Journal of Financial Economics* 126, 592–613.

Graves, Daniel, 2024, What Lies Beneath Zero: Censoring, Demand Estimation, and Hidden Beliefs, Working Paper.

Haddad, Valentin, Zhiguo He, Paul Huebner, Peter Kondor, and Erik Loualiche, 2025, Causal Inference for Asset Pricing, Working Paper.

He, Zhiguo, and Wei Xiong, 2013, Delegated Asset Management and Investment Mandates, *Journal of Finance* 68, 719–765.

Huang, Shiyang, Xu Lu, Yang Song, and Hong Xiang, 2026, Remeasuring Scale in Active Management, *The Review of Financial Studies*.

Jiang, Hao, Yi Li, Zheng Sun, and Ashley Wang, 2022, Does Mutual Fund Illiquidity Introduce Fragility into Asset Prices? Evidence from the Corporate Bond Market, *Journal of Financial Economics* 143, 277–302.

Kashyap, Anil, Natalia Kovrijnykh, Jian Li, and Anna Pavlova, 2021, The Benchmark Inclusion Subsidy, *Journal of Financial Economics,* 142, 756–774.

Koijen, Ralph S.J., and Motohiro Yogo, 2019, A Demand System Approach to Asset Pricing, *Journal of Political Economy* 127, 1475–1515.

Koont, Naz, Yiming Ma, Lubos Pastor, and Yao Zeng, 2025, Steering a Ship in Illiquid Waters: Active Management of Passive Funds, *Review of Financial Studies* 38, 2887–2935.

Laipply, Stephen, Ananth Madhavan, James Mauro, and Nogie Udevbulu, 2025, Fixed Income Index Funds: Demystifying Portfolio Construction and Rebalancing, *Journal of Investment Management* 23, 56–73.

Li, Jian, and Haiyue Yu, 2026, Investor Composition and the Liquidity Component in the U.S. Corporate Bond Market, *Journal of Finance* 81, 871–922.

Li, Jessica S., Maureen O'Hara, Andreas C. Rapp, and Xing Alex Zhou, 2025, Bond Market Illiquidity: Is Portfolio Trading the Solution?, Working Paper.

Li, Ziang, 2023, The Long-term Interest Rate and Corporate Bond Credit Spreads, Working Paper.

Ma, Linlin, Yuehua Tang, and Juan-Pedro Gomez, 2019, Portfolio Manager Compensation in the U.S. Mutual Fund Industry, *Journal of Finance* 74, 587–638.

Ma, Yiming, Kairong Xiao, and Yao Zeng, 2022, Mutual Fund Liquidity Transformation and Reverse Flight to Liquidity, *Review of Financial Studies* 35, 4674–4711.

Merton, Robert C., 1987, A Simple Model of Capital Market Equilibrium with Incomplete

Information, *Journal of Finance* 42, 483–510.

Mota, Lira, and Kerry Y. Siani, 2025, Financially Sophisticated Firms, Working Paper.

Ottonello, Giorgio, 2019, The Impact of Benchmarking in Fixed-Income Markets, Working Paper.

Pavlova, Anna, and Taisiya Sikorskaya, 2023, Benchmarking Intensity, *Review of Financial Studies* 36, 859–903.

Pegoraro, Stefano, Marco Sammon, and John J. Shim, 2026, Optimal Index-Linked Rebalancing with Anticipatory Trading, Working Paper.

Shim, John J., and Karamfil Todorov, 2022, ETFs, Illiquid Assets, and Fire Sales, Working Paper.

Shin, Sean Seunghun, Xing (Alex) Zhou, and Qifei Zhu, 2025, The Growing Index Effect in the Corporate Bond Market, Working Paper.

Sialm, Clemens, Sean Seunghun Shin, and Qifei Zhu, 2025, Recurring Demand for Corporate Bonds, Working Paper.

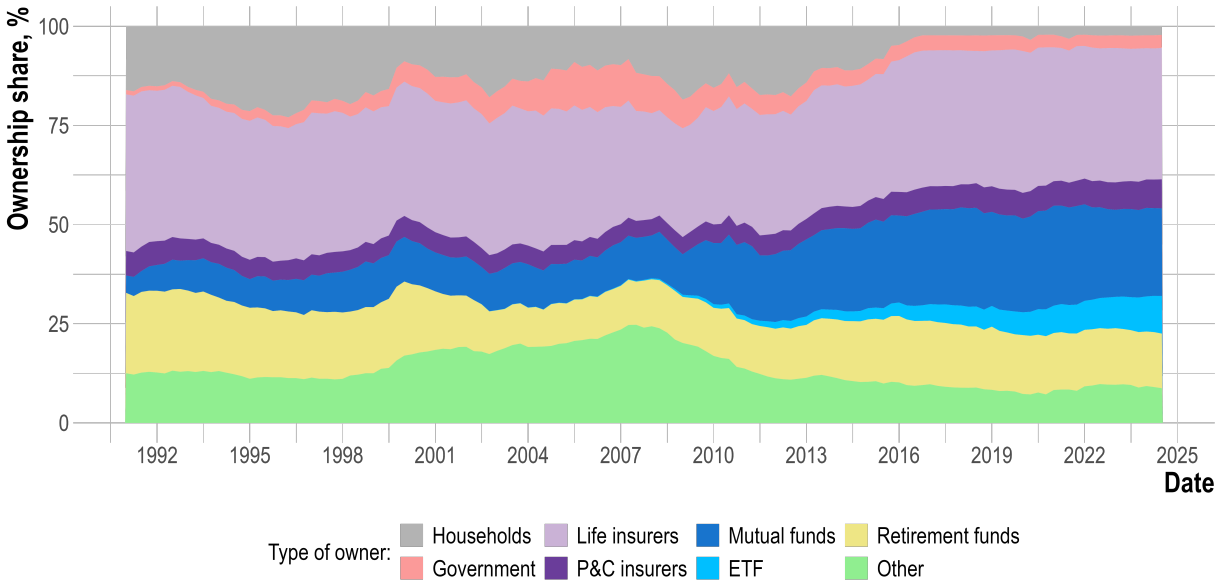
Van Nieuwerburgh, Stijn, and Laura Veldkamp, 2010, Information Acquisition and Under-Diversification, *Review of Economic Studies* 77, 779–805.

Wurgler, Jeffrey, and Ekaterina Zhuravskaya, 2002, Does Arbitrage Flatten Demand Curves for Stocks?, *Journal of Business* 75, 583–608.

A Appendix: Stylized Facts

A.1 Ownership of U.S. Corporate Bonds

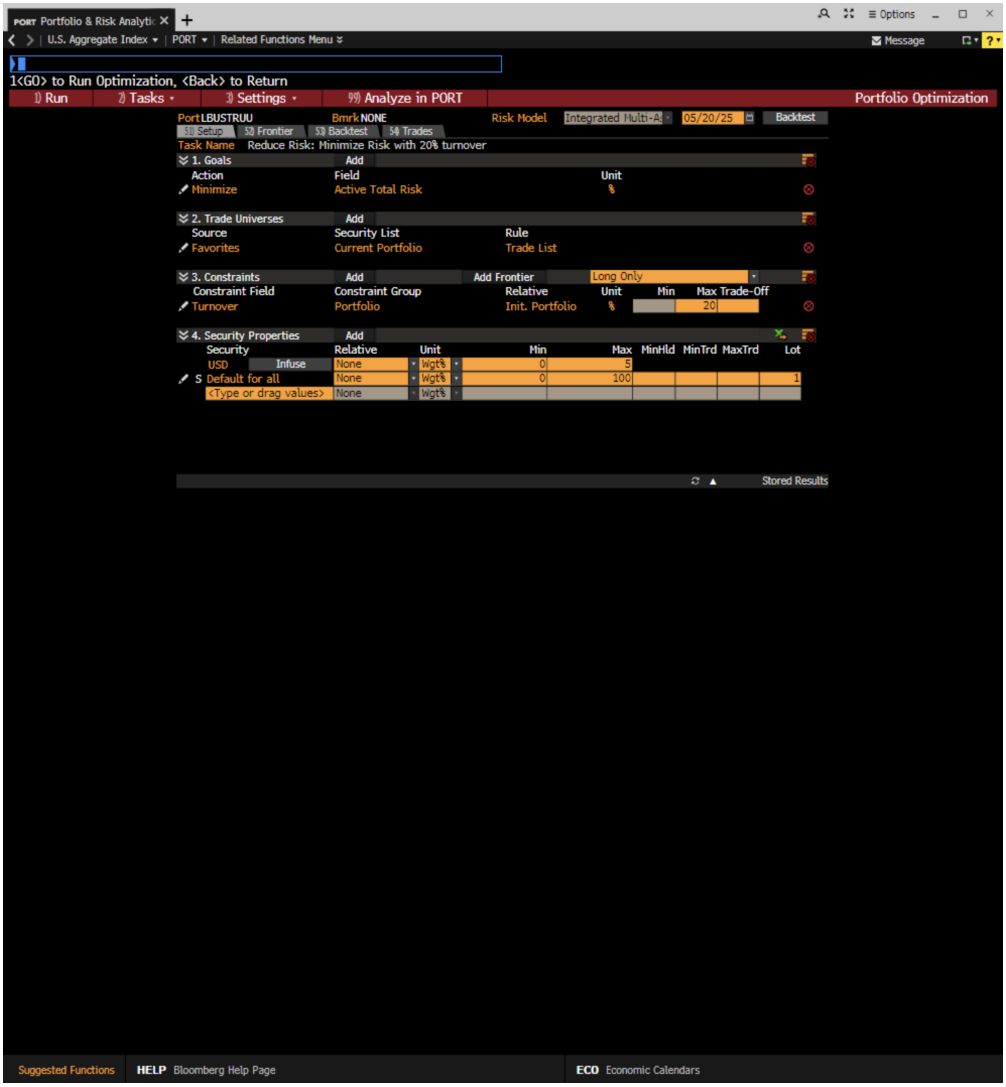
Figure A1: Ownership of U.S. Corporate Bonds by Domestic Investors



This figure illustrates the evolution of U.S. corporate bond holdings by domestic investor categories from 1990 to 2024. The data, sourced from the Federal Reserve Board's Financial Accounts of the U.S. (Z.1 release), exclude holdings by foreign investors and foreign monetary authorities. The depicted categories encompass households, mutual funds, pension funds, insurance companies, banks, and other domestic financial institutions. The data series can be accessed at: https://www.federalreserve.gov/releases/z1/default.htm.

A.2 Bloomberg portfolio allocation tool

Figure A2: Bloomberg portfolio allocation tool (PORT)



(a) Bloomberg PORT optimizer.

(Continued on the next page.)



(b) Bloomberg PORT with corporate bond classifications.

This figure presents screenshots of the Bloomberg portfolio allocation tool. Subfigure (a) demonstrates the default trade-off between tracking error and turnover in the tool, and subfigure (b) highlights the salience of Bloomberg industry classifications.

A.3 Fund samples summary statistics

Table A1: Descriptive statistics for our Canadian sample of funds

	Counts			TNA (billion USD)			Avg. # Holdings		
Year	**All**	**Passive**	**ETF**	**All**	**Passive**	**ETF**	**All**	**Passive**	**ETF**
2003	45	2	0	21.5	7.0	0.0	45.5	364.5	–
2004	44	2	0	26.4	7.9	0.0	51.4	409.0	–
2005	48	2	0	31.1	8.0	0.0	51.1	415.5	–
2006	52	2	0	36.6	8.5	0.0	50.5	436.0	–
2007	62	6	4	50.3	11.2	1.6	55.5	194.3	70.8
2008	67	6	4	39.0	9.1	1.8	62.3	227.8	122.8
2009	77	6	4	58.1	13.1	4.0	72.2	273.0	173.3
2010	78	7	5	70.0	15.1	5.0	86.2	290.1	197.0
2011	89	9	7	71.1	14.9	5.3	87.3	267.6	189.3
2012	97	11	9	98.1	18.3	8.1	94.9	283.3	214.7
2013	111	12	9	88.2	16.4	6.5	97.8	302.8	235.3
2014	116	12	9	80.0	15.7	6.1	97.7	309.3	246.3
2015	124	14	11	80.2	14.1	5.7	103.4	294.7	235.9
2016	128	14	11	93.8	16.3	7.2	104.7	312.4	261.2
2017	133	15	12	94.9	16.7	7.8	106.0	307.9	259.1
2018	138	19	16	89.0	16.3	7.7	111.2	304.4	263.6
2019	143	23	20	108.7	20.4	10.8	125.7	293.6	258.2
2020	155	28	24	162.6	23.1	12.7	141.1	277.2	249.7
2021	160	31	27	173.7	24.9	14.7	154.3	288.6	262.9
2022	161	32	28	153.4	21.4	13.4	157.7	298.2	275.7
2023	167	35	31	154.5	21.8	14.7	157.5	286.3	263.9

Portfolio size includes bond holdings only. We report the number of fund portfolios, not unique share classes. Average number of holdings is bond count per portfolio at year-end. A dash indicates not applicable.

Table A2: Descriptive statistics for our U.S. sample of funds

	Counts			TNA (billion USD)			Avg. # Holdings		
Year	**All**	**Passive**	**ETF**	**All**	**Passive**	**ETF**	**All**	**Passive**	**ETF**
2010	288	31	17	1784.4	218.5	147.4	148.5	513.9	542.8
2011	311	35	20	1955.1	270.9	178.4	166.8	612.7	670.5
2012	319	39	24	2266.6	325.3	212.4	196.4	729.3	788.2
2013	340	41	25	2171.7	335.1	214.9	193.6	791.6	876.0
2014	353	42	26	2186.7	410.3	268.2	206.1	881.1	1001.2
2015	381	52	36	2151.3	450.2	307.4	230.3	863.6	915.1
2016	410	56	39	2390.3	539.6	370.1	243.4	904.4	953.7
2017	423	56	43	2774.8	649.6	443.8	243.4	940.2	939.4
2018	430	69	52	2771.5	674	453.3	256.1	853.4	856.7
2019	447	73	55	3265.3	809.7	527.9	266.7	877.8	865.6
2020	465	77	61	3767.4	991.6	669.1	306.2	958.7	903.7
2021	459	79	63	3984.6	1088.8	722.5	307.0	972.3	931.6
2022	447	78	63	3237.6	956.2	626.8	301.2	934.9	892.1
2023	438	79	64	3519.5	1094.8	715.2	326.5	1032.3	987.2

TNA figures include matched separate accounts. Short positions are excluded. We report the number of fund portfolios, not unique share classes. Average number of holdings is corporate bond count per portfolio at year-end.

A.4 Bond samples summary statistics

Table A3: Descriptive statistics for the Canadian sample

Variable	N	Mean	SD	Min	p10	p50	p90	Max
Panel A: Prices and returns								
Yield spread	138,835	1.492	0.637	0.188	0.769	1.403	2.294	4.916
Yield spread change	137,244	−0.011	0.129	−0.595	−0.128	−0.010	0.106	0.734
Total return	136,681	0.346	1.953	−6.842	−1.792	0.284	2.442	8.034
Panel B: BMI								
Total BMI	138,835	9.186	3.468	0.819	5.841	8.957	12.877	16.534
Active BMI	138,835	7.051	2.916	0.000	4.266	6.737	10.687	12.660
Passive BMI	138,835	2.135	0.700	0.735	1.365	2.071	3.083	3.891
Total BMI change	137,244	−0.023	0.839	−9.095	−0.258	0.003	0.334	2.002
Active BMI change	137,244	−0.025	0.772	−8.417	−0.228	−0.003	0.287	1.532
Passive BMI changes	137,244	0.003	0.103	−0.678	−0.055	0.005	0.064	0.823
Panel C: Ownership								
Total ownership	138,835	12.167	10.049	0.000	0.000	10.310	26.730	100.000
Active ownership	138,835	10.338	9.455	0.000	0.000	8.194	24.214	100.000
Passive ownership	138,835	1.829	1.433	0.000	0.000	1.689	3.529	17.797
Total ownership change	137,244	0.003	1.297	−9.389	−0.458	0.000	0.465	7.816
Active ownership change	137,244	−0.002	1.127	−7.658	−0.449	0.000	0.412	7.001
Passive ownership change	137,244	0.008	0.219	−1.715	−0.059	0.000	0.099	1.215
Panel D: Other bond characteristics								
log(Amount outstanding)	138,835	19.526	0.731	15.893	18.644	19.519	20.436	22.045
Rating (numeric)	138,835	1.631	0.653	1.000	1.000	2.000	2.000	4.000
Bid-ask spread quartile	138,835	2.525	1.097	1.000	1.000	3.000	4.000	4.000
Time to maturity (years)	138,835	11.009	8.855	1.500	2.415	7.110	25.988	29.999
Modified duration	138,835	7.327	4.671	0.689	2.273	5.896	14.773	21.068

This table reports summary statistics of the main variables used in the analysis for the monthly sample of Canadian corporate bonds in 2004–2023. All variables in panels A–C are in percentage points.

Table A4: Descriptive statistics for the U.S. sample

Variable	N	Mean	SD	Min	p10	p50	p90	Max
Panel A: Prices and returns								
Yield spread	617,208	127.515	72.898	−165.780	48.800	114.860	220.550	424.310
Yield spread change	617,208	−0.974	21.564	−95.780	−19.940	−1.040	16.440	141.520
Total return	617,208	0.273	2.547	−9.520	−2.460	0.200	2.910	10.160
Panel B: BMI								
Total BMI	617,208	17.186	3.380	0.000	13.644	16.585	23.017	29.743
Active BMI	617,208	12.936	2.729	0.000	10.245	12.180	17.659	22.279
Passive BMI	617,208	4.250	1.579	0.000	2.309	3.844	6.616	7.484
Total BMI change	617,208	0.075	0.671	−1.124	−0.245	0.020	0.273	7.465
Active BMI change	617,208	0.040	0.582	−1.007	−0.227	−0.007	0.181	6.556
Passive BMI changes	617,208	0.036	0.126	−0.263	−0.035	0.027	0.095	1.434
Panel C: Ownership								
Total ownership	617,208	12.095	11.054	0.000	2.833	9.041	24.608	100.000
Active ownership	617,208	7.739	10.605	0.000	0.042	4.190	19.378	100.000
Passive ownership	617,208	4.363	2.458	0.000	1.432	4.031	7.655	100.000
Total ownership change	617,208	−0.009	1.360	−10.049	−0.560	0.001	0.495	10.624
Active ownership change	617,208	−0.049	1.253	−8.948	−0.500	0.000	0.283	9.798
Passive ownership change	617,208	0.040	0.309	−3.507	−0.083	0.005	0.211	1.979
Panel D: Other bond characteristics								
log(Amount outstanding)	617,208	13.423	0.616	12.389	12.612	13.305	14.221	16.524
Rating (numeric)	617,208	7.383	1.921	1.000	5.000	8.000	10.000	10.000
Bid-ask spread quartile	617,208	2.491	1.129	1.000	1.000	2.000	4.000	4.000
Time to maturity (years)	617,208	11.180	9.016	1.510	2.460	7.410	26.580	30.000
Option-adjusted duration	617,208	7.749	4.894	0.000	2.319	6.367	15.427	22.718

This table reports summary statistics of the main variables used in the analysis for the monthly sample of U.S. corporate bonds in 2012–2023. Yield spread and yield spread change are in basis points, and all other variables in panels A to D are in percentage points.

A.5 Bond ownership transition matrices

Table A5: Transition matrix of bond fund holdings

	1 Quarter		2 Quarters		1 Year		3 Years	
	Not held	Held	Not held	Held	Not held	Held	Not held	Held
Panel A: Canada								
Not held in last period	98.7%	1.3%	97.4%	2.6%	95.3%	4.7%	90.3%	9.7%
Held in last period	6.0%	94.0%	10.1%	89.9%	17.6%	82.4%	35.3%	64.7%
Panel B: U.S.								
Not held in last period	99.5%	0.5%	99.0%	1.0%	98.3%	1.7%	96.5%	3.5%
Held in last period	7.8%	92.2%	13.3%	86.7%	22.1%	77.9%	45.1%	54.9%

This table reports transition probabilities for corporate bond holdings of mutual funds and ETFs (Canada and U.S.). Rows are conditioned on last period's holding status; columns show the status after the indicated horizon. The reference universe is the broadest benchmark available in each geography, i.e., Bloomberg Aggregate index in the U.S. and Universe Overall Bond Index in Canada.

A.6 Fund tracking errors

Table A6: Tracking errors of bond funds

	Rolling 3-year TE			Rolling 5-year TE			
	Mean	**Median**	**St. dev.**	**Mean**	**Median**	**St. dev.**	**N**
Panel A: U.S.							
Active Fund	2.029	1.361	2.023	2.029	1.445	1.776	59,727
Passive Fund	0.906	0.323	1.250	0.780	0.303	1.071	9,984
Panel B: Canada							
Active Fund	1.373	0.749	1.573	1.401	0.846	1.522	28,046
Passive Fund	0.828	0.212	1.382	0.779	0.180	1.350	4,448

This table reports tracking errors (TE) of funds in our U.S. (Panel A) and Canadian (Panel B) samples, all in percent. TEs are computed from monthly deviations of each fund's return from its benchmark return using rolling windows: "3-year" is a 36-month window and "5-year" is a 60-month window. N reports fund/portfolio-month observations over which statistics are computed.

A.7 Aggregate sparsity

Table A7: Characteristics of the aggregate portfolios of Canadian funds with the same benchmark

	Share of index bonds held (%)	**Active Share (%)**	**Tracking Error (bps)**
Panel A: Active funds			
50/50 Short/Mid Corporate Bond Index	19	72	220
50/50 Short/Mid Overall Bond Index	8	88	57
61/39 Universe/Long Overall Bond Index	5	95	108
35/65 Short/Mid Overall Bond Index	34	64	149
Long Overall Bond Index	48	63	58
Short Corporate Bond Index	19	77	68
Short Overall Bond Index	68	45	68
Universe Corporate BBB Bond Index	5	93	313
Universe Corporate Bond Index	43	60	183
Universe Overall Bond Index	77	39	156
Weighted Average	72	43	140
Panel B: Passive funds			
Long Corporate Bond Index	56	45	36
Long Overall Bond Index	60	40	14
Mid Corporate Bond Index	74	28	18
Short Corporate Bond Index	71	32	92
Short Mid Corporate BBB Bond Index	58	44	–
Short Overall Bond Index	74	29	158
Universe Corporate Bond Index	75	29	24
Universe Overall Bond Index	78	34	89
Weighted Average	75	32	91

This table shows characteristics of the aggregate portfolios of active mutual funds (Panel A) and passive mutual funds and ETFs (Panel B) with the same benchmark in Canada. Each portfolio is a value-weighted sum of funds benchmarked to the respective index. Tracking error is reported in basis points. Share of bonds held and Active Share are expressed in percent. The weighted average uses total fund AUM benchmarked to each index (averaged over the sample), computed separately for active and passive funds.

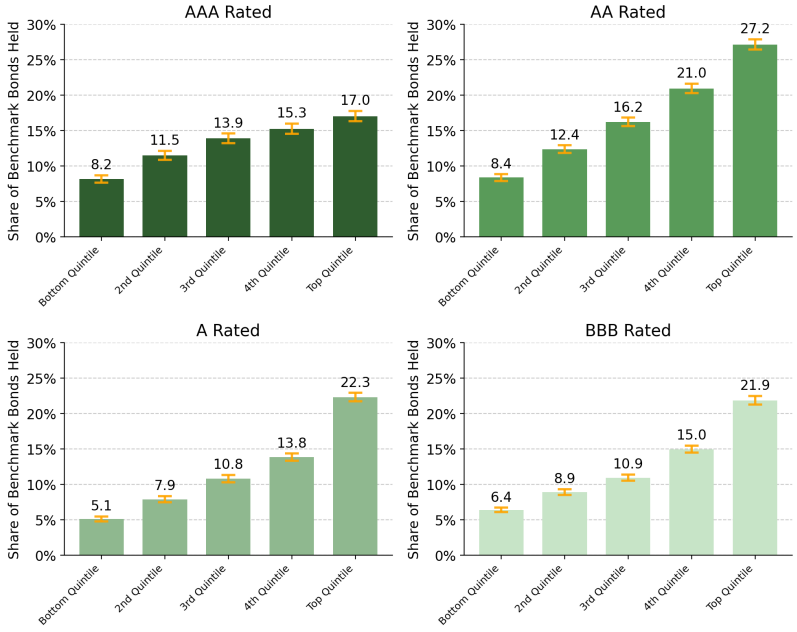
Table A8: Characteristics of the aggregate portfolios of U.S. funds with the same benchmark

	Share of index bonds held (%)	**Active Share (%)**	**Tracking Error (bps)**
Panel A: Active funds			
Aggregate 1-10 year Ex-BAA	27	68	73
Aggregate 1-10 year	9	88	185
Aggregate 1-3 year	7	94	212
Aggregate 3-5 year	0	100	479
Aggregate	78	50	229
Corporate 1-10 year	2	97	119
Corporate 1-3 year	6	91	88
Corporate 1-5 year	7	91	94
Corporate 10+ year	24	75	207
Corporate	17	81	132
Credit 1-10 year	8	90	147
Credit 1-3 year	14	84	71
Credit 1-5 year	49	52	142
Credit 10+ year A+	39	52	–
Credit 10+ year	17	82	222
Credit 5-10 year	45	52	113
Credit	30	72	167
Govt/Credit 1-10 year A+	2	87	68
Govt/Credit 1-10 year	34	67	123
Govt/Credit 1-3 year Ex-BAA	20	49	59
Govt/Credit 1-3 year	62	65	142
Govt/Credit 1-5 year	22	78	200
Govt/Credit 10+ year	15	83	216
Govt/Credit	9	90	369
Weighted Average	66	56	211
Panel B: Passive funds			
Aggregate 1-3 year	3	96	184
Aggregate 1-5 year	13	84	132
Aggregate	93	23	108
Corporate 1-10 year	69	34	28
Corporate 1-3 year	83	24	7
Corporate 1-5 year	88	20	17
Corporate 10+ year	89	19	75
Corporate 5-10 year	90	17	43
Corporate	42	66	185
Credit 1-10 year	55	48	7
Credit 1-3 year	45	55	41
Credit 10+ year	88	19	66
Credit	41	55	59
Govt/Credit 1-10 year	51	48	5
Govt/Credit 1-3 year	24	74	112
Govt/Credit 1-5 year	81	28	51
Govt/Credit 10+ year	88	17	77
Govt/Credit 5-10 year	85	20	34
Govt/Credit	20	75	9
Weighted Average	90	24	91

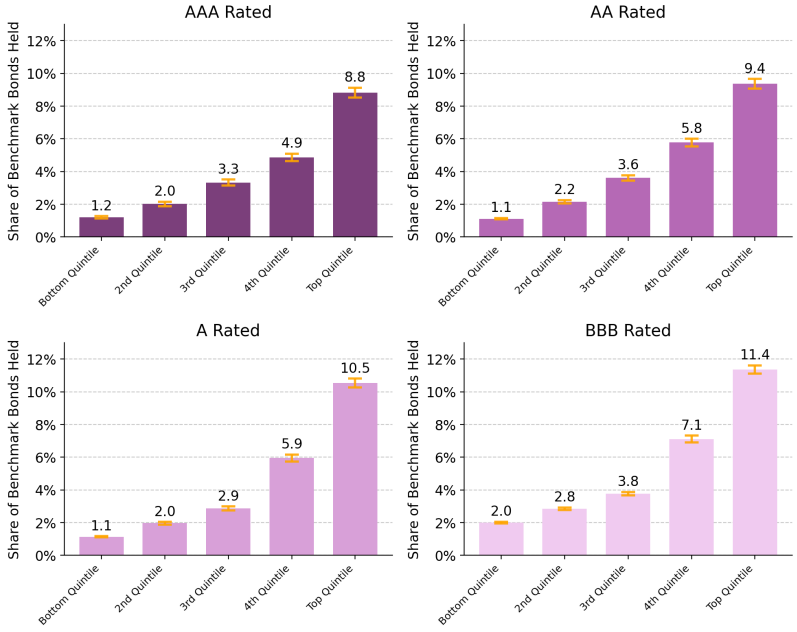
This table reports characteristics of aggregate portfolios of U.S. active funds (Panel A) and passive funds and ETFs (Panel B) sharing the same benchmark. Each portfolio is a value-weighted sum of funds benchmarked to the index. Tracking error is in basis points; share of bonds held and Active Share are in percent. Weighted averages use total fund AUM benchmarked to each index (averaged over the sample), computed separately for active and passive funds.

A.8 Share of benchmark bonds held across benchmark weight quintiles by credit ratings

Figure A3: Share of benchmark bonds held across benchmark weight quintiles by credit ratings



(a) Canada

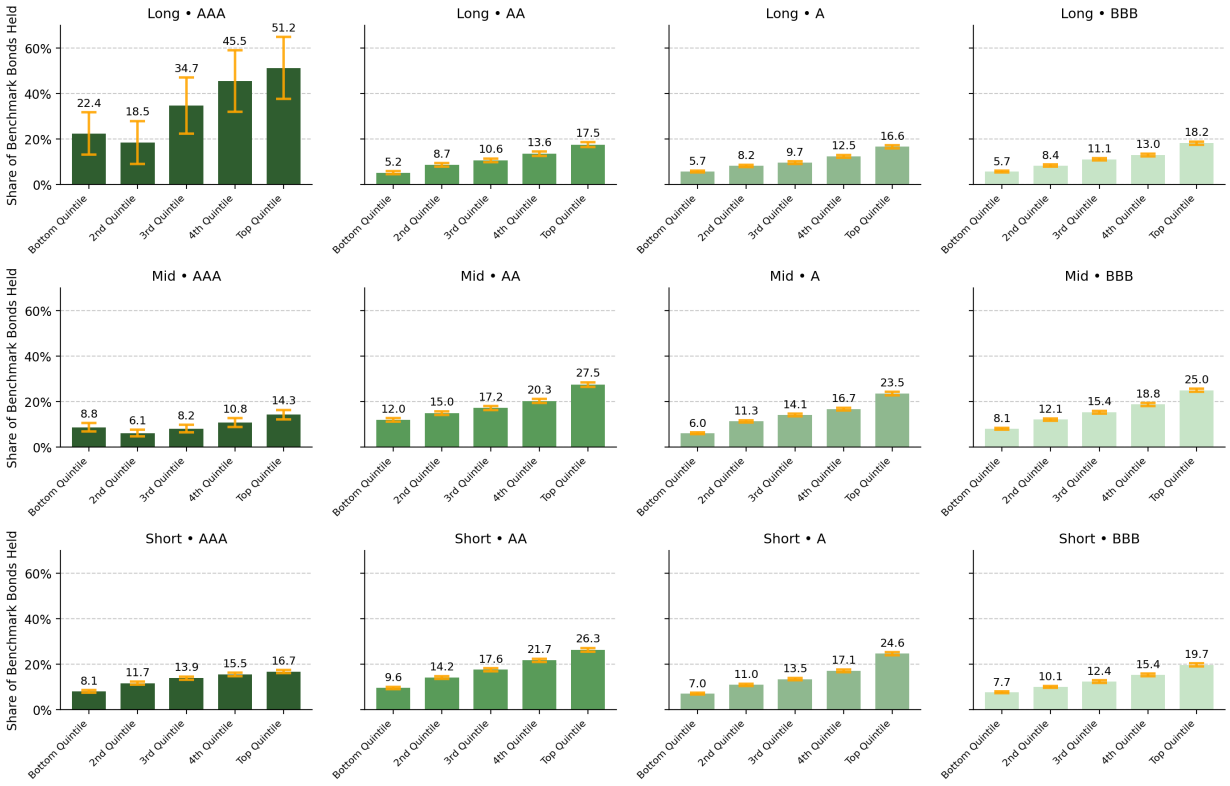


(b) U.S.

This figure illustrates the share of benchmark bonds held across benchmark weight quintiles by rating bucket.

A.9 Share of benchmark bonds held across benchmark weight quintiles by credit rating and maturity buckets

Figure A4: Share of benchmark bonds held across benchmark weight quintiles by rating-maturity buckets



(a) Canada

Figure continues on the next page

Figure A4: Share of benchmark bonds held across benchmark weight quintiles by rating-maturity buckets



(b) U.S.

This figure illustrates the share of benchmark bonds held across benchmark weight quintiles by rating and maturity buckets.

A.10 Panel regressions of fund holdings on benchmark index membership and weights

Table A9: Regressions of active and passive fund holdings on benchmark index membership and weights for Canada and U.S.

	Fund weight | fund weight $\neq 0$			**Dummy(fund weight > 0)**		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A1: Canada – Passive funds						
Dummy(within benchmark)		0.005***	0.002*		0.567***	0.437***
		(5.00)	(1.70)		(48.01)	(34.43)
Benchmark weight			0.925***			44.35***
			(18.88)			(16.51)
Observations	182,061	182,061	182,061	613,675	613,675	613,675
Adj. R^2	62.7%	62.8%	63.9%	39.3%	55.5%	56.3%
Panel A2: Canada – Active funds						
Dummy(within benchmark)		0.002***	0.0001		0.155***	0.094***
		(5.25)	(0.19)		(25.03)	(17.39)
Benchmark weight			0.615***			24.05***
			(5.31)			(12.90)
Observations	377,498	377,498	377,498	4,039,999	4,039,999	4,039,999
Adj. R^2	71.9%	71.9%	71.9%	12.5%	13.9%	14.1%
Panel B1: U.S. – Passive funds						
Dummy(within benchmark)		0.0005***	−0.0001		0.544***	0.486***
		(5.52)	(−1.06)		(69.68)	(57.86)
Benchmark weight			0.782***			111.9***
			(24.50)			(12.19)
Observations	2,635,568	2,635,568	2,635,568	17,781,083	17,781,083	17,781,083
Adj. R^2	81.5%	81.5%	81.7%	17.1%	34.0%	34.2%
Panel B2: U.S. – Active funds						
Dummy(within benchmark)		0.001***	0.0004***		0.054***	0.024***
		(16.74)	(4.00)		(67.25)	(23.70)
Benchmark weight			0.633***			51.99***
			(7.13)			(33.65)
Observations	2,450,524	2,450,524	2,450,524	99,265,445	99,265,445	99,265,445
Adj. R^2	77.5%	77.5%	77.6%	5.9%	6.5%	6.7%

This table reports estimates of the conditional correlation between fund weight or propensity to hold and benchmark membership dummy variable and benchmark index weight separately for passive and active funds. Panels A1 and A2 report the results for Canada, and Panels B1 and B2 for the U.S. All regressions are saturated by including fund-by-date and bond-by-date fixed effects. t-statistics clustered at the bond and year-month levels are presented in parentheses. Significance levels are marked as: $^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

A.11 Construction of historical benchmark data for U.S. funds

We assembled a dataset of historical benchmarks for U.S. mutual funds and ETFs from the database of historical fund prospectuses available on the website of the U.S. Securities and Exchange Commission (SEC)[22]. We further validated them using two additional sources: (1) a snapshot of benchmarks in Morningstar as of March 2025, and (2) SEC Mutual Fund Prospectus Risk/Return Summary datasets (MFRR)[23]. Benchmarks are mentioned in the annual return summary published in prospectuses.

A.11.1 Extract Benchmarks

For the crsp_cik_map table in WRDS (containing a total of 2,586 company CIKs), we downloaded all 485APOS and 485BPOS files for 2,504 company CIKs using the sec_edgar_downloader Python package. By comparing several samples, we confirmed that all required file types were successfully downloaded. A manual check verified that the remaining company CIKs without downloads do not have data available on the EDGAR website.

From the 485APOS and 485BPOS files, we extracted the table(s) containing or immediately following the string "(average) annual (total) return(s)" (with parentheses indicating optional terms), marked by the `<table></table>` tags. If a table contained any of the following terms, we retained the table:

- s&p, russell, crsp, msci, dj, dow jones, nasdaq, ftse, topix, tse, schwab, barclays, wilshire, bridgeway, guggenheim, calvert, kaizen, lipper, redwood, w.e. donoghue, essential treuters, barra, ice bofaml, bbgbarc, cboe
- benchmark, index (note: does not match *index fund* or *index series*)

We used the Gemini API to determine whether each extracted table was an average annual return table. If it was, we extracted the corresponding benchmark for each fund and labeled the benchmark in sequence.

[22] https://www.sec.gov/edgar/searchedgar/mutualsearch.html
[23] https://www.sec.gov/dera/data/mutual-fund-prospectus-risk-return-summary-data-sets

If a file listed all series names (fund names) at the beginning using the `<series-name>` tag, we identified the closest preceding series name for each table.

For tables where the closest preceding series name was not found in the previous step, we identified the most similar series name and its corresponding series ID from the SEC-provided mapping using the fund name(s) in the table.

If neither of the above methods succeeded, we used the Gemini API to analyze the text preceding each table and identify the string closest to the table that was most likely to be a fund name.

For each table, we prioritized the selection of the extracted series name as follows:

1. First, use the closest preceding series name as the extracted series name.
2. If no series name is found, use the closest preceding string most likely to be a fund name, as identified from the text preceding the table.
3. If neither of the above is available, use the fund name within the table as the extracted series name.

The SEC provides the company CIK, company name, series ID, and series name for each fund for every year from 2010 to 2024. For each "average annual return" table in year t, we extracted the SEC mapping for the corresponding company CIK for years $t-1$, t, and $t+1$. If the table was from a year earlier than the earliest available year for the company CIK, we retrieved the minimum available year and the next year's SEC mapping. For each extracted series name, we identified the exactly matched series name and its corresponding series ID from the SEC-provided mapping.

A.11.2 Fund Categories

We classified U.S. mutual funds and ETFs into active and passive categories following Pavlova and Sikorskaya (2023).

Active vs. Passive and ETF

1. If et_flag is empty and index_fund_flag is empty, it is an "Active Fund".

2. If et_flag is empty and index_fund_flag is “B” (Index Based Fund), it is an “Active Fund”.
3. If index_fund_flag is D or E, it is an “Index Fund”.
4. If the fund name contains keywords: ‘index’, ‘indx’, ‘ idx ’, ‘s&p’, ‘ sp’, ‘nasdaq’, ‘msci’, ‘crsp’, ‘ftse’, ‘barclays’, ‘ dj ’, ‘ dow ’, ‘jones’, ‘russell’, ‘ nyse ’, ‘wilshire’, ‘400’, ‘500’, ‘600’, ‘1000’, ‘1500’, ‘2000’, ‘2500’, ‘3000’, ‘5000’, ‘ dfa ’, ‘program’, it is an “Index Fund”.
5. If et_flag is not empty, it is an “ETF”.
6. If the fund name contains keywords: ‘spdr’, ‘trackers’, ‘holdrs’, ‘powershares’, ‘street-tracks’, ‘etf’, ‘exchange traded’, ‘exchange-traded’, it is an ETF.

Asset Types

1. If crsp_obj_cd starts with ‘ED’, it is “Equity Domestic”.
2. If crsp_obj_cd starts with ‘EF’, it is “Equity Foreign”.
3. If crsp_obj_cd starts with ‘IF’, it is “Fixed Income Foreign”.
4. If crsp_obj_cd starts with ‘IM’, it is “Fixed Income Domestic Money Market”.
5. If crsp_obj_cd starts with ‘I’ but does not start with ‘IF’ or ‘IM’, it is “Fixed Income Domestic”.
6. If crsp_obj_cd starts with ‘M’, it is “Mixed FI & Equity”.
7. All other categories are “Other”.

We used only fixed-income domestic funds with their primary benchmarks.

A.11.3 Clean Benchmarks

We used the Gemini API to create a mapping, manually checked it, and then used it to unify different names of the same benchmark into a single name, such as unifying ‘lehman aggregate bond index’, ‘bloomberg us agg bond tr usd’, ‘bloomberg capital u.s. aggregate tr usd’ as ‘bloomberg u.s aggregate bond tr usd’.

A.11.4 Correct Mismatched Funds and Benchmarks

If an SEC file contained multiple funds and their corresponding benchmarks in succession, for example, "fund 1, benchmark 1, fund 2, benchmark 2, fund 3, benchmark 3", but the SEC did not provide the `seriesId` for fund 2, or the name of fund 2 in the file did not match the name of fund 2 in the SEC-provided mapping, then both benchmark 1 and benchmark 2 might be assigned to fund 1, leading to incorrect matching. Given the low probability of this situation occurring, we adopted the following method to correct it: if a fund consistently used benchmark A but suddenly showed benchmark B midway (like A A A A B A A A A), benchmark B was clearly a mismatched entry, and we replaced B with A.

A.11.5 Validation

Validation Against MFRR. We validated primary benchmarks extracted from SEC filings against the MFRR time-series benchmark data. Because MFRR data are incomplete, $4.4 trillion in fund assets appear only in our sample, $0.3 trillion appear only in MFRR, and the validation is based on $2.1 trillion in funds that coexist for the same quarter. The matching rate is 81% (based on the latest total net assets); that is, these funds have the same benchmark for the same quarter between MFRR and our extracted data.

Validation Against Morningstar. We included only primary benchmarks for both SEC and Morningstar (MS) benchmarks.

We excluded 205 series IDs (SEC fund IDs) that have more than one primary prospectus benchmark in Morningstar. This is usually because the mapping file incorrectly maps different funds' `secid` (Morningstar ID) to the same `seriesId`.

Because MS benchmarks are the latest snapshot, for each `seriesId` (fund), we compared only the latest benchmark extracted from SEC with MS.

There were 1,982 series ($2.82 trillion) that had benchmarks only from SEC and 248 series ($0.28 trillion) that had benchmarks only from MS. We could compare only 1,360 series ($3.63 trillion) that had benchmarks in both SEC and MS.

The matching rate was 86% by latest total net assets.

B Appendix: Proofs

B.1 Investor Optimization Problem

We first state the optimization problems of a direct investor and a fund manager and then provide the relevant derivations. The direct investor's optimization problem at time $t = 0$ is standard

$$\max_{\theta_D} E[-\exp\{-\gamma W\}]. \tag{B1}$$

As we show below (Appendix B.1.1), this problem reduces to the standard mean-variance optimization:

$$\max_{\theta_D} \quad \theta_D'(\mu - p) - \frac{1}{2}\gamma \underbrace{\theta_D' \Sigma \theta_D}_{\text{Variance}}. \tag{B2}$$

The problem of a fund manager is the following mean-tracking-error-variance optimization problem, reported in the main text and reproduced here:

$$\max_{\theta} \quad (a+b)\theta'(\mu - p) - \frac{1}{2}\gamma \underbrace{((a+b)\theta - b\omega)'\Sigma((a+b)\theta - b\omega)}_{\text{Tracking-error variance}} - Cn. \tag{B3}$$

The first two terms imply that the manager should select assets with higher expected returns and lower tracking-error variance. Unlike a direct investor, the manager is concerned with the fund's tracking-error variance rather than its total variance.[24] This is because the manager is benchmarked to an index, meaning that the zero-risk portfolio from her perspective is one that perfectly replicates the benchmark (specifically, $\theta = \frac{b}{a+b}\omega$). The final component of the manager's trade-off is the portfolio management cost, which discourages including too many assets in the fund's portfolio.

[24] To keep the model elegant and tractable, note that we define returns in per-share terms. In the empirical analyses, however, we use standard per-dollar returns and, strictly speaking, tracking-error variance is not the same object. As in much of the literature using CARA preferences, we acknowledge this inconsistency, but we still prefer to keep our theoretical results in terms of per-share returns, for expositional clarity.

B.1.1 Reduction to Mean–Variance Problems

To evaluate the expectations in the direct investors' and managers' problems, we need the following property.

A useful Gaussian identity. Let $Y \in \mathbb{R}^N$ be Gaussian with mean vector $m \in \mathbb{R}^N$ and covariance matrix $V \in \mathbb{R}^{N\times N}$, i.e., $Y \sim \mathcal{N}(m, V)$. Then for any constant vector $x \in \mathbb{R}^N$ and scalar $\alpha \in \mathbb{R}$,

$$E\Big[\exp^{\alpha x' Y}\Big] = e^{\alpha x' m + \frac{\alpha^2}{2} x' V x}. \tag{B4}$$

Therefore, maximizing $E[e^{\alpha x' Y}]$ is equivalent to maximizing $\alpha x' m + \frac{\alpha^2}{2} x' V x$.

In our application, the vector of asset payoffs D is Gaussian with mean vector μ and covariance matrix Σ, so the identity above applies with $Y = D$, $m = \mu$, and $V = \Sigma$.

Direct investor. For the direct investor,

$$W = W_0 + \theta_D'(D - p),$$

where $D \sim \mathcal{N}(\mu, \Sigma)$. Hence,

$$E[W] = W_0 + \theta_D'(\mu - p), \qquad \mathrm{Var}(W) = \theta_D' \Sigma \theta_D.$$

Applying the property (B4) and dropping constants, it follows that the problem is equivalent to

$$\max_{\theta_D} \ \theta_D'(\mu - p) - \frac{\gamma}{2} \theta_D' \Sigma \theta_D,$$

which yields (B2).

Fund manager. The manager's compensation is

$$w = (a + b)\theta'(D - p) - b\omega'(D - p) - Cn + c.$$

Since D is normal,

$$E[w] = (a+b)\theta'(\mu - p) - b\omega'(\mu - p) - Cn + c,$$

and

$$\text{Var}(w) = ((a+b)\theta - b\omega)'\Sigma((a+b)\theta - b\omega).$$

Applying the property (B4) and dropping constants, it follows that the problem is equivalent to

$$\max_{\theta} \ (a+b)\theta'(\mu - p) - \frac{\gamma}{2}((a+b)\theta - b\omega)'\Sigma((a+b)\theta - b\omega) - Cn,$$

which is equation (3).

B.2 Proof of Proposition 1

Proof of Proposition 1. The solution to the direct investor's problem (B2) is well known and yields the following demand function:

$$\theta_D = \frac{1}{\gamma}\Sigma^{-1}(\mu - p). \tag{B5}$$

Specializing this solution to the case of $\beta_{m,i} = 0$, for which the matrix $\Sigma = \sigma_\epsilon^2 I_N$ is diagonal (with I_N denoting the $N \times N$ identity matrix), gives (4).

The rest of the proof specializes to this uncorrelated case. Take a price vector p as given.

Step 1 (Manager's optimal demand for a given set of assets). Fix any subset $\mathsf{I} \subset \{1, \ldots, N\}$ with $|\mathsf{I}| = n$ and restrict the manager to positions $\theta_i = 0$ for all $i \notin \mathsf{I}$. Since $\Sigma = \sigma_\epsilon^2 I_N$, the manager's objective (3) can be written as

$$\sum_{i \in \mathsf{I}} \left\{ (a+b)\theta_i(\mu_i - p_i) - \frac{\gamma}{2}\sigma_\epsilon^2\Big((a+b)\theta_i - b\omega_i\Big)^2 - C \right\} - \frac{\gamma}{2}\sigma_\epsilon^2 b^2 \sum_{i \notin \mathsf{I}} \omega_i^2.$$

For each $i \in \mathsf{I}$, the first-order condition is

$$(a+b)(\mu_i - p_i) - \gamma\sigma_\epsilon^2(a+b)\Big((a+b)\theta_i - b\omega_i\Big) = 0.$$

Thus

$$\theta_i = \frac{1}{\gamma\sigma_\epsilon^2(a+b)}(\mu_i - p_i) + \frac{b}{a+b}\omega_i, \qquad i \in \mathsf{I}, \tag{B6}$$

and $\theta_i = 0$ for $i \notin \mathsf{I}$.

Step 2 (Manager's value function and marginal benefit). Let $V_p(\mathsf{I})$ denote the manager's value function: the maximized value of the objective in (3) when prices are evaluated at p and the manager is restricted to hold only assets in I. Thus V_p is a static value function, indexed by the price vector that the atomistic manager takes as given. Substituting the optimal holdings from Step 1 gives

$$V_p(\mathsf{I}) = -\frac{\gamma}{2}\sigma_\epsilon^2 b^2 \sum_{k=1}^{N} \omega_k^2 + \sum_{i\in\mathsf{I}} \left[\frac{1}{2\gamma\sigma_\epsilon^2}\Big(\mu_i - p_i + \gamma\sigma_\epsilon^2 b\omega_i\Big)^2 - C\right]. \tag{B7}$$

The first term in (B7) is common across all feasible sets. Therefore, because an individual manager is atomistic and takes prices as given, the incremental value of adding an asset $j \notin \mathsf{I}$ is

$$\Delta_j(p) \equiv V_p(\mathsf{I}\cup\{j\}) - V_p(\mathsf{I}) = \frac{\gamma}{2}\sigma_\epsilon^2\left(\frac{\mu_j - p_j}{\gamma\sigma_\epsilon^2} + b\omega_j\right)^2 - C. \tag{B8}$$

This expression includes the reduction in tracking-error variance that comes from replacing the excluded position $\theta_j = 0$ with the optimal included position. Equivalently, it equals

$$\frac{1}{2\gamma\sigma_\epsilon^2}(\mu_j - p_j)^2 + b\omega_j(\mu_j - p_j) + \frac{\gamma}{2}\sigma_\epsilon^2 b^2\omega_j^2 - C.$$

Step 3 (Equilibrium prices for a conjectured cutoff n). Conjecture that all managers hold exactly the first n assets, ordered by decreasing benchmark weights, i.e., $\mathsf{I} = \{1,\dots,n\}$. Let p^n denote the corresponding candidate equilibrium price vector.

For $i \le n$, use (B6) and the direct-investor demand in market clearing, $\lambda_D\theta_D + \lambda_F\theta =$

$\overline{\theta}$:

$$\lambda_D \frac{1}{\gamma\sigma_\epsilon^2}(\mu_i - p_i^n) + \lambda_F \left(\frac{1}{\gamma\sigma_\epsilon^2(a+b)}(\mu_i - p_i^n) + \frac{b}{a+b}\omega_i \right) = \overline{\theta}_i.$$

Solving gives

$$\mu_i - p_i^n = \gamma\sigma_\epsilon^2 \Big(\lambda_D + \tfrac{\lambda_F}{a+b} \Big)^{-1} \Big(\overline{\theta}_i - \tfrac{b}{a+b}\lambda_F \omega_i \Big), \qquad i \le n, \tag{B9}$$

which yields the first line of (6). For $i > n$, manager demand is zero, so market clearing implies $\lambda_D \theta_{D,i} = \overline{\theta}_i$. Hence

$$\mu_i - p_i^n = \gamma\sigma_\epsilon^2 \frac{\overline{\theta}_i}{\lambda_D}, \qquad i > n, \tag{B10}$$

which yields the second line of (6).

Step 4 (Verification of the cutoff rule and determination of n^*). Using the manager's value function from Step 2, and since managers are atomistic, the relevant no-deviation comparison is made at the candidate equilibrium prices p^n. A pure cutoff equilibrium with cutoff n requires

$$\Delta_i(p^n) \ge 0 \quad \text{for all } i \le n, \qquad \Delta_i(p^n) < 0 \quad \text{for all } i > n.$$

Let

$$A \equiv \left[\lambda_D + \frac{\lambda_F}{a+b} \right]^{-1}.$$

For an included asset $i \le n$, (B9) implies

$$\frac{\mu_i - p_i^n}{\gamma\sigma_\epsilon^2} + b\omega_i = A \left(\overline{\theta}_i + b\lambda_D \omega_i \right).$$

For an excluded asset $i > n$, (B10) implies

$$\frac{\mu_i - p_i^n}{\gamma\sigma_\epsilon^2} + b\omega_i = \frac{\overline{\theta}_i}{\lambda_D} + b\omega_i.$$

Therefore the no-deviation conditions are

$$\frac{\gamma}{2}\sigma_\epsilon^2 \left(A \left(\overline{\theta}_i + b\lambda_D \omega_i \right) \right)^2 - C \ge 0, \qquad i \le n, \tag{B11}$$

and

$$\frac{\gamma}{2}\sigma_\epsilon^2\left(\frac{\overline{\theta}_i}{\lambda_D}+b\omega_i\right)^2-C<0,\qquad i>n. \tag{B12}$$

The maintained assumption in this section is that benchmark weights are proportional to asset supply, $\omega_i=\overline{\theta}_i/W_F$, and assets are ordered so that $\omega_1\geq\cdots\geq\omega_N$. Hence both left-hand sides above are weakly decreasing in the asset rank i. It is therefore sufficient to check the last included asset and the first excluded asset. Thus a pure cutoff n^* is characterized by

$$\frac{\gamma}{2}\sigma_\epsilon^2\left(A\left(\overline{\theta}_{n^*}+b\lambda_D\omega_{n^*}\right)\right)^2-C\geq 0, \tag{B13}$$

and, if $n^*<N$,

$$\frac{\gamma}{2}\sigma_\epsilon^2\left(\frac{\overline{\theta}_{n^*+1}}{\lambda_D}+b\omega_{n^*+1}\right)^2-C<0. \tag{B14}$$

The second condition is vacuous when $n^*=N$. These inequalities verify that the manager's value function is maximized by the cutoff portfolio $\{1,\ldots,n^*\}$: no manager wants to drop an included asset and no manager wants to add an excluded asset.

Finally, substituting the equilibrium prices from Step 3 into (B6) delivers the manager demand (5), while the direct-investor demand is (4). Together with (6) and the cutoff characterization above, this proves the proposition. □

B.3 Mixed equilibrium in Section 3.1

There may exist a mixed equilibrium in which a fraction of managers holds n^*+1 assets (i.e., $1,\ldots,n^*+1$) while the remaining fraction holds n^* assets (i.e., $1,\ldots,n^*$).

Specifically, consider the uncorrelated case of Proposition 1. Suppose that a mass $\varphi\in(0,1)$ of managers holds the top n^*+1 assets, while the remaining mass $1-\varphi$ holds the top n^* assets (assets are ordered by decreasing benchmark weights). For any asset $i\leq n^*$, both manager types hold it and therefore have the same demand given by (5). For $i=n^*$+1, only a fraction φ of managers participates, while for $i>n^*$+1 no managers participate.

Market clearing, for each asset i, is given by

$$\lambda_D \theta_{D,i} + \lambda_F \widetilde{\theta}_{M,i} = \overline{\theta}_i,$$

where $\theta_{D,i}$ is given by (4) and $\widetilde{\theta}_{M,i}$ is aggregate manager demand. Hence equilibrium prices coincide with (6) for all assets except n^*+1. In particular,

$$\mu_{n^*+1} - p_{n^*+1} = \gamma \sigma_\epsilon^2 \left(\lambda_D + \frac{\varphi \lambda_F}{a+b} \right)^{-1} \left(\overline{\theta}_{n^*+1} - \frac{b}{a+b} \varphi \lambda_F \omega_{n^*+1} \right). \tag{B15}$$

To pin down φ, impose indifference between including n^* and n^*+1 assets:

$$V(n^*+1) - V(n^*) = 0.$$

Because managers are atomistic and take prices as given, the value difference from expanding the portfolio from $\{1, \ldots, n^*\}$ to $\{1, \ldots, n^*+1\}$ is equal to the incremental contribution of asset n^*+1 evaluated at equilibrium prices. In the uncorrelated cash-flow case ($\Sigma = \sigma_\epsilon^2 I_N$), this incremental value can be computed in closed form by substituting the manager's optimal holding of asset n^*+1 into the objective function. The resulting indifference condition is therefore equivalent to

$$\frac{\gamma}{2} \sigma_\epsilon^2 \left(\frac{\mu_{n^*+1} - p_{n^*+1}}{\gamma \sigma_\epsilon^2} + b \omega_{n^*+1} \right)^2 - C = 0. \tag{B16}$$

Equivalently, writing $x \equiv \mu_{n^*+1} - p_{n^*+1}$,

$$\frac{1}{2\gamma\sigma_\epsilon^2} x^2 + b \omega_{n^*+1} x + \frac{\gamma}{2} \sigma_\epsilon^2 b^2 \omega_{n^*+1}^2 - C = 0.$$

Equation (B16) implies

$$x = \sqrt{2\gamma\sigma_\epsilon^2 C} - \gamma \sigma_\epsilon^2 \, b \, \omega_{n^*+1},$$

where we take the root that makes $x/(\gamma\sigma_\epsilon^2) + b\omega_{n^*+1}$ positive. Define $y \equiv x/(\gamma\sigma_\epsilon^2)$, so

$$y = \sqrt{\frac{2C}{\gamma\sigma_\epsilon^2}} - b\omega_{n^*+1}.$$

Equating this y to the market-clearing expression in (B15) yields the mixing probability in closed form:

$$\varphi = \frac{(a+b)\Big(\overline{\theta}_{n^*+1} - \lambda_D y\Big)}{\lambda_F\Big(y + b\,\omega_{n^*+1}\Big)}. \tag{B17}$$

Verification. By construction, managers are indifferent between portfolios of sizes n^* and n^*+1 at the mixed-equilibrium prices. Assets $i \leq n^*$ are held by all managers and are priced as in (6); asset n^*+1 is priced according to (B15); and assets $i > n^*+1$ are held by no managers and are therefore priced as in the second line of (6). The remaining no-deviation condition is that no manager wants to add asset n^*+2 (or any lower-ranked asset). Since, under the maintained proportionality assumption, the marginal value of adding an excluded asset is weakly decreasing in rank, it is enough to require

$$\frac{\gamma}{2}\sigma_\epsilon^2\left(\frac{\overline{\theta}_{n^*+2}}{\lambda_D} + b\omega_{n^*+2}\right)^2 - C < 0, \tag{B18}$$

with this condition omitted when $n^*+1 = N$. Together with the condition that the solution in (B17) lies in $(0,1)$, these inequalities deliver the mixed equilibrium described above.

B.4 Proofs for the Case of Correlated Cash Flows

Proof of Lemma 1.

Assume the manager is restricted to include only Asset 1, so $\theta_2 = 0$. Terms in the objective that are constant in θ_1, including the per-asset portfolio-management cost, do not affect the first-order condition. The relevant one-dimensional problem is therefore

$$\max_{\theta_1}\ (a+b)\theta_1(\mu_1 - p_1) - \frac{\gamma}{2}\Big((a+b)\theta - b\omega\Big)'\Sigma\Big((a+b)\theta - b\omega\Big),$$

where $\theta = (\theta_1, 0)'$ and $\omega = (\omega_1, \omega_2)'$. Let

$$y \equiv (a+b)\theta - b\omega = \Big((a+b)\theta_1 - b\omega_1,\ -b\omega_2\Big)'.$$

The first-order condition with respect to θ_1 is

$$(a+b)(\mu_1 - p_1) - \gamma(a+b)\, e_1'\Sigma y = 0,$$

where $e_1 \equiv (1,0)'$. Hence,

$$\mu_1 - p_1 = \gamma\, e_1'\Sigma y = \gamma\Big(\sigma_1^2\Big((a+b)\theta_1 - b\omega_1\Big) + \rho_{12}\sigma_1\sigma_2(-b\omega_2)\Big).$$

Solving for θ_1 yields

$$\theta_1 = \frac{\mu_1 - p_1}{\gamma\sigma_1^2(a+b)} + \frac{b}{a+b}\omega_1 + \frac{b}{a+b}\frac{\rho_{12}\sigma_1\sigma_2}{\sigma_1^2}\,\omega_2,$$

and $\theta_2 = 0$, which gives (10).

If the manager can hold both assets, the first-order conditions take the standard mean–variance form and yield

$$\theta = \frac{1}{\gamma(a+b)}\Sigma^{-1}(\mu - p) + \frac{b}{a+b}\omega,$$

as stated. □

Proof of Lemma 2.

Economy with portfolio management costs (manager includes only Asset 1). By Lemma 1, the manager holds $\theta_2 = 0$. Market clearing in Asset 2 is therefore

$$\lambda_D\theta_{D,2} = \overline{\theta}_2 \quad \Rightarrow \quad \theta_{D,2} = \frac{\overline{\theta}_2}{\lambda_D}.$$

Using $\mu - p = \gamma\Sigma\theta_D$, we have

$$\mu_1 - p_1 = \gamma\Big(\sigma_1^2\theta_{D,1} + \rho_{12}\sigma_1\sigma_2\theta_{D,2}\Big) = \gamma\Big(\sigma_1^2\theta_{D,1} + \rho_{12}\sigma_1\sigma_2\frac{\overline{\theta}_2}{\lambda_D}\Big), \tag{B19}$$

$$\mu_2 - p_2 = \gamma\Big(\rho_{12}\sigma_1\sigma_2\theta_{D,1} + \sigma_2^2\theta_{D,2}\Big) = \gamma\Big(\rho_{12}\sigma_1\sigma_2\theta_{D,1} + \sigma_2^2\frac{\overline{\theta}_2}{\lambda_D}\Big). \tag{B20}$$

Market clearing in Asset 1 is

$$\lambda_D\theta_{D,1} + \lambda_F\theta_1 = \overline{\theta}_1.$$

From Lemma 1, the manager's (restricted) demand for Asset 1 is

$$\theta_1 = \frac{\mu_1 - p_1}{\gamma\sigma_1^2(a+b)} + \frac{b}{a+b}\omega_1 + \frac{b}{a+b}\frac{\rho_{12}\sigma_1\sigma_2}{\sigma_1^2}\,\omega_2.$$

Substituting (B19) into this expression and then into market clearing for Asset 1 yields

$$\Big(\lambda_D + \tfrac{\lambda_F}{a+b}\Big)\theta_{D,1} = \overline{\theta}_1 - \frac{b\lambda_F}{a+b}\Big(\omega_1 + \tfrac{\rho_{12}\sigma_1\sigma_2}{\sigma_1^2}\omega_2\Big) - \frac{\lambda_F}{a+b}\frac{\rho_{12}\sigma_1\sigma_2}{\sigma_1^2}\frac{\overline{\theta}_2}{\lambda_D}.$$

Multiplying both sides by σ_1^2 and rearranging gives

$$\sigma_1^2\theta_{D,1} = A\Big[\sigma_1^2\overline{\theta}_1 + \rho_{12}\sigma_1\sigma_2\overline{\theta}_2 - \frac{b\lambda_F}{a+b}\Big(\sigma_1^2\omega_1 + \rho_{12}\sigma_1\sigma_2\omega_2\Big)\Big] - \rho_{12}\sigma_1\sigma_2\frac{\overline{\theta}_2}{\lambda_D},$$

where $A \equiv \Big[\lambda_D + \frac{\lambda_F}{a+b}\Big]^{-1}$, as in the lemma. Plugging this into (B19) yields the expression in (12) .

Next, substitute $\theta_{D,1}$ into (B20). After collecting terms, this gives

$$p_2 = \mu_2 - \gamma A\Big[\rho_{12}\sigma_1\sigma_2\overline{\theta}_1 + \rho_{12}^2\sigma_2^2\overline{\theta}_2 - \frac{b\lambda_F}{a+b}\Big(\rho_{12}\sigma_1\sigma_2\omega_1 + \rho_{12}^2\sigma_2^2\omega_2\Big)\Big] - \frac{\gamma}{\lambda_D}\sigma_2^2(1-\rho_{12}^2)\overline{\theta}_2,$$

which is the expression for p_2 in (13).

Economy without portfolio management costs (manager includes both assets). When the

manager includes both assets,

$$\theta = \frac{1}{\gamma(a+b)}\Sigma^{-1}(\mu - p) + \frac{b}{a+b}\omega.$$

Substituting this expression together with the direct investor's demand θ_D into market clearing and solving for $(\mu - p)$, we arrive at

$$\mu - p = \gamma A\,\Sigma\Big(\overline{\theta} - \frac{b\lambda_F}{a+b}\omega\Big),$$

Hence, p_1 coincides with (12), and

$$p_2 = \mu_2 - \gamma A\Big[\rho_{12}\sigma_1\sigma_2\overline{\theta}_1 + \sigma_2^2\overline{\theta}_2 - \frac{b\lambda_F}{a+b}\big(\rho_{12}\sigma_1\sigma_2\omega_1 + \sigma_2^2\omega_2\big)\Big],$$

which is the expression stated in the lemma. □

Proof of Lemma 3.

Let p denote the equilibrium price vector in Lemma 2 for the sparse economy in which managers hold only Asset 1, and set $x \equiv \mu - p$. The manager is atomistic and takes p as given. Let $V_p(\mathsf{I})$ denote the manager's value function when the admissible set of assets is I. The proof below assumes that Σ is nonsingular ($|\rho_{12}| < 1$); the limiting case follows by continuity. The lemma suppresses the price subscript and writes $V(\mathsf{I})$ for $V_p(\mathsf{I})$.

Write

$$y \equiv (a+b)\theta - b\omega.$$

Since $(a+b)\theta = y + b\omega$, the objective in (3) can be written as

$$b\omega' x + \Big\{y'x - \frac{\gamma}{2}y'\Sigma y\Big\} - C|\mathsf{I}|.$$

The term $b\omega' x$ is common across admissible sets and therefore cancels from value differences.

Step 1 (Value function when $\mathsf{I} = \{1,2\}$). When the manager can trade both assets, y is

unrestricted. The maximizer is

$$y^* = \frac{1}{\gamma}\Sigma^{-1}x,$$

and hence

$$V_p(\{1,2\}) = b\omega' x + \frac{1}{2\gamma}x'\Sigma^{-1}x - 2C.$$

Step 2 (Value function when $\mathsf{I} = \{1\}$*).* When the manager can trade only Asset 1, $\theta_2 = 0$, or equivalently $y_2 = -b\omega_2$. Let $\sigma_{12} \equiv \rho_{12}\sigma_1\sigma_2$ and $\sigma^2_{2|1} \equiv \sigma_2^2 - \sigma_{12}^2/\sigma_1^2 = \sigma_2^2(1-\rho_{12}^2)$. For a fixed y_2, optimizing over y_1 gives

$$y_1(y_2) = \frac{x_1 - \gamma\sigma_{12}y_2}{\gamma\sigma_1^2}.$$

After substituting out y_1, the concentrated objective is quadratic in y_2 with curvature $-\gamma\sigma^2_{2|1}$. Its unrestricted maximizer is

$$y_2^* = \frac{x_2 - (\sigma_{12}/\sigma_1^2)x_1}{\gamma\sigma^2_{2|1}}.$$

Therefore, the gross value loss from fixing $y_2 = -b\omega_2$ rather than choosing y_2 freely is

$$\frac{\gamma}{2}\sigma^2_{2|1}\Big(y_2^* + b\omega_2\Big)^2.$$

It follows that

$$V_p(\{1\}) = b\omega' x + \frac{1}{2\gamma}x'\Sigma^{-1}x - \frac{\gamma}{2}\sigma^2_{2|1}\Big(y_2^* + b\omega_2\Big)^2 - C.$$

It remains to express y_2^* at the sparse-equilibrium prices. Since Asset 2 is held only by direct investors in that equilibrium, market clearing gives $\theta_{D,2} = \overline{\theta}_2/\lambda_D$. The direct investor's first-order condition is $x = \gamma\Sigma\theta_D$. Hence

$$x_2 - \frac{\sigma_{12}}{\sigma_1^2}x_1 = \gamma\left(\sigma_2^2 - \frac{\sigma_{12}^2}{\sigma_1^2}\right)\theta_{D,2} = \gamma\sigma^2_{2|1}\frac{\overline{\theta}_2}{\lambda_D},$$

so

$$y_2^* = \frac{\overline{\theta}_2}{\lambda_D}.$$

Substituting this expression and $\sigma^2_{2|1} = \sigma^2_2(1-\rho^2_{12})$ into $V_p(\{1\})$ gives

$$V_p(\{1\}) = b\omega' x + \frac{1}{2\gamma} x'\Sigma^{-1}x - \frac{\gamma}{2}\sigma^2_2(1-\rho^2_{12})\left(\frac{\overline{\theta}_2}{\lambda_D} + b\omega_2\right)^2 - C.$$

Step 3 (Difference). Because the manager takes prices as given when comparing admissible sets, we can subtract the two value functions evaluated at the same price vector p. The common term $b\omega' x$ cancels, yielding

$$V(\{1,2\}) - V(\{1\}) = (1-\rho^2_{12})\,\frac{\gamma}{2}\,\sigma^2_2\left(\frac{\overline{\theta}_2}{\lambda_D} + b\omega_2\right)^2 - C,$$

which proves the lemma. □

B.5 Extension of Lemma 3: Asset-Specific Portfolio Management Costs

This subsection extends Lemma 3 to the case in which portfolio management costs are asset-specific. Throughout the subsection, value functions are evaluated at the sparse-equilibrium price vector from Lemma 2, at which managers hold Asset 1 but not Asset 2. Let $C_i \geq 0$ denote the fixed cost of including asset i in the manager's portfolio. For a fixed admissible set $\mathsf{I} \subset \{1,2\}$, define the manager's value function at these prices as

$$V_C(\mathsf{I}) = \max_{\theta:\,\theta_i=0\ \forall i\notin\mathsf{I}} \left\{(a+b)\theta'(\mu-p) - \frac{\gamma}{2}\Big((a+b)\theta - b\omega\Big)'\Sigma\Big((a+b)\theta - b\omega\Big) - \sum_{i\in\mathsf{I}} C_i\right\}. \tag{B21}$$

Equivalently, the manager's compensation contains the cost term $-\sum_{i\in\mathsf{I}} C_i$ rather than $-C|\mathsf{I}|$.

Lemma B1. *Consider the two-asset economy of Lemma 3, but replace the common per-asset portfolio management cost C with asset-specific costs C_1 and C_2. Let $V_C(\mathsf{I})$ denote the value function in (B21), evaluated at the sparse-equilibrium prices described above. The*

incremental value of adding Asset 2 to a portfolio that already contains Asset 1 is

$$V_C(\{1,2\}) - V_C(\{1\}) = (1-\rho_{12}^2)\,\frac{\gamma}{2}\,\sigma_2^2\left(\frac{\overline{\theta}_2}{\lambda_D} + b\omega_2\right)^2 - C_2. \tag{B22}$$

Thus, conditional on already holding Asset 1, the manager adds Asset 2 if and only if

$$(1-\rho_{12}^2)\,\frac{\gamma}{2}\,\sigma_2^2\left(\frac{\overline{\theta}_2}{\lambda_D} + b\omega_2\right)^2 \geq C_2. \tag{B23}$$

Proof of Lemma B1. Let $V^0(\mathsf{I})$ denote the manager's value function at the same given price vector when all portfolio-management costs are set to zero. The calculation in the proof of Lemma 3 implies that the gross gain from adding Asset 2 to a portfolio that already contains Asset 1 is

$$V^0(\{1,2\}) - V^0(\{1\}) = (1-\rho_{12}^2)\,\frac{\gamma}{2}\,\sigma_2^2\left(\frac{\overline{\theta}_2}{\lambda_D} + b\omega_2\right)^2.$$

This gross gain is unchanged when portfolio-management costs are asset-specific. As in Lemma 3, all value functions are evaluated at the same price vector, so the common price-dependent term in the objective cancels from the comparison. The costs C_i are fixed with respect to portfolio weights and therefore do not affect the first-order conditions or the optimal holdings conditional on a given admissible set; they only affect the comparison across admissible sets. In particular,

$$V_C(\mathsf{I}) = V^0(\mathsf{I}) - \sum_{i\in\mathsf{I}} C_i.$$

Moving from $\{1\}$ to $\{1,2\}$ leaves the cost C_1 unchanged and adds the cost C_2. Therefore,

$$\begin{aligned} V_C(\{1,2\}) - V_C(\{1\}) &= V^0(\{1,2\}) - V^0(\{1\}) - C_2 \\ &= (1-\rho_{12}^2)\,\frac{\gamma}{2}\,\sigma_2^2\left(\frac{\overline{\theta}_2}{\lambda_D} + b\omega_2\right)^2 - C_2, \end{aligned}$$

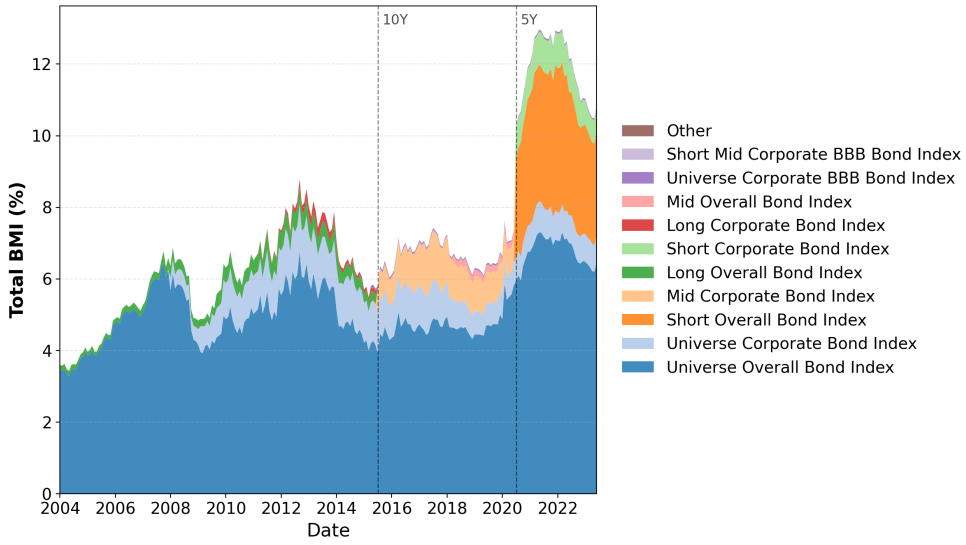
which proves (B22). The inclusion condition follows immediately from requiring the incre-

mental value to be nonnegative. □

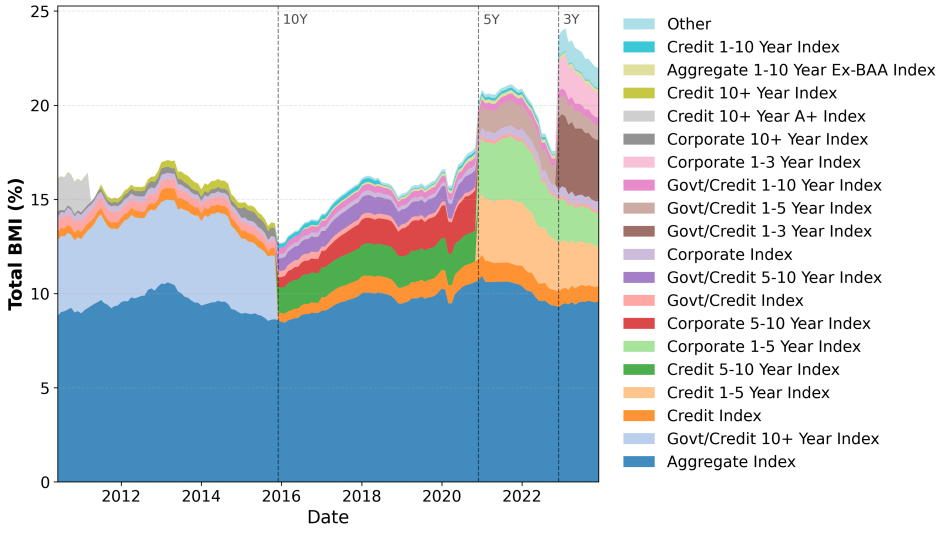
C Appendix: Empirical Analysis

C.1 An illustration of BMI composition

Figure C1: BMI composition of individual bonds



(a) Canada



(b) U.S.

This figure illustrates the composition of BMI by a bond's time to maturity in (a) Canada and (b) the U.S. For the illustration, we use ENBWE 8.85% 07/21/2025 bond in Canada and IBM 7% 10/30/25 bond in the U.S.

C.2 Changes in BMI around maturity cutoffs

Table C1: Changes in BMI around maturity cutoffs

	All bonds	Included	Excluded
Panel A: U.S.			
Total BMI	2.057***	2.077***	2.029***
	(26.41)	(25.57)	(26.43)
Observations	95,772	53,869	41,654
Adj. R^2	0.861	0.863	0.864
Panel B: Canada			
Total BMI	1.860***	1.848***	1.881***
	(48.63)	(45.47)	(44.40)
Observations	21,851	13,307	8,498
Adj. R^2	0.987	0.987	0.988

This table reports the estimates of ΔBMI regressed on the 5-year switch dummy in the U.S. (Panel A) and Canadian (Panel B) samples of bonds with time to maturity between 4 and 6 years. We consider the full sample as well as the samples of included and excluded bonds. A bond is considered *excluded* if it is in the bottom three deciles by size (par value) within its sector-rating-maturity bucket in all indexes to which it belongs and is considered *included* otherwise. All regressions include baseline controls from specification (16), including the logarithm of the bond's par value, numeric credit rating, bid-ask spread quartile, time to maturity, duration, as well as bond and year–month fixed effects. t-statistics based on the wild cluster bootstrap with clustering by issuer and month are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$ based on wild cluster bootstrap p-values.

C.3 BMIs and bond prices in a pooled specification

Table C2: Changes in BMI and bond prices in a pooled sample of included and excluded bonds

	Change in the pricing variable over horizon				
	$t+1$	$t+3$	$t+6$	$t+9$	$t+12$
Panel A: Canada					
Panel A.1: Total return (%)					
$\hat{\beta}_{\text{Excluded}}$	0.025	0.042	0.058	0.066	0.054
	(1.35)	(1.27)	(1.41)	(1.33)	(1.30)
$\hat{\beta}_{\text{Included}}$	0.025	0.094***	0.089***	0.105***	0.125***
	(1.25)	(3.78)	(3.24)	(2.94)	(3.42)
$\hat{\beta}_{\text{Included}} - \hat{\beta}_{\text{Excluded}}$	−0.001	0.052	0.031	0.038	0.071
	(−0.02)	(1.28)	(0.64)	(0.68)	(1.44)
Observations	21,736	21,364	20,830	20,277	19,717
Panel A.2: Yield spread (bps)					
$\hat{\beta}_{\text{Excluded}}$	0.036	−0.429	−1.049	−1.059	−1.208
	(0.15)	(−0.79)	(−1.61)	(−1.53)	(−1.62)
$\hat{\beta}_{\text{Included}}$	−0.337	−0.602*	−1.489***	−2.189***	−2.788***
	(−1.17)	(−1.72)	(−2.71)	(−3.27)	(−3.42)
$\hat{\beta}_{\text{Included}} - \hat{\beta}_{\text{Excluded}}$	−0.372	−0.173	−0.439	−1.129	−1.579
	(−0.98)	(−0.26)	(−0.55)	(−1.30)	(−1.45)
Observations	21,802	21,430	20,896	20,343	19,783
Panel B: U.S.					
Panel B.1: Total return (%)					
$\hat{\beta}_{\text{Excluded}}$	0.031**	0.038**	0.026	0.041	0.017
	(2.15)	(2.08)	(1.09)	(1.51)	(0.59)
$\hat{\beta}_{\text{Included}}$	0.049***	0.063***	0.056***	0.071**	0.074***
	(3.82)	(3.64)	(2.40)	(2.50)	(2.54)
$\hat{\beta}_{\text{Included}} - \hat{\beta}_{\text{Excluded}}$	0.019*	0.025*	0.030	0.030	0.057**
	(1.54)	(1.72)	(1.34)	(1.19)	(2.34)
Observations	95,523	93,711	90,826	87,962	84,969
Panel B.2: Yield spread (bps)					
$\hat{\beta}_{\text{Excluded}}$	−0.370**	−0.709***	−0.318	−0.432	0.208
	(−2.11)	(−2.36)	(−0.79)	(−0.98)	(0.44)
$\hat{\beta}_{\text{Included}}$	−0.789***	−1.188***	−0.976***	−0.784*	−0.862*
	(−3.98)	(−4.29)	(−2.66)	(−1.85)	(−1.83)
$\hat{\beta}_{\text{Included}} - \hat{\beta}_{\text{Excluded}}$	−0.420**	−0.480	−0.658	−0.352	−1.071**
	(−1.89)	(−1.54)	(−1.45)	(−0.72)	(−2.01)
Observations	95,523	93,712	90,836	87,979	84,991

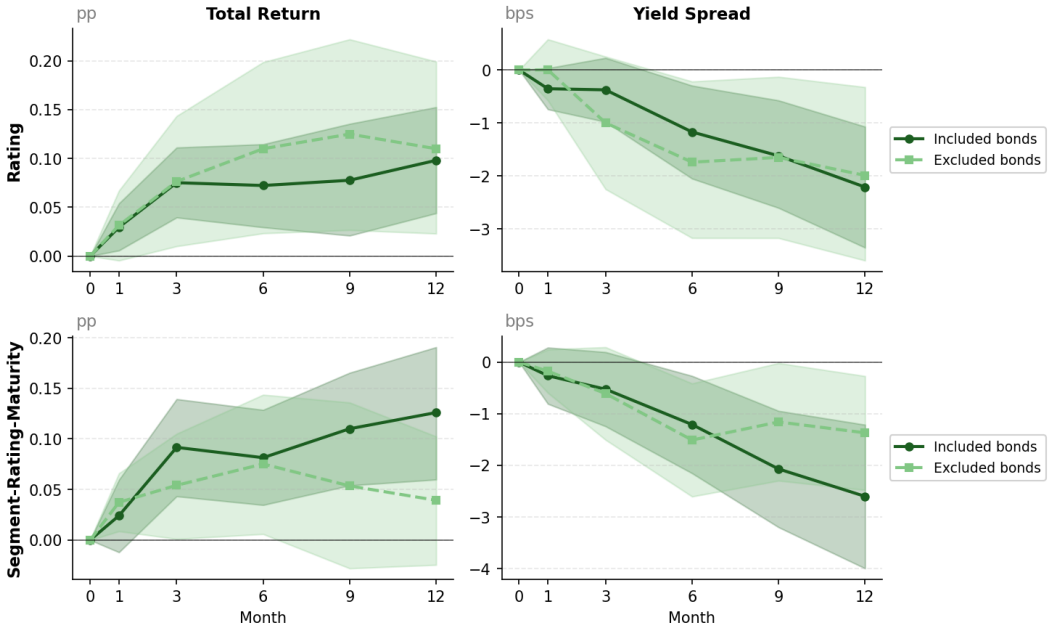
This table tests whether the ΔBMI sensitivity differs between included and excluded bonds in the Canadian (panel A) and U.S. (panel B) samples, for bonds with time to maturity between 4 and 6 years. In each subpanel, each column corresponds to a single pooled regression of the form

$$\Delta Y_{i,t+h} = \beta_E \, \Delta BMI_{i,t} + \delta \left(\Delta BMI_{i,t} \times D\{\text{Included}_{i,t}\} \right) + \gamma_E' \bar{X}_{i,t} + \gamma_D' \left(\bar{X}_{i,t} \times D\{\text{Included}_{i,t}\} \right) + \alpha_i^g + \tau_t^g + \varepsilon_{i,t+h},$$

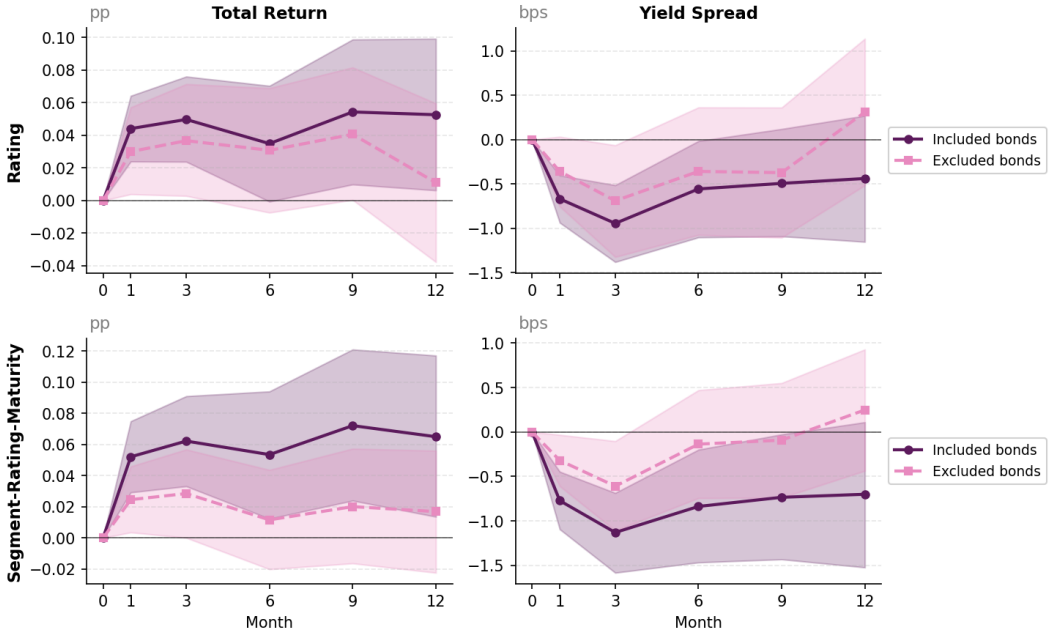
where the dependent variable is the total return or the change in yield spread from month t to the stated horizon, X collects the baseline controls, and g indexes group-specific bond and year–month fixed effects. The reported difference is $\hat{\delta}$, and $\hat{\beta}_{\text{Included}} = \hat{\beta_E} + \hat{\delta}$. A bond is considered *excluded*, i.e., $D\{\text{Included}\} = 0$, if it is in the bottom three deciles by size (par value) within its sector-rating-maturity bucket in all indexes to which it belongs and is considered *included* otherwise. t-statistics based on the wild cluster bootstrap with clustering by issuer and month are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$ based on wild cluster bootstrap p-values.

C.4 BMIs and bond prices with alternative sparsity adjustments

Figure C2: The estimated effect of BMI on bond prices with alternative substitute buckets



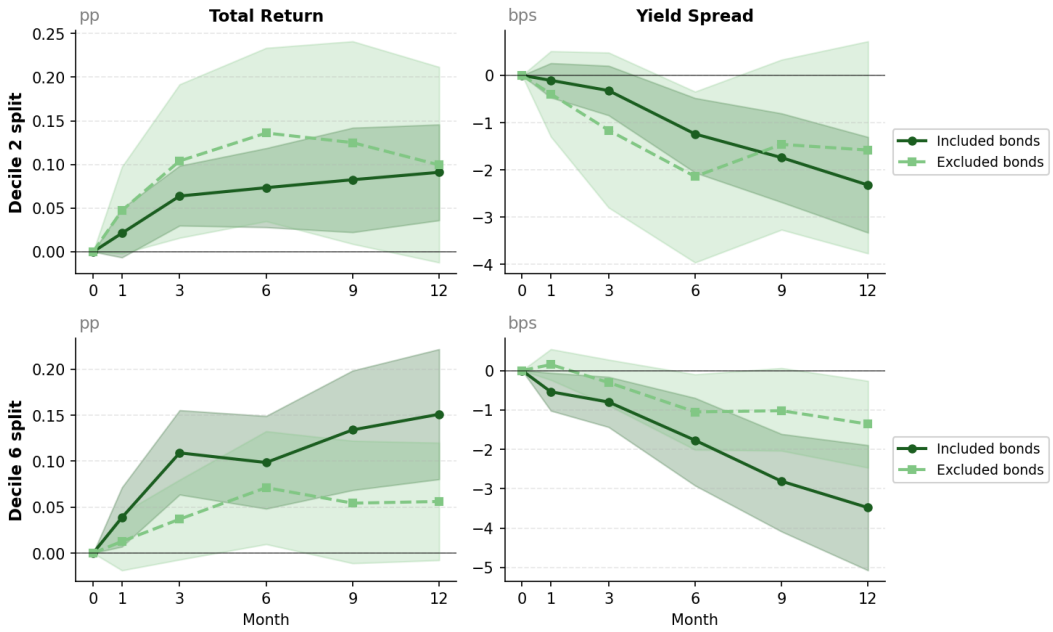
(a) Canada



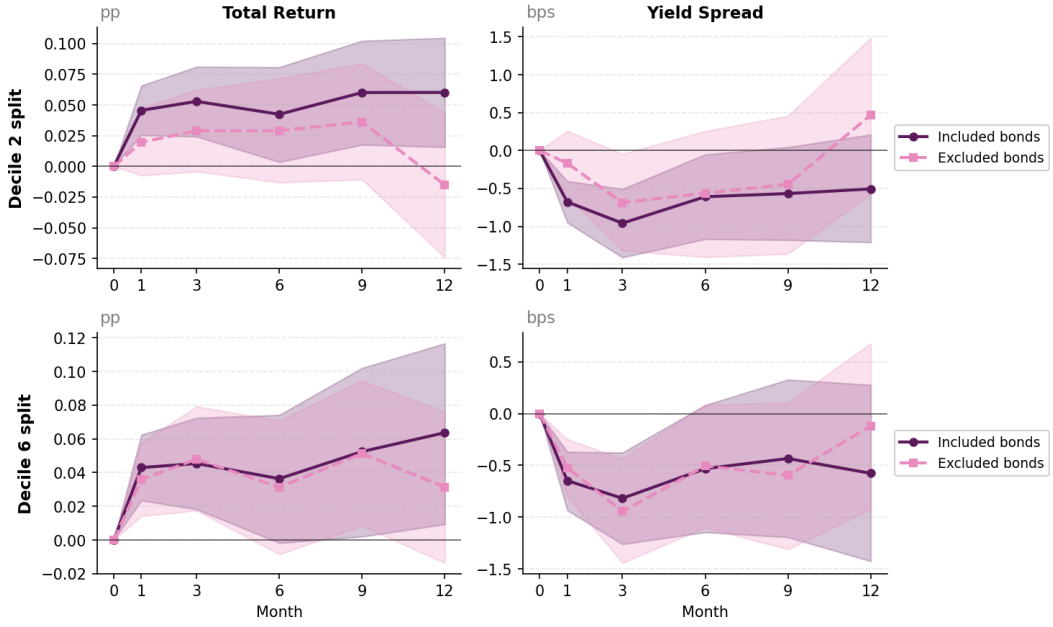
(b) U.S.

This figure plots the estimates and 90% confidence intervals based on regression (16) in (a) Canada and (b) the U.S. In contrast to Figure 4, we use rating and segment-maturity-rating buckets to classify bonds as excluded or included, while keeping the size decile threshold unchanged.

Figure C3: The estimated effect of BMI on bond prices with alternative decile thresholds



(a) Canada

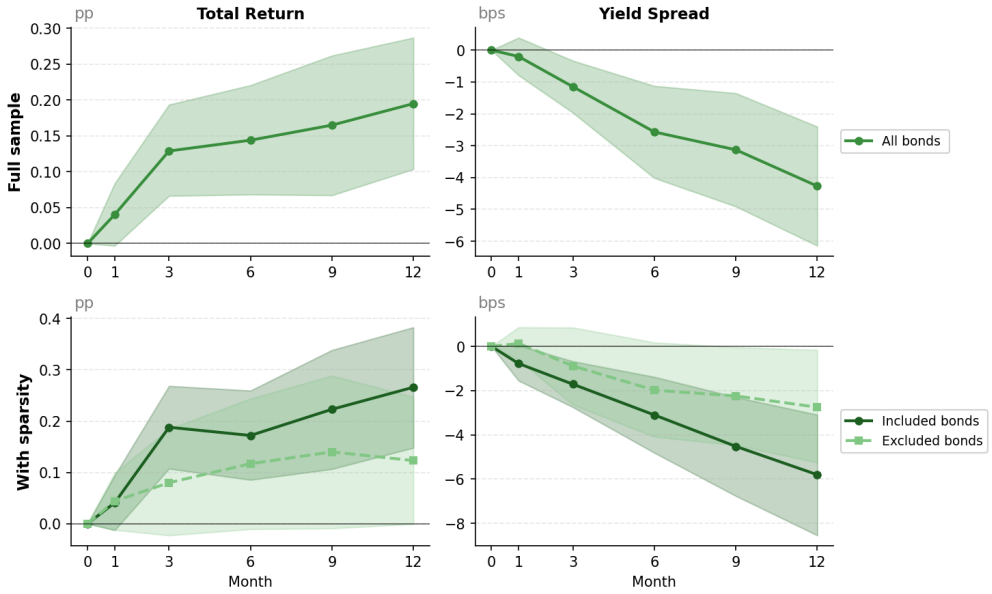


(b) U.S.

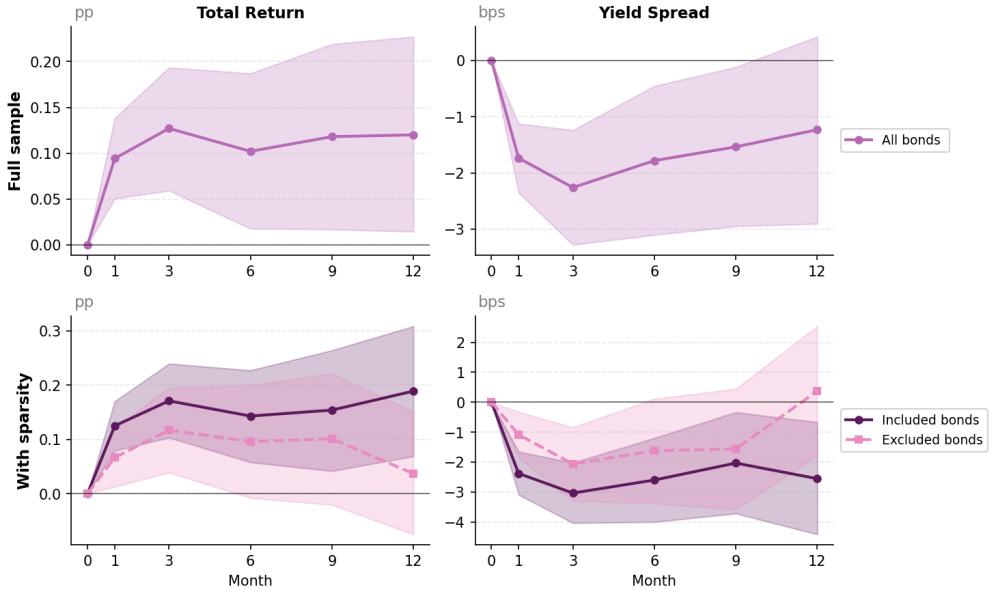
This figure plots the estimates and 90% confidence intervals based on regression (16) in (a) Canada and (b) the U.S. In contrast to Figure 4, we use the first decile and the median of the bond-size distribution within sector-maturity-rating buckets to classify bonds as excluded or included.

C.5 Estimation results for a switch dummy variable

Figure C4: The estimated effect of crossing the 5-year cutoff on corporate bond prices



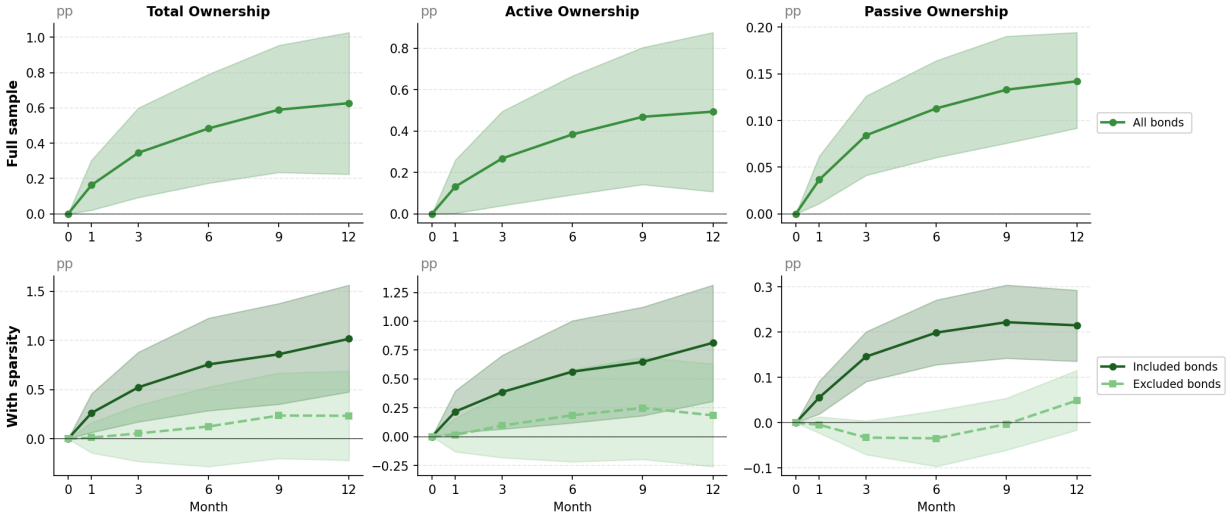
(a) Canada



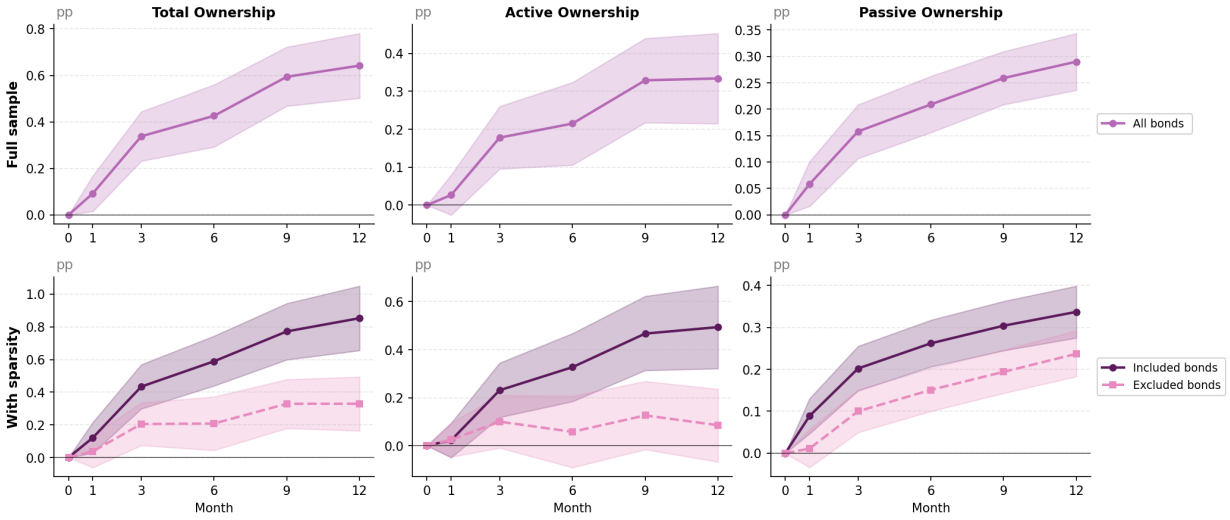
(b) U.S.

This figure plots the estimates and 90% confidence intervals from regression (16) using a switch dummy variable instead of BMI change in (a) Canada and (b) the U.S. The switch dummy variable is 1 in the month when the bond crosses the 5-year maturity cutoff. A bond is considered *excluded* if it is in the bottom three deciles by size (par value) within its sector-rating-maturity bucket in all indexes to which it belongs and is considered *included* otherwise.

Figure C5: The estimated effect of crossing the 5-year cutoff on corporate bond fund ownership



(a) Canada

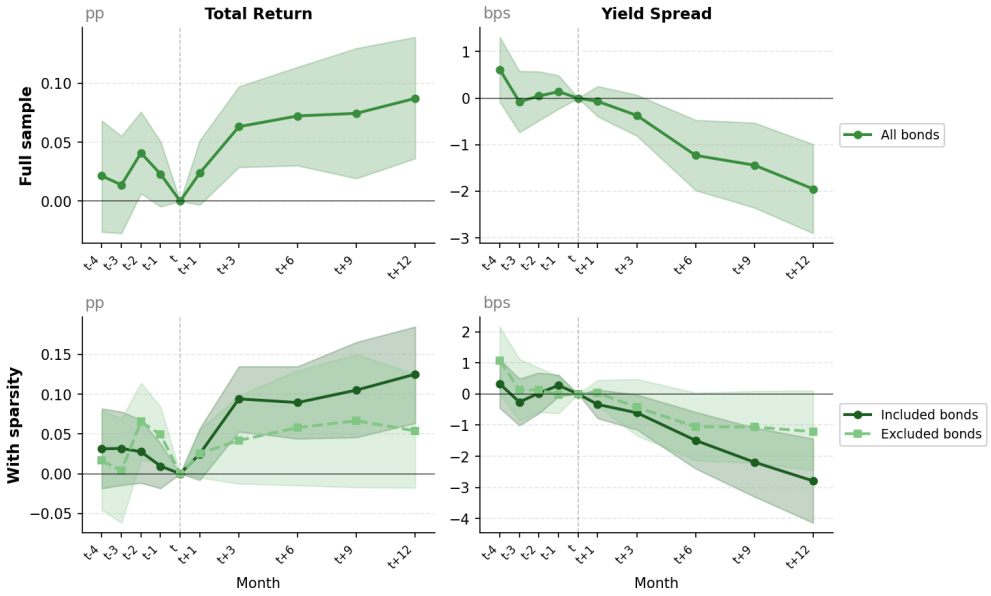


(b) U.S.

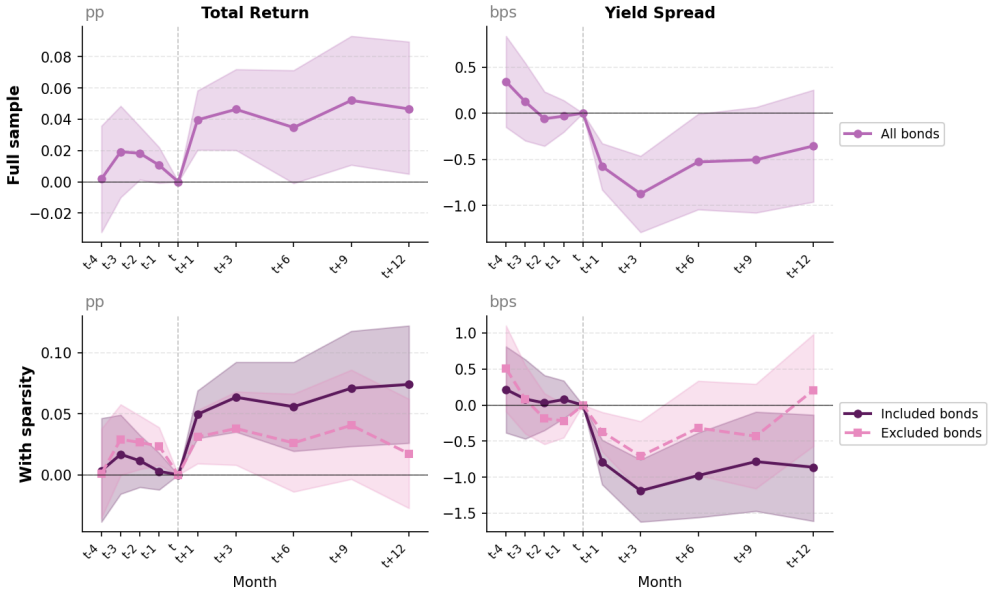
This figure plots the estimates and 90% confidence intervals from regression (16) using a switch dummy variable instead of BMI change in (a) Canada and (b) the U.S. The switch dummy variable is 1 in the month when the bond crosses the 5-year maturity cutoff. A bond is considered *excluded* if it is in the bottom three deciles by size (par value) within its sector-rating-maturity bucket in all indexes to which it belongs and is considered *included* otherwise.

C.6 Bond prices and ownership before crossing the 5-year cutoff

Figure C6: The estimated effect of BMI on corporate bond prices (with lags)



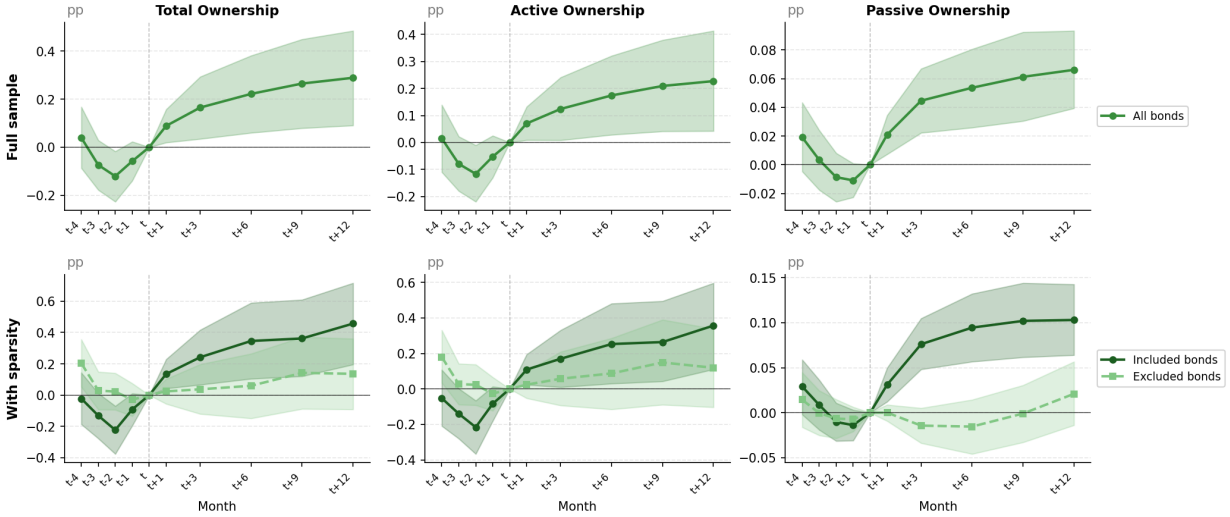
(a) Canada



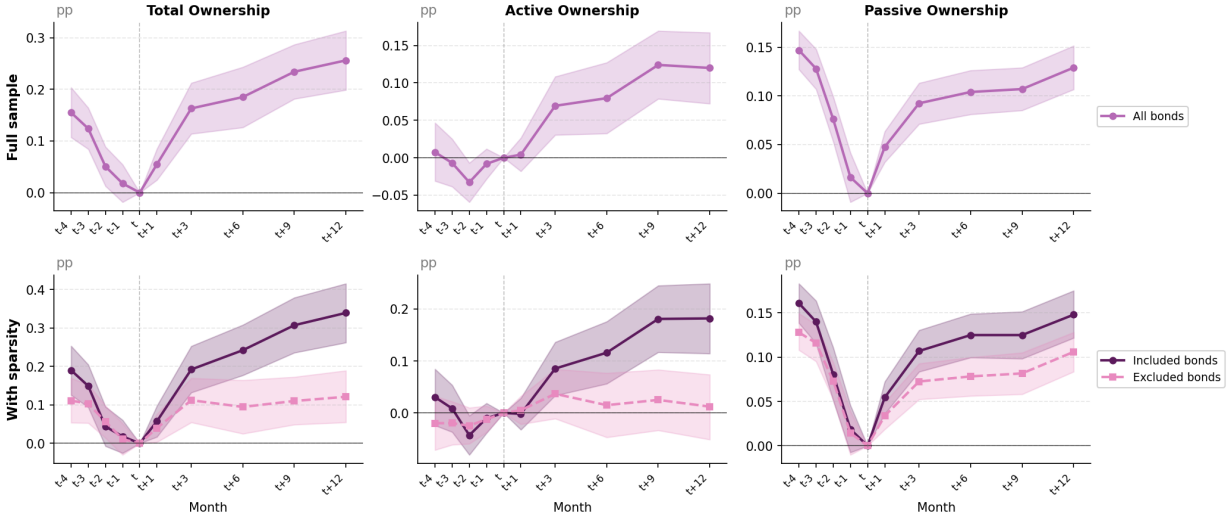
(b) U.S.

This figure plots the estimates and 90% confidence intervals from regression (16) with additional lags of the dependent variables in (a) Canada and (b) the U.S. Control variables are lagged to match the lags of the dependent variables. A bond is considered *excluded* if it is in the bottom three deciles by size (par value) within its sector-rating-maturity bucket in all indexes to which it belongs and is considered *included* otherwise.

Figure C7: The estimated effect of BMI on corporate bond fund ownership (with lags)



(a) Canada

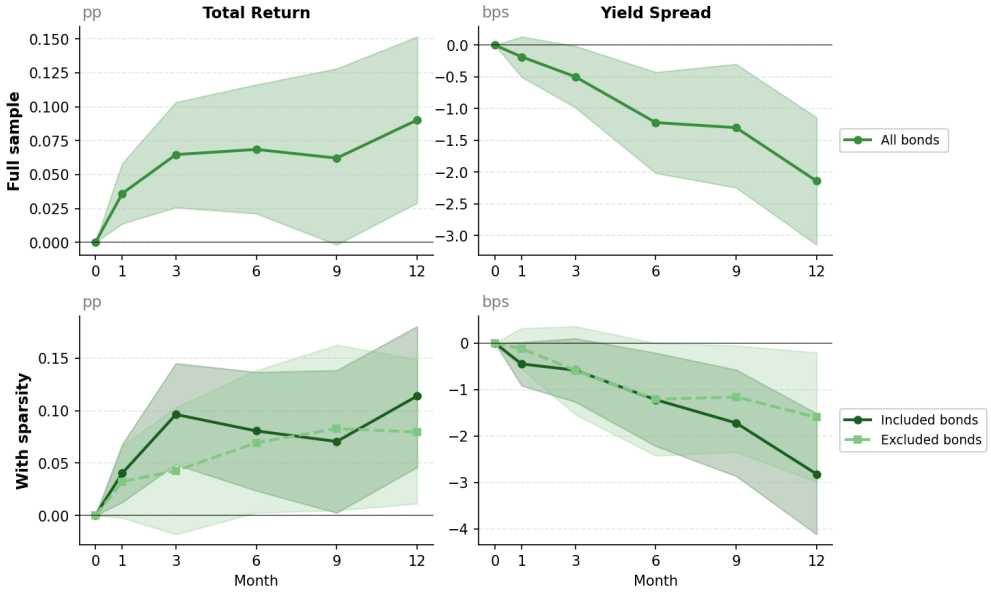


(b) U.S.

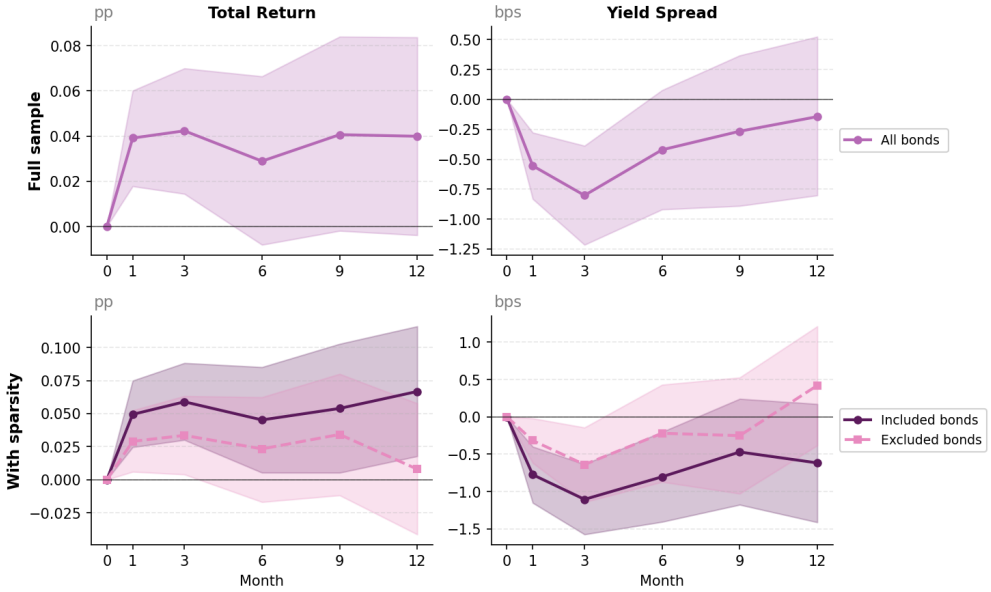
This figure plots the estimates and 90% confidence intervals from regression (16) with additional lags of the dependent variables in (a) Canada and (b) the U.S. Control variables are lagged to match the lags of the dependent variables. A bond is considered *excluded* if it is in the bottom three deciles by size (par value) within its sector-rating-maturity bucket in all indexes to which it belongs and is considered *included* otherwise.

C.7 Estimation results removing recently issued bonds

Figure C8: The estimated effect of BMI on corporate bond prices (removing new issues)



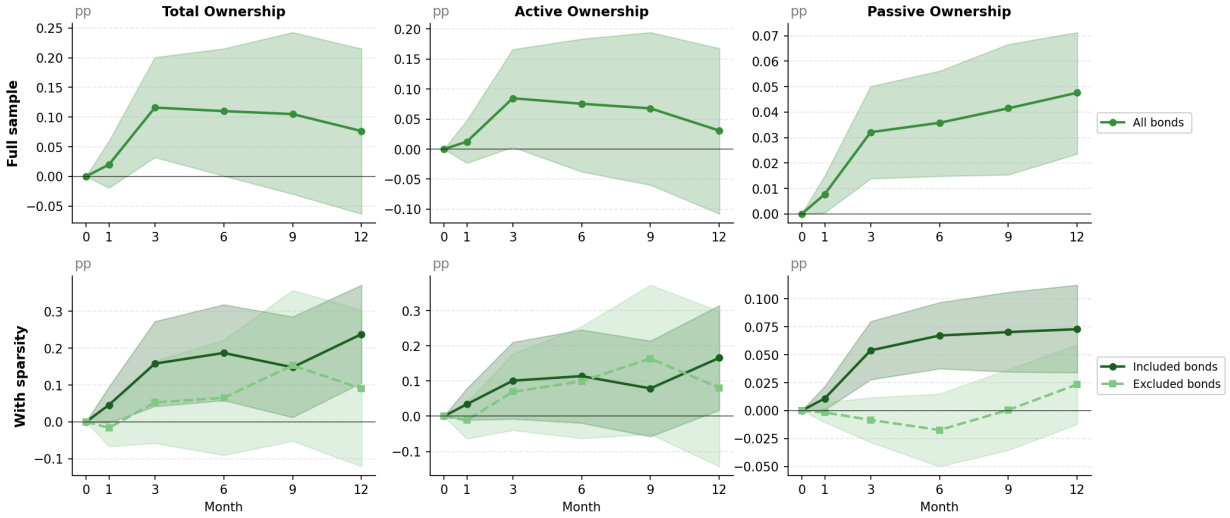
(a) Canada



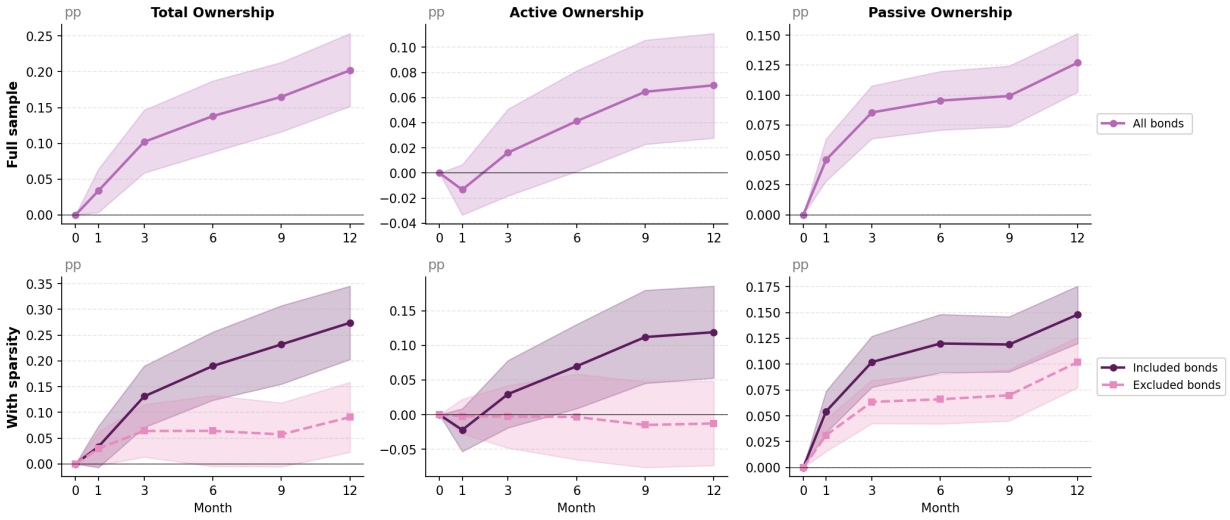
(b) U.S.

This figure plots the estimates and 90% confidence intervals from regression (16) in (a) Canada and (b) the U.S. Bonds are removed from the sample if they were issued within the previous 12 months.

Figure C9: The estimated effect of BMI on corporate bond fund ownership (removing new issues)



(a) Canada



(b) U.S.

This figure plots the estimates and 90% confidence intervals from regression (16) in (a) Canada and (b) the U.S. Bonds are removed from the sample if they were issued within the previous 12 months.

C.8 BMIs and bond ownership in a pooled specification

Table C3: Changes in BMI and bond ownership in a pooled sample of included and excluded bonds

	Change in the ownership variable over horizon				
	t+1	t+3	t+6	t+9	t+12
Panel A: Canada					
Panel A.1: Total ownership					
$\hat{\beta}_{\text{Excluded}}$	0.024	0.037	0.060	0.142	0.136
	(0.50)	(0.40)	(0.50)	(0.99)	(1.00)
$\hat{\beta}_{\text{Included}}$	0.135**	0.242**	0.346**	0.362**	0.457***
	(2.39)	(2.24)	(2.38)	(2.41)	(2.98)
$\hat{\beta}_{\text{Included}} - \hat{\beta}_{\text{Excluded}}$	0.111	0.205	0.286	0.221	0.321*
	(1.54)	(1.44)	(1.47)	(1.08)	(1.68)
Panel A.2: Active ownership					
$\hat{\beta}_{\text{Excluded}}$	0.024	0.057	0.087	0.149	0.119
	(0.51)	(0.64)	(0.75)	(1.04)	(0.89)
$\hat{\beta}_{\text{Included}}$	0.109**	0.170*	0.253*	0.264*	0.357**
	(2.09)	(1.70)	(1.87)	(1.94)	(2.45)
$\hat{\beta}_{\text{Included}} - \hat{\beta}_{\text{Excluded}}$	0.085	0.113	0.165	0.115	0.239
	(1.24)	(0.83)	(0.90)	(0.57)	(1.29)
Panel A.3: Passive ownership					
$\hat{\beta}_{\text{Excluded}}$	0.000	−0.014	−0.015	−0.001	0.021
	(0.02)	(−1.19)	(−0.87)	(−0.05)	(0.98)
$\hat{\beta}_{\text{Included}}$	0.031***	0.076***	0.095***	0.102***	0.103***
	(2.89)	(4.39)	(4.26)	(4.10)	(4.26)
$\hat{\beta}_{\text{Included}} - \hat{\beta}_{\text{Excluded}}$	0.031***	0.090***	0.110***	0.103***	0.082**
	(2.79)	(4.49)	(4.05)	(3.49)	(2.43)
Observations	21,802	21,430	20,896	20,343	19,783
Panel B: U.S.					
Panel B1: Total ownership					
$\hat{\beta}_{\text{Excluded}}$	0.040*	0.112***	0.095**	0.110***	0.121***
	(1.84)	(3.13)	(2.16)	(2.69)	(2.85)
$\hat{\beta}_{\text{Included}}$	0.057**	0.192***	0.242***	0.307***	0.339***
	(2.24)	(5.09)	(5.78)	(6.79)	(7.25)
$\hat{\beta}_{\text{Included}} - \hat{\beta}_{\text{Excluded}}$	0.018	0.080*	0.147***	0.197***	0.217***
	(0.62)	(1.74)	(2.84)	(3.52)	(3.88)
Panel B2: Active ownership					
$\hat{\beta}_{\text{Excluded}}$	0.005	0.037	0.015	0.025	0.012
	(0.31)	(1.20)	(0.38)	(0.65)	(0.31)
$\hat{\beta}_{\text{Included}}$	−0.002	0.085***	0.116***	0.181***	0.182***
	(−0.09)	(2.66)	(3.12)	(4.48)	(4.26)
$\hat{\beta}_{\text{Included}} - \hat{\beta}_{\text{Excluded}}$	−0.007	0.049	0.101*	0.156***	0.169***
	(−0.28)	(1.14)	(1.99)	(2.87)	(3.06)
Panel B3: Passive ownership					
$\hat{\beta}_{\text{Excluded}}$	0.034***	0.073***	0.078***	0.082***	0.106***
	(3.59)	(5.74)	(5.33)	(5.64)	(7.50)
$\hat{\beta}_{\text{Included}}$	0.055***	0.107***	0.125***	0.125***	0.148***
	(4.82)	(7.09)	(7.82)	(7.83)	(9.37)
$\hat{\beta}_{\text{Included}} - \hat{\beta}_{\text{Excluded}}$	0.021***	0.034***	0.046***	0.043***	0.042***
	(3.14)	(3.42)	(3.78)	(3.37)	(3.03)
Observations	95,523	93,712	90,836	87,979	84,991

This table tests whether the ΔBMI sensitivity differs between included and excluded bonds in the Canadian (panel A) and U.S. (panel B) samples, for bonds with time to maturity between 4 and 6 years. In each subpanel, each column corresponds to a single pooled regression of the form

$$\Delta Y_{i,t+h} = \beta_E \, \Delta BMI_{i,t} + \delta \left(\Delta BMI_{i,t} \times D\{\text{Included}_{i,t}\} \right) + \gamma_E' \bar{X}_{i,t} + \gamma_D' \left(\bar{X}_{i,t} \times D\{\text{Included}_{i,t}\} \right) + \alpha_i^g + \tau_t^g + \varepsilon_{i,t+h},$$

where the dependent variable is the change in fund ownership (total, active, or passive) in percentage points from month t to the stated horizon, X collects the baseline controls, and g indexes group-specific bond and year–month fixed effects. The reported difference is $\hat{\delta}$, and $\hat{\beta}_{\text{Included}} = \hat{\beta_E} + \hat{\delta}$. A bond is considered *excluded*, i.e., $D\{\text{Included}\} = 0$, if it is in the bottom three deciles by size (par value) within its sector-rating-maturity bucket in all indexes to which it belongs and is considered *included* otherwise. t-statistics based on the wild cluster bootstrap with clustering by issuer and month are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$ based on wild cluster bootstrap p-values.

C.9 Additional price impact estimates

Table C4: Instrumented price impact with rating–maturity–time fixed effects

	Price impact for								
	All bonds			Included bonds			Excluded bonds		
	t+3	t+6	t+12	t+3	t+6	t+12	t+3	t+6	t+12
Panel A: U.S.									
Total return (%)	0.74**	0.69*	0.48**	1.02*	0.53**	0.47**	0.70	4.83	0.79
	(2.49)	(1.96)	(2.43)	(1.82)	(2.15)	(2.55)	(1.38)	(0.23)	(0.99)
Yield spread (bps)	−13.24**	−11.01*	−5.73*	−18.24*	−9.76**	−6.21**	−12.09	−59.06	−8.59
	(−2.44)	(−1.91)	(−1.94)	(−1.77)	(−2.08)	(−2.18)	(−1.30)	(−0.23)	(−0.80)
$\widehat{\Delta Own}$	0.07***	0.06***	0.11***	0.06*	0.09***	0.15***	0.06*	0.01	0.05
	(2.95)	(2.57)	(3.57)	(1.97)	(2.79)	(3.24)	(1.72)	(0.23)	(1.15)
KP rk Wald F-stat.	8.76	6.72	12.91	3.83	7.84	10.71	2.89	0.054	1.28
Observations	86,062	83,394	77,913	49,088	47,722	44,811	36,144	34,855	32,323
Panel B: Canada									
Total return (%)	0.51*	0.40	0.31*	0.57*	0.34*	0.30**	0.81	2.54	0.42
	(1.80)	(1.65)	(1.88)	(1.74)	(1.71)	(2.06)	(0.60)	(0.20)	(0.67)
Yield spread (bps)	−3.77	−7.03*	−7.07*	−3.29	−5.46*	−6.57**	−15.13	−52.68	−10.34
	(−1.40)	(−1.67)	(−1.94)	(−1.28)	(−1.69)	(−2.05)	(−0.59)	(−0.21)	(−0.79)
$\widehat{\Delta Own}$	0.14**	0.18*	0.26**	0.19*	0.27*	0.39***	0.06	0.03	0.11
	(2.00)	(1.87)	(2.54)	(1.90)	(1.95)	(2.64)	(0.66)	(0.24)	(0.92)
KP rk Wald F-stat.	3.83	3.45	6.43	3.60	3.81	7.04	0.37	0.042	0.83
Observations	20,575	20,051	18,969	12,500	12,176	11,510	7,969	7,768	7,351

This table reports 2SLS estimates of specification (18) across 3-, 6-, and 12-month horizons. It also reports the first-stage estimates of (17) and the corresponding Kleibergen–Paap rk Wald F-statistics. All regressions include baseline controls as well as bond fixed effects and rating–maturity–year–month fixed effects. The U.S. specification uses one-year maturity buckets; the Canadian specification uses 2.5-year maturity buckets. The dependent variable is either the total return or the change in yield spread from month t to the stated horizon. t-statistics, based on standard errors clustered by issuer and rating–maturity–year–month, are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table C5: Instrumented price impact in the 4–5.5 year maturity window

	Price impact for								
	All bonds			Included bonds			Excluded bonds		
	t+3	t+6	t+12	t+3	t+6	t+12	t+3	t+6	t+12
Panel A: U.S.									
Total return (%)	0.50**	0.21	0.27	0.58**	0.26*	0.31**	0.55	0.38	0.37
	(2.11)	(1.09)	(1.47)	(2.33)	(1.83)	(2.25)	(0.99)	(0.43)	(0.42)
Yield spread (bps)	−9.58**	−2.89	−2.65	−11.00**	−4.68*	−3.82*	−9.48	−1.27	0.34
	(−2.38)	(−0.97)	(−0.99)	(−2.43)	(−1.91)	(−1.71)	(−0.98)	(−0.10)	(0.02)
$\widehat{\Delta Own}$	0.08***	0.10***	0.14***	0.11***	0.17***	0.21***	0.05	0.03	0.04
	(3.36)	(3.40)	(4.25)	(3.26)	(4.37)	(4.81)	(1.44)	(0.80)	(0.95)
KP rk Wald F-stat.	11.19	11.13	18.49	10.59	18.64	22.81	2.06	0.66	0.93
Observations	67,151	65,195	60,880	39,031	37,989	35,639	27,868	26,958	24,998
Panel B: Canada									
Total return (%)	1.36	1.32	0.64	0.66	0.48	0.53	−0.67	−0.74	−0.60
	(0.65)	(0.60)	(1.14)	(1.29)	(1.22)	(1.55)	(−0.47)	(−0.56)	(−0.30)
Yield spread (bps)	−14.93	−26.89	−14.86	−5.78	−9.44	−11.75	14.73	16.62	19.44
	(−0.67)	(−0.61)	(−1.10)	(−1.20)	(−1.27)	(−1.56)	(0.49)	(0.62)	(0.34)
$\widehat{\Delta Own}$	0.05	0.06	0.13	0.14	0.19	0.25*	−0.05	−0.07	−0.04
	(0.71)	(0.63)	(1.21)	(1.46)	(1.32)	(1.79)	(−0.48)	(−0.58)	(−0.31)
KP rk Wald F-stat.	0.44	0.35	1.43	2.10	1.75	3.19	0.30	0.43	0.13
Observations	16,559	16,119	15,216	10,155	9,881	9,334	6,348	6,181	5,825

This table reports 2SLS estimates of specification (18) across 3-, 6-, and 12-month horizons. It also reports the first-stage estimates of (17) and the corresponding Kleibergen–Paap rk Wald F-statistics. We restrict the sample to bonds with time to maturity between 4 and 5.5 years. All regressions include the baseline controls as well as bond and year–month fixed effects. The dependent variable is either the total return or the change in yield spread from month t to the stated horizon. t-statistics, based on standard errors clustered by issuer and year–month, are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

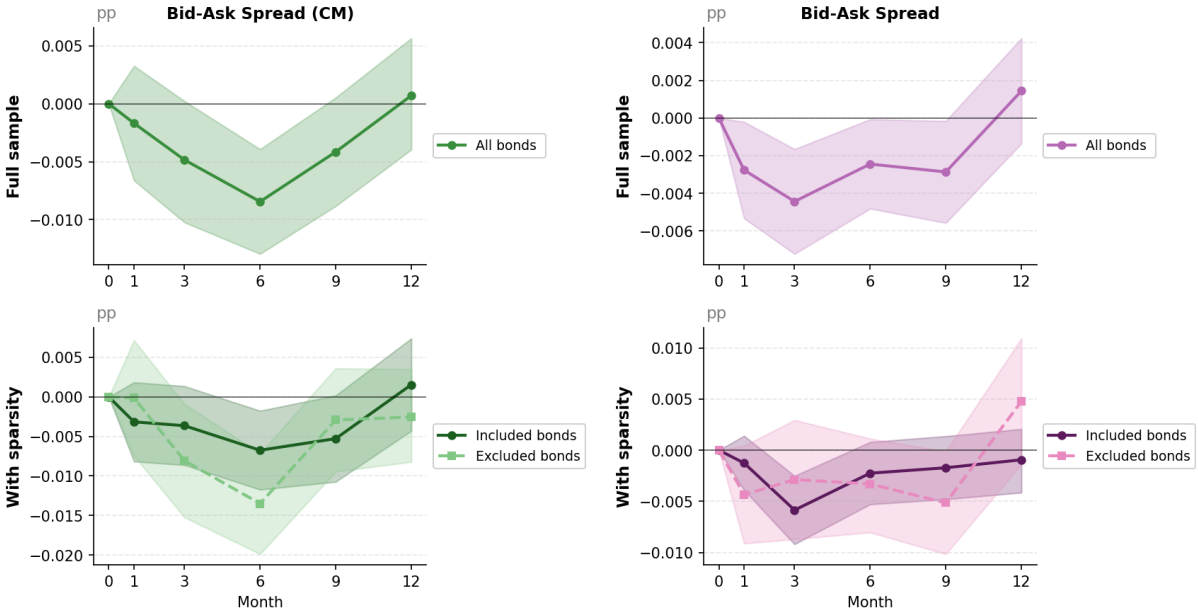
Table C6: Instrumented price impact with time-varying control slopes

	Price impact for								
	All bonds			Included bonds			Excluded bonds		
	t+3	t+6	t+12	t+3	t+6	t+12	t+3	t+6	t+12
Panel A: U.S.									
Total return (%)	0.53**	0.45*	0.58**	0.47**	0.26*	0.40**	0.72	1.92	1.96
	(2.35)	(1.74)	(2.07)	(2.22)	(1.95)	(2.37)	(1.39)	(0.56)	(0.67)
Yield spread (bps)	−9.04**	−6.59	−7.31*	−8.37**	−4.90**	−5.28*	−11.02	−22.96	−23.03
	(−2.35)	(−1.65)	(−1.79)	(−2.13)	(−2.04)	(−1.86)	(−1.35)	(−0.54)	(−0.62)
$\widehat{\Delta Own}$	0.08**	0.08*	0.09**	0.10	0.13**	0.14**	0.06	0.03	0.02
	(2.42)	(1.92)	(2.12)	(1.51)	(2.39)	(2.44)	(1.22)	(0.42)	(0.43)
KP rk Wald F-stat.	9.69	6.64	8.43	7.29	12.61	11.20	2.95	0.34	0.53
Observations	86,370	83,694	78,198	49,516	48,144	45,217	36,592	35,302	32,743
Panel B: Canada									
Total return (%)	0.47***	1.34	0.63	1.88	0.97	1.08	−0.06***	0.13***	−0.53
	(3.31)	(1.34)	(1.24)	(0.50)	(0.67)	(0.74)	(−3.32)	(4.64)	(−0.88)
Yield spread (bps)	−6.62***	−19.44	−3.80	−23.64	−22.22	−24.92	0.88**	1.05*	14.15
	(−3.32)	(−1.34)	(−0.70)	(−0.50)	(−0.69)	(−0.74)	(2.59)	(1.78)	(0.94)
$\widehat{\Delta Own}$	0.04	0.01	0.01	0.05	0.09	0.10	0.02	−0.05	−0.15
	(0.48)	(0.11)	(0.04)	(0.38)	(0.52)	(0.53)	(0.13)	(−0.32)	(−0.79)
KP rk Wald F-stat.	0.35	0.02	0.01	0.25	0.46	0.56	0.01	0.23	1.26
Observations	20,645	20,118	19,031	12,567	12,243	11,577	8,035	7,831	7,410

This table reports 2SLS estimates of specification (18) across 3-, 6-, and 12-month horizons. It also reports the first-stage estimates of (17) and the corresponding Kleibergen–Paap rk Wald F-statistics. All regressions include bond and year–month fixed effects and interact the controls for the logarithm of the bond's par value, numeric credit rating, bid-ask spread quartile, time to maturity, and duration with year–month fixed effects. The dependent variable is either the total return or the change in yield spread from month t to the stated horizon. t-statistics, based on standard errors clustered by issuer and year–month, are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

C.10 BMIs and bond liquidity

Figure C10: The estimated effect of BMI on corporate bond liquidity



(a) Canada (b) U.S.

This figure plots the estimates and 90% confidence intervals from regression (16) in (a) Canada and (b) the U.S. using changes in bid-ask spread as a dependent variable.

C.11 Benchmark issue-size eligibility cutoff and fund ownership

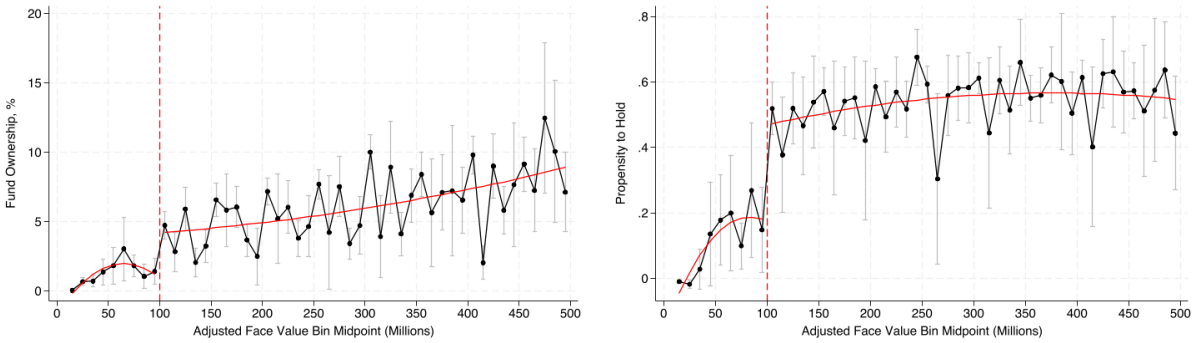
In this section, we study fund ownership patterns around the issue-size cutoffs applied by the benchmark providers. Bonds below the size cutoffs are deemed ineligible for benchmark index inclusion. In our theory, such bonds have benchmarking intensities of zero and are predicted to have low ownership shares of benchmarked funds.

In Figures C11 and C12, we present binned coefficient plots, reflecting the average level of fund ownership as well as propensity to hold bonds of different sizes around the index cutoffs. Figures C14 and C13 report the same analysis within active and passive funds. The binned coefficient plots show that fund ownership and the propensity to hold a bond both rise with issue size, and that ownership sharply increases as bonds cross the eligibility threshold. Away from the cutoff the relationship between ownership and size is smooth, while near the cutoff there is a visible upward jump which is present for both fund types. In the U.S., the passive jump is the more pronounced, whereas in Canada, the active jump is the larger of the two, on both the intensive and the extensive margin.

We complement the binned coefficient plots with formal discontinuity estimates from a global-polynomial specification, reported in Table C7. The estimates use the full corporate-bond sample—every bond-quarter, with no size restriction. For each outcome we regress ownership on an indicator that a bond's amount outstanding is above the eligibility cutoff, together with a polynomial in the running variable (issue size relative to the cutoff) interacted with the above-cutoff indicator, plus time fixed effects; the coefficient on the above-cutoff indicator is the level jump at the cutoff net of the smooth size gradient. Because the polynomial is fit on the entire size distribution, it absorbs the size–ownership trend, and we report both a linear and a quadratic polynomial. As in the binned coefficient plots, we study the intensive margin (ownership share) and the extensive margin (an indicator that at least one fund of the given type holds the bond), separately for total, active, and passive funds.

Table C7 confirms the visual pattern: crossing the cutoff raises fund ownership discontinuously in the full sample as well as in the active and passive subsamples. These jumps are positive and predominantly significant on the intensive margin, and on the extensive

Figure C11: Ownership around the FTSE size cutoff in Canada

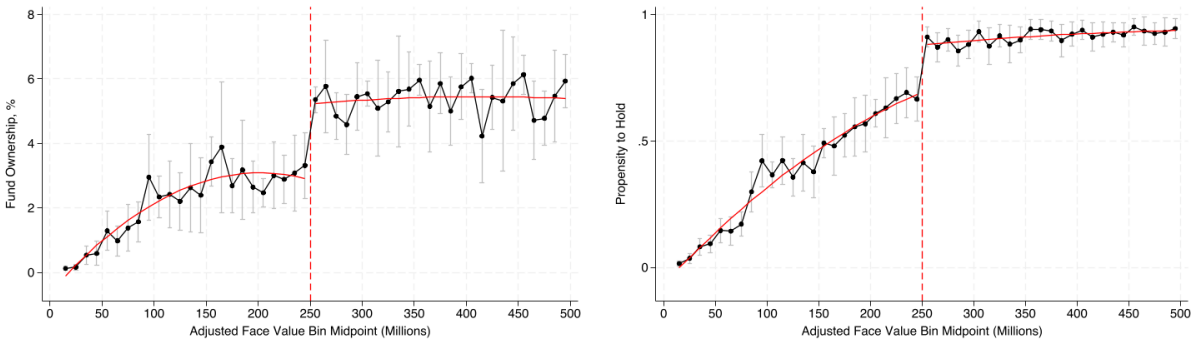


(a) Fund ownership

(b) Propensity to hold

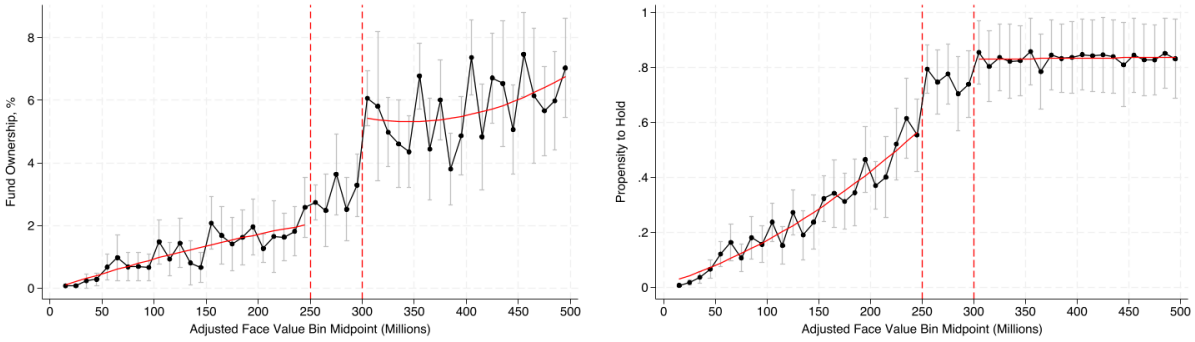
This figure illustrates fund ownership and propensity to hold corporate bonds around the FTSE index issue-size eligibility cutoff of 100 million CAD. The sample includes the union of corporate bonds in the FTSE Canada Overall Universe Bond Index and corporate bonds in the Bloomberg Canada universe.

Figure C12: Ownership around the Bloomberg size cutoff in the U.S.



(a) Fund ownership, pre Q2-2017

(b) Propensity to hold, pre Q2-2017



(c) Fund ownership, from Q2-2017

(d) Propensity to hold, from Q2-2017

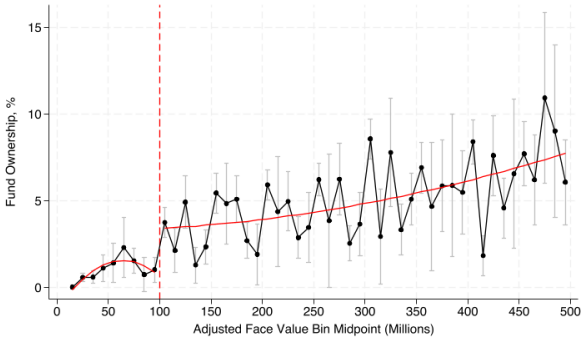
This figure illustrates fund ownership and propensity to hold corporate bonds around the Bloomberg index issue-size eligibility cutoff of 250 million USD before the second quarter of 2017 and around the new size cutoff of 300 million USD after the second quarter of 2017. The sample includes the union of corporate bonds in the Bloomberg Aggregate Bond Index and corporate bonds in the WRDS sample.

margin everywhere except the post-2017Q2 U.S. market, which we discuss below. On the intensive margin in the U.S., total ownership jumps by about 0.6–0.8 pp before 2017Q2 and 1.4–2.4 pp after; the passive jump exceeds the active jump but is of the same order of magnitude. Canada shows a much larger active response: crossing the 100M CAD cutoff raises active ownership by about 1.5–2.3 pp—roughly seven times the corresponding passive jump of 0.2–0.3 pp. The extensive margin mirrors this pattern: the propensity to hold jumps most for passive funds in the U.S. (+0.08 to +0.22 before 2017Q2) and most for active funds in Canada (+0.15 to +0.21). The ownership discontinuity is therefore economically meaningful for active funds in both countries and it is especially large in Canada.

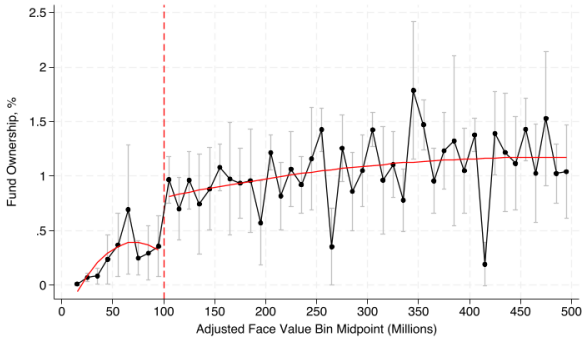
The U.S. results after 2017Q2 come with one interpretation caveat. When Bloomberg raised the issue-size eligibility cutoff from 250M to 300M in 2017Q2, bonds in the 250–300M band that had been index members became ineligible, yet funds—especially active funds—tended to keep positions they had already held. This inertia mechanically sustains ownership just below the new cutoff and flattens the observed ownership patterns, obscuring the discontinuity; it is most visible on the extensive margin, where the all-positions estimate in Table C7 is economically negligible and, under the quadratic control, slightly negative. To separate forward-looking behavior from grandfathered holdings, Table C8 re-estimates the post-2017Q2 discontinuity from new positions only—restricting each bond-quarter's ownership to funds that did not already hold the bond before the cutoff moved. Focusing on new positions confirms a positive and significant eligibility jump for all funds on both margins: the extensive-margin quadratic-polynomial discontinuity increases from essentially zero to +0.10 for active and +0.14 for passive funds, and the active intensive jump is, if anything, larger than in the all-positions sample.[25]

[25]This asymmetry connects our evidence to Feldhütter, Poulsen, and Woschitz (2026), who study the same 2017 change in Bloomberg's issue-size eligibility requirement and document that passive funds sell bonds that become ineligible. Our results are consistent with theirs on the passive side and add the active-fund margin: active funds largely do not sell the positions they already hold in the affected size segment, but they also stop initiating new positions in it.

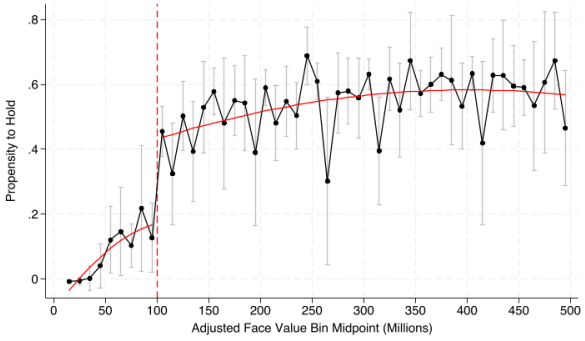
Figure C13: Active and passive ownership around the FTSE size cutoff in Canada



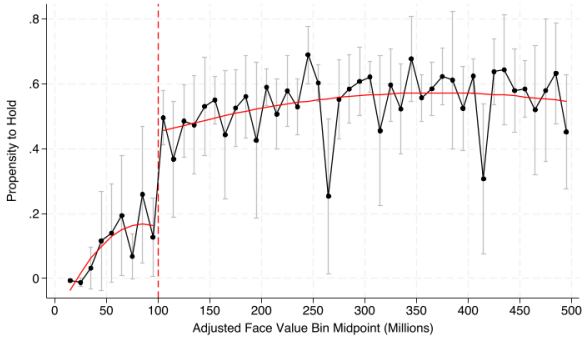
(a) Active fund ownership



(b) Passive fund ownership



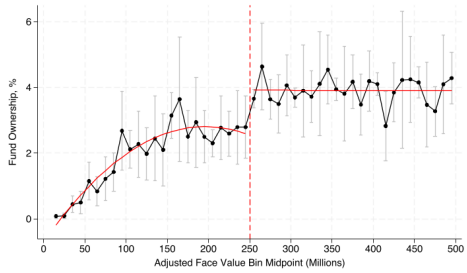
(c) Active fund propensity to hold



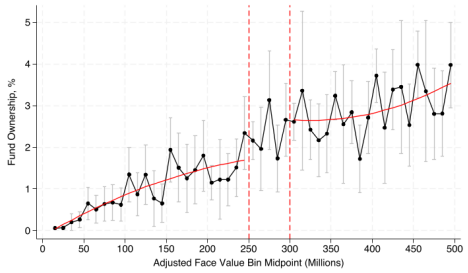
(d) Passive fund propensity to hold

This figure splits Canadian fund ownership of corporate bonds by fund type and margin around the FTSE index issue-size eligibility cutoff of 100 million CAD. The intensive margin ("fund ownership") is holdings as a share of amount outstanding; the extensive margin ("propensity to hold") is an indicator that at least one fund of the given type holds the bond. Each point is a size-bin coefficient from a regression with report-date fixed effects, which we complement with 95% confidence intervals with standard errors clustered by CUSIP and date.

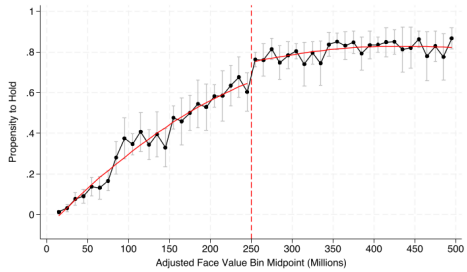
Figure C14: Active and passive ownership around the Bloomberg size cutoff in the U.S.



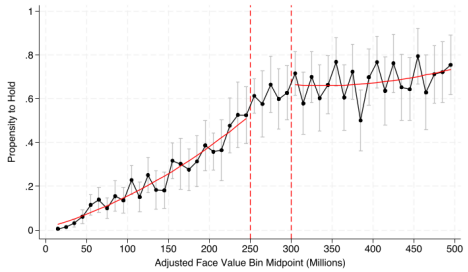
(a) Active fund ownership, pre Q2-2017



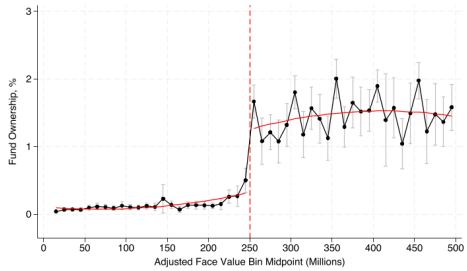
(b) Active fund ownership, from Q2-2017



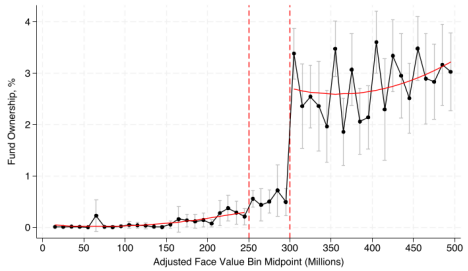
(c) Active fund propensity to hold, pre Q2-2017



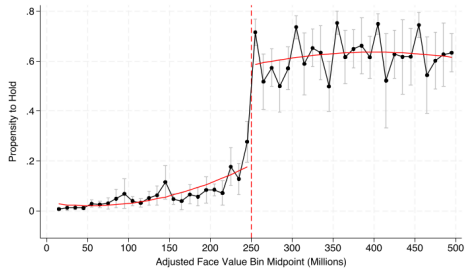
(d) Active fund propensity to hold, from Q2-2017



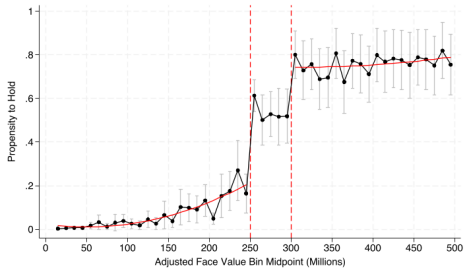
(e) Passive fund ownership, pre Q2-2017



(f) Passive fund ownership, from Q2-2017



(g) Passive fund propensity to hold, pre Q2-2017



(h) Passive fund propensity to hold, from Q2-2017

This figure splits U.S. fund ownership of corporate bonds by fund type and margin around the Bloomberg index issue-size eligibility cutoff (250 million USD before Q2-2017, 300 million USD after). The intensive margin ("fund ownership") is holdings as a share of amount outstanding; the extensive margin ("propensity to hold") is an indicator that at least one fund of the given type holds the bond. Each point is a size-bin coefficient from a regression with quarter fixed effects, which we complement with 95% confidence intervals with standard errors clustered by CUSIP and quarter.

Table C7: Ownership discontinuity at the benchmark size-eligibility cutoff

	Intensive margin (ownership %)		**Extensive margin (P(held))**	
	Linear	Quadratic	Linear	Quadratic
Panel A: Canada (cutoff 100M CAD)				
Total	2.569***	1.688***	0.151***	0.103**
	(5.97)	(3.33)	(4.07)	(2.46)
Active	2.256***	1.482***	0.211***	0.150***
	(6.00)	(3.30)	(6.00)	(3.65)
Passive	0.309***	0.192*	0.166***	0.110***
	(3.16)	(1.68)	(4.40)	(2.59)
Observations		471,503		
Panel B: U.S., pre-2017Q2 (cutoff 250M USD)				
Total	0.798***	0.616***	0.079***	0.049***
	(6.19)	(4.75)	(9.10)	(7.23)
Active	0.140	0.271**	0.097***	0.080***
	(1.39)	(2.53)	(8.40)	(7.44)
Passive	0.694***	0.347***	0.224***	0.083***
	(11.02)	(6.61)	(11.58)	(6.08)
Observations		458,502		
Panel C: U.S., post-2017Q2 (cutoff 300M USD)				
Total	2.446***	1.440***	0.033**	-0.036***
	(8.40)	(5.68)	(2.50)	(-3.24)
Active	0.748***	0.603***	0.085***	0.024
	(4.95)	(3.89)	(5.11)	(1.51)
Passive	1.729***	0.847***	0.188***	0.016
	(10.50)	(6.69)	(8.78)	(1.14)
Observations		862,503		

Each entry is the estimated level jump g at the index issue-size eligibility cutoff, i.e. the coefficient on the above-cutoff indicator in the global-polynomial regression

$$y_{i,t} = \alpha + g \cdot \mathbf{1}(size_{i,t} > c) + \sum_{k=1}^{K} \left[\beta_k \, r_{i,t}^k + \gamma_k \, r_{i,t}^k \cdot \mathbf{1}(size_{i,t} > c)\right] + \delta_t + \varepsilon_{i,t},$$

where $r_{i,t} = size_{i,t} - c$ is amount outstanding relative to the cutoff c, "Linear" sets the polynomial order $K = 1$ and "Quadratic" sets $K = 2$, and δ_t is a time fixed effect. The polynomial in size (interacted with the above-cutoff indicator) absorbs the smooth size–ownership gradient, so g is the discontinuity at the cutoff. On the intensive margin the dependent variable is a bond's fund ownership as a percentage of amount outstanding, so g is in percentage points; on the extensive margin it is an indicator that at least one fund of the given type holds the bond, so g is a change in the holding probability. t-statistics (in parentheses) are based on standard errors double-clustered by CUSIP and time. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table C8: New positions versus all positions: ownership discontinuity at the U.S. benchmark size-eligibility cutoff, post-2017Q2

	Intensive margin (ownership %)		**Extensive margin (P(held))**	
	Linear	Quadratic	Linear	Quadratic
Panel A: All positions				
Total	2.446***	1.440***	0.033**	-0.036***
	(8.40)	(5.68)	(2.50)	(-3.24)
Active	0.748***	0.603***	0.085***	0.024
	(4.95)	(3.89)	(5.11)	(1.51)
Passive	1.729***	0.847***	0.188***	0.016
	(10.50)	(6.69)	(8.78)	(1.14)
Observations		862,503		
Panel B: New positions only				
Total	2.222***	1.455***	0.187***	0.071***
	(7.63)	(6.13)	(8.65)	(4.01)
Active	0.989***	0.707***	0.186***	0.104***
	(7.11)	(5.37)	(9.14)	(5.57)
Passive	1.242***	0.752***	0.298***	0.137***
	(7.77)	(6.41)	(10.69)	(6.88)
Observations		862,503		

The table reports the level jump g at the post-2017Q2 U.S. issue-size eligibility cutoff (300M USD), i.e. the coefficient on the above-cutoff indicator in the global-polynomial regression of Table C7 ("Linear" sets the polynomial order $K = 1$, "Quadratic" sets $K = 2$; time fixed effects are included). Panel A uses all fund positions. Panel B measures ownership from new positions only: a bond-quarter's holdings are restricted to funds that did not hold that bond in any report before 2017Q2, so grandfathered holdings retained after Bloomberg raised the cutoff from 250M to 300M do not contribute. The corporate-bond sample, ownership definitions, and passive/active split are identical to Table C7. t-statistics (in parentheses) are based on standard errors double-clustered by CUSIP and time. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.